

루이스의 상대역 이론(Counterpart Theory)과 본질주의*

김준영
(서울대 철학과)

1. 들어가며

데이빗 루이스(David Lewis : 1941~2001)가 제시한 상대역 이론(Counterpart Theory)은 자연언어의 양상문장을 일차술어논리로 번역하는 형식체계이자, 그것에 대한 의미를 정해주는 의미론이다. 예를 들면, ‘헤겔은 필연적으로 존재한다’라는 문장은 ‘ $(x)(Wx \rightarrow \neg(\exists y)(Iyx \& Cya))$ ’와 같이 상대역 이론으로 번역될 수 있으며, 이는 ‘헤겔의 상대역이 모든 가능 세계에 존재한다’라는 식으로 풀이될 수 있다. 루이스는 기존에 존재하던 양상문장에 대한 형식체계와 의미론이 가진 문제점들을 극복하기 위해 이러한 이론을 구상하게 되었다.

본 논문은 크게 두 가지의 목적을 가지고 있다. 일차적인 목적은 루이스가 주장하는 상대역 이론의 형이상학적 체계를 살펴보고, 상대역 관계에 대해 고찰해보는 것이다. 상대역 이론의 형이상학적 체계는 그가 제시하는 8개의 공준만으로 충분히 설명될 수 있는 명료한 체계이다. 그러나 상대역을 결정해주는 상대역 관계(counterpart relation)는 그렇게 명료하지만은 않은데, 이러한 문제는 펠드먼(Feldman)에 의해 제기되었다. 루이스는 상대역 관계를 맥락 의존적인 개념으로 수정함으로써 이러한 문제를 해결할 수 있었으나, 결과적으로 상대역 개념이 상당히 모호한 개념이 되었다. 특히, 이러한 수정된 상대역 이론으로는 'Counterpart Theory and Quantified Modal

* 본고는 2012년도 학사학위 논문을 축약한 것임.

Logic'(1968)에서 그가 제시한 본질적 속성 정의에 약간의 문제가 생기는 것 같다. 이에 본 논문의 또 다른 목적은, 상대역 이론을 바탕으로 한 새로운 본질적 속성 정의를 제시하고자 하는 것이다.

이에 이 논문은 다음의 순서를 따른다. 2장에서는 양상논리에 대한 기본적인 이해를 바탕으로, 형이상학 체계로서의 상대역 이론에 대해 살펴볼 것이다. 필자는 루이스가 제시한 8개의 공준을 바탕으로 그의 형이상학 체계를 설명하고자 한다. 3장에서는 상대역 이론에서 중요한 부분을 차지하는 상대역 관계 개념에 대해서 살펴볼 것이다. 상대역 관계는 유사성을 기준으로 성립되는 관계로, 루이스의 상대역 이론이 통세계 동일성의 문제를 피할 수 있게 해준다. 4장에서는 펠드먼의 반론을 살펴보고, 그로 인해 수정된 상대역 관계 개념과 루이스의 본질적 속성 정의가 갖고 있는 문제를 살펴본다. 필자에 따르면, 그의 본질적 속성 정의는 그 자체로는 별다른 정보성을 지니지 못한다. 5장에서는 루이스의 본질적 속성 정의가 아닌 새로운 본질적 속성 정의를 제시한다. 이 본질적 속성 정의는 본질을 정도의 개념으로 치환시키는데, 이를 위해 필자는 '본질도'(degree of *essence*)라는 새로운 개념을 도입할 것이다. 6장에서는 이 정의가 기존의 본질 논의에서 가질 수 있는 장점을 설명함으로써, 이 새로운 본질적 속성 정의를 옹호할 것이다. 마지막으로 7장에서는, 지금까지 이 논문에서 제기된 주장에 의해서 특권적 지위를 상실한 크립키 식의 형이상학적 필연성에 대해 살펴볼 것이다. 상대역 이론에 따르면 대상 동일성은 반드시 성립하는 것은 아니며, 그러한 점에서 형이상학적 필연성은 논리적 필연성과 같은 수준의 지위를 가질 수 없을 것이다.

2. 양화양상논리(Quantified Modal Logic)와 상대역 이론

이 논문은 형이상학 체계로서의 상대역 이론을 그 주제로 한다.¹⁾ 그러나 형이상학으로서의 상대역 이론은 의미론, 더 나아가 형식 체계로서의 상대역 이론과 밀접히 연관되어 있기 때문에, 형식 체계로서의 상대역 이론을 먼저 살펴보는 것이 적절할 것 같다.

그보다 먼저 우리 자연 언어의 양상문장과 그에 대한 양화양상논리(QML) 체계를 간단히 살펴보자. 우리의 자연언어체계는 다음과 같은 양상 문장을 지니고 있다.

- (1) 매두사가 사람인 것은 가능하다.
- (2) 총각이 결혼하지 않은 남자라는 것은 필연적이다.

이러한 양상문장은 우리의 일상생활에서 흔하게 쓰이는 문장들이다. 그러나 기존의 일차술어논리는 이러한 양상언어를 형식화할 수 없었으며, 이에 자연스럽게 양상언어를 형식화할 수 있는 체계를 고안하게 되었다. 그 중 우리가 가장 쉽게 접할 수 있는 체계는 기존의 일차술어논리에 양상 오퍼레이터 '□'와 '◇'를 추가한 양화양상논리(QML) 체계이다.²⁾ 이러한 양화양상논리는 20세기 초 루이스(C. I. Lewis)에 의해 처음으로 발전되었다.³⁾ 그는 기초논리학에 양상적 공리들을 추가시켜 “공리적으로 엄밀하게 체계화된 양상논리체계들을 고안해 낸다.”⁴⁾ 이러한 QML 체계 하에서 위의 두 문장은 다음과

1) 선우환은 루이스의 상대역 이론을 (1) 형식 체계로서의 상대역 이론, (2) 양상언어에 대한 의미론으로서의 상대역 이론, (3) 가능세계 형이상학으로서의 상대역 이론이라는 세 가지 서로 다른 의미로 구분한다. 이는 타당한 분류라고 여겨지며, 따라서 본고에서도 이 분류를 그대로 적용하도록 하겠다.

2) '□'는 '~는 필연적이다'로, '◇'는 '~는 가능하다'로 각각 번역된다. 또한 '□A = df ~◇~A' 혹은 '◇A = df ~□~A'로 상호 정의될 수 있다. 따라서 양상 오퍼레이터 '□'와 '◇' 중 하나만 추가된 체계로 봐도 무방하다.

3) C. I. Lewis, *A Survey of Symbolic Logic*, California Univ. Press, 1918.

4) 손병홍, 『가능세계의 철학』, 소피아, 2004, p.22.

같이 번역될 수 있다.

(1') $\Diamond Pa$

(2') $\Box(x)(Bx \rightarrow Ux)$

그러나 루이스는 이런 QML문장들을 어떻게 해석해야할지, 특히 진리조건을 어떻게 부여해야 하는지에 대해서는 설명하지 못했다. 이는 그 이후, 크립키(Kripke), 몽태규(Montague), 힌티카(Hintikka) 등이 양상문장을 해석하는 형식 의미론은 제시하면서부터 알 수 있게 되었다. 이들은 QML 문장을 해석하기 위해서 ‘가능 세계’라는 개념을 도입했다. 아이디어만 간단히 살펴보면 다음과 같다. 우리 세계에서 접근 가능한 어떤 가능 세계들을 상정했을 때, 그 가능세계들에서 그 존재자의 상태를 살펴봄으로써 이러한 양상문장의 진리조건을 가려낼 수 있다는 것이다. 이에 따라 (1)은, 우리 세계에서 접근 가능한 가능 세계 중 한 세계에서라도 매두사가 사람인 세계가 있다면, 참인 문장이다. 그리고 (2)는 우리 세계에서 접근 가능한 모든 가능 세계에서 총각이 결혼하지 않는 남자라면 참인 문장이라고 할 수 있다. 크립키 등은 이와 같은 방식으로 가능 세계를 도입한 QML의 의미를 창안하여, 양상문장을 적절히 해석해낼 수 있었다. 그러나 이와 같은 의미론들은 통세계 동일성의 문제 등 몇몇 문제점을 안고 있다.⁵⁾ 이에 루이스는 이러한 문제점들을 피해갈 수 있는 새로운 형식 체계인 ‘상대역 이론’을 제시하게 된다.

형식체계로서의 상대역 이론은 양상양화논리와 달리 일차술어논리 내에서 이루어진다. 즉, 양상 오퍼레이터와 같은 것들을 추가하지 않고 양상문장을 번역해낼 수 있다는 것이다. 이러한 상대역 이론은 아래에 주어진 네 개의 원초적 술어를 가지고 있다.⁶⁾

5) 통세계 동일성 문제는 논문의 뒷부분에서 다룰 것이다. 그 밖의 문제들에 대해서는, *Ibid.*, pp.24-32 참조.

6) 이 부분은 Lewis, D., “Counterpart Theory and Quantified Modal Logic”, 1968, p.114을 기본으로, 김영배, 「D.루이스의 양상실재론 옹호」, 서울대학교 대학원, 1994, pp.51-52 와 선우환, 「양화양상논리(QUANTIFIED MODAL LOGIC) 의미론과 루이스의 상대역이론(COUNTERPART THEORY) -양상의

- Wx (x 는 가능세계이다)
- Ixy (x 는 가능세계 y 속에 있다)
- Ax (x 는 현실적이다)
- Cxy (x 는 y 의 상대역이다)

그리고 상대역 이론은 아래의 8개의 공준을 지닌다.

- P1 : $(x)(y)(Ixy \rightarrow Wy)$
(무언가를 그 안에 담고 있는 것은 모두 세계이다)
- P2 : $(x)(y)(z)(Ixy \& Ixz \rightarrow y = z)$
(어떠한 것도 두 세계 속에 있지 않다)
- P3 : $(x)(y)(Cxy \rightarrow (\exists z)Ixz)$
(무언가의 상대역인 것은 어떤 세계 속에 있다)
- P4 : $(x)(y)(Cxy \rightarrow (\exists z)Iyz)$
(무언가를 상대역으로 지니는 것은 어떤 세계 속에 있다)
- P5 : $(x)(y)(z)(Ixy \& Izy \& Cxz \rightarrow x = z)$
(어떠한 것도 그것이 속한 세계에서 다른 어떤 것의 상대역이 아니다)
- P6 : $(x)(y)(Ixy \rightarrow Cxx)$
(세계 속에 있는 어떤 것도 자기 자신의 상대역이다)
- P7 : $(\exists x)(Wx \& (y)(Iyx \leftrightarrow Ay))$
(어떠한 세계는 모든 현실적인 것 그리고 오직 그것들만을 담고 있다)
- P8 : $(\exists x) Ax$
(어떤 것은 현실적이다)

예를 들어 설명하는 것이 빠를 것이다. 앞의 (1),(2)를 상대역 이론으로 번역해보자. (1), (2)는 다음과 같이 번역될 수 있다.

- (1') $(\exists x)(Cxa \& Px)$
- (2') $(x)(Wx \rightarrow (y)((Iyx \& By) \rightarrow Uy))$

미론의 두 대안과 또 다른 대안 -, 서울대학교 대학원, 1993, pp.17-19를 참조하였다.

(1'')는 ‘메두사의 상대역이면서 사람인 존재가 존재한다’는 의미이고, (2'')는 ‘모든 가능 세계에서 총각은 결혼하지 않은 남자이다’라는 의미이다. 이와 같이 상대역 이론은 ‘가능 세계’와 그 안에 존재하는 ‘상대역’이라는 개념을 통해 양상문장을 번역하고, 해석하는 이론이다.

이제 이를 바탕으로 형이상학 체계로서의 상대역 이론을 살펴보자. 상대역 이론의 가장 특징적인 점은, 이 이론에서는 모든 가능 세계가 실제로 존재한다고 간주된다는 점이다. 즉, 다른 가능 세계에 존재하는 나의 상대역은 마치 내가 존재하는 것과 마찬가지로 존재한다는 것이다. 루이스 자신의 표현을 빌리자면, 그들은 “단적으로”(simpliciter)⁷⁾ 존재한다. 루이스는 이런 자신의 입장을 “양상실재론”(modal realism)이라고 명명한다.⁸⁾ 다른 철학자들이 가능 세계를 허구 내지 규약적인 것으로 여기고 있다는 것에 비하여 볼 때, 루이스의 입장은 극단적 실재론(extreme realism)이라고 할 수 있을 것이다.

그렇다면 이런 실재하는 가능 세계에 거주하는 자들은 어떤 존재들인가? 루이스에 따르면 “사물들이 그러했을 수도 있는 방식들”(way things could have been) 중에 하나로 이해된다.⁹⁾ 각 사물이 그러했을 수도 있는 방식들은 모두 각자의 가능 세계에 존재하는 것이다. 예를 들어, 안철수는 현실 세계에서는 대선에 출마했지만¹⁰⁾, 다른 어떤 가능 세계에 사는 안철수의 상대역은 대선에 출마하지 않은 채로 존재한다. 마찬가지로 안철수의 상대역이 힐링캠프에 출연하지 않은 가능 세계도 존재하고, 안철수의 상대역이 황금어장에 출연하지 않은 가능 세계도 존재한다. 즉, 우리는 “사물들이 그러했을 수도 있는 방식들”이라는 말에서 직관적으로 사물들이 그러하지 않음을, 즉 그런 방식으로 존재하지 않음을 추론해낼 수 있지만, 루이스

7) Lewis, D., *On the Plurality of Worlds*, Blakwell, 1986, p.3.

8) 그러나 이후에 루이스 자신도 인정하듯이 ‘양상실재론’이라는 이름은 호도적이다. 상대역 이론에서 실재하는 것은 양상이 아니라 가능 세계들이다.

9) Lewis, D., *Counterfactuals*, Blackwell, 1973, p.84.

10) 본고를 제출할 당시 안철수 씨는 아직 18대 대선 후보직을 사퇴하기 전이었다.

의 상대역 이론에 따르면 결코 그렇지 않다. “사물들이 그러했을 수도 있는 방식들”은 사물들이 실제로 그러한 방식과 마찬가지로 존재한다. 어떤 것의 상대역이란 다른 어떤 세계에서 실제로 존재하는 존재자이다. 이것이 “양상실재론”이란 말로 루이스가 말하고자 하는 바이다.

그런데 위에서 살펴본 것처럼 사물들이 현실에서 존재하는 것과 마찬가지로 존재한다면, 공준 P8에서 말하고자 하는 것은 무엇인가? 즉, ‘현실적이다’라는 말은 어떤 표현인가? 루이스는 P8을 통해 현실적인 것과 가상적인 것의 경계를 나누고 있는 것은 아닌가? 그렇지 않다. 루이스에 따르면 현실적이라는 것은 단지 ‘여기’ 혹은 ‘지금’과 같은 지표사일 뿐이다. 일상 언어에서 현실적인 것과 가상적인 것은 존재론적인 차이가 있다. 그러나 루이스에게는 이러한 존재론적 차이가 무의미하다. 모든 가능 세계의 모든 존재자들이 빼와 살을 가지고 존재하고 있기 때문이다.

‘현실적’이라는 용어에 있어서 한 가지 더 주의해야 할 점이 있다. 앞서 말했듯이 ‘현실적’이라는 말은 지표사라고 볼 수 있다. 따라서 ‘현실 세계’라는 용어를 사용할 때, 그것은 우리가 속한 현실 세계를 지칭하는 것이 아니라, 그 용어를 사용한 개체가 속한 세계를 지칭하는 것이다. 당연히 이 세계에 속한 우리는 ‘현실 세계’라는 용어로 우리의 현실 세계를 지칭하겠지만, 예컨대 다른 세계에 존재하는 나의 상대역이 사용하는 ‘현실 세계’가 내가 살고 있는 이 세계를 지칭하지는 않는다는 것이다.

이제 루이스의 상대역 이론에서 세계라는 것이 수 없이 존재하며, 또 우리와 꼭 같은 방식으로 존재한다는 점은 명확해졌을 것이다. 그런데 세계라는 것이 다수로서 존재하려면, 각 세계는 각각의 세계들에 대해 고립되어 있어야 한다. 그들은 결코 겹쳐지거나, 인과적 관계를 맺을 수 없다. 만약 그러한 일이 일어난다면, 그 두 세계는 두 세계가 아니라 하나의 세계라고 봐야한다. 루이스에 따르면, 세계라는 것은 그 세계 밖의 것들은 영향을 끼치지 못하고, 오직 그 세계 안의 개체들만이 영향을 주고 받을 수 있는 체계이다. 따라서 두 세계가 서로 영향을 끼치는 것, 즉 인과적 관계를 맺는다는 것은 세계

의 정의상 어긋난다. 따라서 각 세계는 고립적이다. 그에 따라 당연히 어떤 한 개체가 다른 두 세계에 겹쳐져 있는 경우도 불가능하다. 이는 공준 P2로부터도 따라 나온다.

3. 상대역 관계와 유사성

앞 장에서 우리는 루이스의 상대역 이론, 그리고 그를 바탕으로 한 양상실재론을 살펴봤다. 양상실재론에 따르면, 모든 가능 세계는 실제로 존재하며, 각 세계에 대해 고립적으로 존재한다. 이번에는 본 논문에서 가장 핵심적인 상대역 관계(counterpart relation)에 대한 루이스의 설명을 살펴보도록 하겠다. 상대역 관계에 대한 내용은 공준 P3-P6를 통해 알 수 있다. P3, P4 에 따르면, 상대역이란 어떤 세계 속에 속한 개체와 또 다른 어떤 세계 속에 속한 개체 간에 맺는 관계이다. 그리고 P5, P6 에 따르면, 상대역 관계는 같은 세계에서는 자기 자신과만 성립하는 관계이다. 이러한 상대역 관계가 성립되기 위한 필요조건은 ‘유사성’이다. 루이스는 오직 유사성만을 상대역 관계를 성립하게 하는 조건으로 여긴다.

여기서 한 가지 유의해야 할 상대역 관계의 특징이 있다. 한 가능 세계의 어떤 개체가 나와 상당한 정도로 유사하더라도, 그 세계에 그 개체보다 더 나와 유사한 무언가가 있다면 그 개체는 나의 상대역이 될 수 없다는 것이다. 즉, 나의 상대역이란 그가 속한 세계에서 나와 가장 유사한 개체이다. 하지만 그렇다고 어떤 세계에 나의 상대역이 반드시 존재해야 할 필요는 없다. 예컨대, 오직 풀뚜기들만 사는 세계에서 그나마 나와 가장 유사한 풀뚜기가 나의 상대역이 되어야 하는 것은 아니다. 그 세계의 사는 어떤 풀뚜기도 나와 충분할 정도로 유사하지 않기 때문이다. 단지 그 세계에 존재하는 개체 중에서 가장 닮았다는 것만으로는 상대역의 충분조건이 될 수 없다.

또 한 가지 중요한 점은, x 가 y 의 상대역인 것이 y 가 x 의 상대역임을 보장해주지 않는다는 것이다. y 의 입장에서는 그 세계의 x 가

자신과 가장 유사하더라도, x 의 입장에서는 y 의 세계에 y 보다도 더 자신과 유사한 개체가 있을 수 있다. 즉, ‘ Cxy ’가 ‘ Cyx ’에 따라 나오지 않으므로, 상대역 관계는 대칭적(symmetrical)이지 않다. 또한, x 와 y 의 유사성과 y 와 z 의 유사성이 x 와 z 의 유사성을 보장해주지도 않으므로 이행적(transitive)이지도 않다. 즉, ‘ Cxy ’와 ‘ Cyz ’로부터 ‘ Cxz ’를 도출할 수 없다.

상대역 관계의 이러한 성격이 중요한 이유는, 이로 인해 ‘통세계 동일성의 문제’(the problem of trans-world identity)를 피할 수 있기 때문이다. 앞 장에서 간단히 언급했듯이, 통세계 동일성의 문제는 QML 의미론의 큰 난점으로 지적되어 왔다. 루이스의 상대역 이론은 여타 이론들이 처한 이러한 난점을 쉽게 극복할 수 있다는 장점을 지닌다. 그럼 먼저 통세계 동일성 문제에 대해 살펴보자. 이는 흔히 ‘치솜의 역설’이라 불리는 치솜(R. chisolm)의 사고 실험을 통해 잘 알 수 있다. 치솜은 다음과 같이 말한다.¹¹⁾

우리의 현실세계를 $W1$ 이라고 했을 때, $W1$ 에 아담과 노아가 살고 있다고 하자. 당연히게도 아담과 노아는 서로 다른 속성을 가지고 있다. 예컨대, 아담은 930살까지 살았고 노아는 950살까지 살았다는 속성을 각각 가지고 있다. 이제 이들이 $W1$ 에서 아주 약간만 다른 속성을 가지고 살아가는 $W2$ 를 생각해보자. 나이로 다시 예를 들자면, $W2$ 에서 아담은 931살까지 살았고, 노아는 949살까지 살았다는, $W1$ 과 약간만 다른 속성을 가지고 있다. 이 경우에 속성이 아주 약간만 변화했기 때문에, $W1$ 의 아담과 $W2$ 의 아담이 동일한 인물임을 인정할 수 있다. 노아 역시 마찬가지이다. 그리고 또 다시 $W2$ 와 아주 약간만 다른 $W3$ 를 생각해보자. 여기서는 아담은 932살까지 살았고, 노아는 948살까지 살았다. 역시 $W2$ 와 $W3$ 는 아주 약간만 다르기 때문에 $W2$ 의 아담과 $W3$ 의 아담은 동일하다고 간주될 수 있다. 이런 식으로 쭉 진행하다보면, 결국 $W20$ 에서 아담은 950살까지 살고 노아는 930살까지 산 속성을 지니게 된다. 다른 속성들도 이와 마찬가지로 점차 변화해 왔다고 생각해보자. 그렇다면 $W1$ 의 아담은 $W20$ 의

11) Chisolm, R., “Identity through Possible Worlds”, in Loux, 1979,

노아와 완전히 같은 속성을 가지게 되고, W1의 노아 역시 W20의 아담과 완전히 같은 속성을 지니게 된다. 그러나 이는 동일성 관계에 대한 논의에서 널리 받아들여지는 ‘라이프니츠의 법칙’에 어긋난다. 라이프니츠의 법칙은 ‘동일자 식별불가능성의 법칙’과 ‘식별불가능자 동일성의 법칙’의 두 가지로 나뉘는데, 치즘의 역설에서 노아와 아담은 이 중 어느 것도 통과하지 못한다. W1과 W20의 아담은 동일자 이면서 전혀 다른 속성을 가지고 있다. 즉, 식별가능하다. 따라서 동일자 식별불가능성의 법칙을 통과하지 못한다. 또한 W1의 노아와 W20의 아담은 완전히 같은 속성을 가지고 있으면서도 다른 개체로 간주된다. 따라서 식별불가능자 동일성의 법칙을 통과하지 못한다. 결과적으로 치즘의 역설은 통세계 동일성이 모순에 빠지게 됨을 보여준다고 볼 수 있다.

다른 가능 세계 의미론들은 가능 개체가 가능 세계들에서 ‘동일’하기를 요구하기 때문에 이러한 역설이 발생할 수밖에 없다. 동일성 관계는 이행적이기 때문이다. 그러나 상대역 관계에서는 이러한 역설이 발생하지 않는다. 앞서 살펴봤듯이, 상대역 관계는 이행적이지 않다. W1의 노아는 W20는 고사하고 W10의 노아와도 상대역 관계가 성립하지 않을 공산이 크다. 각 세계의 노아들이 W1의 노아의 상대역이 되기 위해서는 모두 각각 W1의 노아와의 유사성을 검토해야 한다. 따라서 통세계 동일성의 문제는 전혀 발생하지 않는다.

그런데 지금까지의 설명에 따르면, 상대역이라는 것은 결국 그 개체와 같지 않다는 것을 함의하는 것 같다. 그럼 당연히 다음과 같은 의문이 제기될 수 있다. 우리에게 필요한 것은 이 개체의 가능한 상태나 변화에 관한 것이지, 이 개체의 상대역이 어떠한지에 관한 내용이 아니다. ‘원래 키가 179cm인 철수의 키가 180cm 일수도 있었다’라고 말할 때, 철수의 상대역이 키가 180cm인 것이 과연 철수랑 무슨 상관이 있다고 할 수 있는가? 이러한 의문은 크립키에 의해 제기되었다.

우리가 ‘험프리가 (이러이러한 일만 했더라면) 선거에 이길 수 있었다’라고 말할 경우, 우리는 험프리에게 일어날 수 있었던 일

에 대해 말하는 것이 아니라 누군가 다른 어떤 사람, 즉 ‘상대역’에게 일어날 수 있었던 일에 대해 말하는 것이 된다. 그러나 아마도 험프리는 아무리 자신과 닮았다고 하더라도 다른 어떤 사람이 다른 가능세계에서 승리를 했을 수 있는지에 대해선 관심을 덜 가질 것이다.¹²⁾

그러나 이러한 견해는 상대역 이론의 언어와 일상적인 양상 언어를 혼동한 것에 기인한 오해이다. 크립키는 문장 ‘험프리가 선거에서 이길 수 있었다’를 ‘어떤 가능세계에서 험프리의 상대역이 선거에서 이길 수 있었다’라고 번역했다. 그러나 이 문장은 ‘어떤 가능세계에서 험프리의 상대역이 선거에서 이긴 그런 험프리의 상대역이 존재한다’라고 번역되어야 하며, 이는 그런 험프리의 상대역이 존재할 때 참을 산출하는 양상문장이다. 다시 말해서, 이길 수 있었던 것은 험프리지 어떤 가능세계의 험프리의 상대역이 아니다. 어떤 가능세계의 험프리의 상대역은 단지 선거에서 이긴 상태로 존재함으로써, 우리의 현실세계에서 문장 ‘험프리가 선거에서 이길 수 있었다’의 진리 조건을 정해주는 역할을 할 뿐이다.

4. Feldman의 반론과 본질적 속성의 문제

지금까지 루이스의 상대역 이론을 개략적으로 살펴봤다. 루이스의 이론의 가장 큰 특징은 가능세계가 정말로 실재한다는 ‘양상실재론’을 주장한다는 점이다. 철학에서는 상식에 어긋나는 주장을 종종 볼 수 있지만, 이 양상실재론은 그 중에서도 가장 상식에 어긋나는 주장인 것 같다. 예컨대 보편자와 같은 추상적 존재가 존재한다는 것보다도, 모든 사물이 그럴 수 있었던 모든 분기점들이 다른 세계에 빼와 살을 가지고 존재한다는 것이 훨씬 더 비상식적이라고 판단된다. 루이스의 이러한 비상식적 이론은 정당화될 수 있는가? 루이스는 이 상대역 이론이 유용하기 때문에 정당화될 수 있다고 본다. 마치 수학

12) Kripke, *Naming and Necessity*, Harvard, 1972, p.45.

의 집합론이 유용하기 때문에 집합이라는 것이 존재한다고 가정되는 것처럼, 상대역 이론 역시 양상문장을 잘 번역할 수 있을 뿐 아니라, 다른 많은 철학적 퍼즐들을 풀어낼 수 있기 때문에 정당화될 수 있다는 것이다.

그렇다면 또 이러한 의문이 제기될 수 있다. 가령 상대역 이론의 유용성을 그대로 유지하면서 가능세계가 반드시 실재해야 하는 것이 아닌 이론을 구성할 수 있다면, 루이스의 상대역 이론보다 그 이론이 더 타당한 이론이라고 봐야하는 것 아닌가? 물론 그렇지만, 루이스에 따르면 그런 이론은 가능하지 않다. 루이스는 이와 같은 ‘대체적’(ersatz) 상대역 이론들이 가진 난점을 하나하나 지적한다.¹³⁾

상대역 이론에 대한 또 다른 반론은 펠드먼에 의해 제기되었다. 펠드먼의 요지는 다음과 같다. 상대역 이론에서 상대역 관계는 유사성에 의존하기 때문에, 그 개체가 질적으로 아주 달랐을 가능성에 대해 말할 수 없게 된다는 것이다.¹⁴⁾ 예를 들어,

(3) 철수는 실제의 철수와 전혀 닮지 않았을 수도 있었다.

와 같은 양상 문장을 생각해보자. 직관적으로 봤을 때, 이 문장은 충분히 참일 수 있다. 그럼 이 문장을 상대역 이론으로 번역해보자.

(3') 현실 세계의 철수와는 전혀 닮지 않은, 그런 철수의 상대역이 존재한다.

아마도 이러한 형태의 문장이 될 것이다. 그러나 이는 모순이다. 상대역이란 그 정의상 그 상대역 관계에 있는 개체와 충분한 정도로 닮았음을 전제하고 있기 때문이다. 좀 더 확실히 하기 위해서 이 문장을 다시 번역해보자.

13) 본 논문에서는 중요한 사항이 아니므로 이에 대한 내용은 생략한다. 이에 대한 설명은, Lewis, D., *On the Plurality of Worlds*, Blakwell, 1986 ch.3. pp.136-191. 또는 김영배, pp.87-108.

14) Feldman, F., "Counterparts", *Journal of Philosophy* 68. 1971, pp.407-408.

(3′) 현실 세계의 철수와는 전혀 닮지 않은, 다른 세계에 존재하는 철수와 충분한 정도로 닮은 그런 개체가 존재한다.

한 개체가 철수와 닮았으면서 동시에 닮지 않은 것은 불가능하다. 즉, 이 문장은 모순이다. 따라서 (3)은 상대역 이론을 통해서 적절히 번역될 수 없다.

이에 대한 루이스의 해결책은 상대역 관계를 맥락 의존적으로 설정하는 것이다.¹⁵⁾ 그에 따르면 상대역 관계는 여러 면에서 불확정적이다. 그는 세 가지 정도의 예시를 든다.

- Ⓐ 어떤 측면에서 유사성을 고려하는가
- Ⓑ 각 유사성의 측면들의 가중치를 어떻게 매기는가
- Ⓒ 어느 정도 이상의 유사성을 상대역이 성립할 정도의 유사성이라고 간주하는가

상대역 관계는 이런 사항들을 어떻게 고려하는가에 따라서 달라질 수 있다는 것이다. 따라서 (3′)은 다음과 같이 설명될 수 있다. 어떤 가능 개체가 철수의 상대역이면서 동시에 그와 전혀 닮지 않을 수 있는 것은, 상대역 관계를 고려할 때의 유사성의 측면과 다른 어떤 측면에서 전혀 닮지 않을 수 있다는 것이다. 즉, 어떤 측면에서는 매우 유사하기 때문에 상대역 관계가 성립할 수 있지만, 다른 어떤 측면에서는 전혀 상이할 수 있다는 것이다. 따라서 (3)은 상대역 이론을 통해서 적절히 번역될 수 있다. 루이스는 이러한 방식으로 펠드먼의 반론을 피해간다.

그러나 펠드먼의 반론을 피하려다보니, 상대역 관계는 매우 모호하고 맥락의존적인 개념이 될 수밖에 없었다. 이러한 상대역 관계 개념은 “양상언어의 진리조건을 부여하는데 있어 공허하거나 자의적이 될 공산이 크다.”¹⁶⁾ 이러한 상대역 개념의 모호한 성격은 그가 상대

15) Lewis, D., *Philosophical Papers, vol.1*, Oxford University Press, 1983, pp.42-43.

Lewis, D., *On the Plurality of Worlds*, Blackwell, 1986, pp.8-9.

16) 선우환, 「양화양상논리(QUANTIFIED MODAL LOGIC) 의미론과 루이스의

역 이론을 처음 제안했을 때에는 전혀 염두에 두지 않았던 것이다. 특히, 이러한 상대역 개념은 'Counterpart Theory and Quantified Modal Logic'에서 그가 내린 본질적 속성 정의에 다소간의 문제를 야기시키는 것 같다. 그의 정의는 다음과 같다.¹⁷⁾

1968년 본질적 속성 정의 : 어떤 것의 본질적 속성이란 그것의 모든 상대역들이 공유하는 속성이다.

또한 루이스는 본질적 속성과 본질을 구분하는데, 본질은 본질적 속성들의 교집합이라고 본다. 본질은 다음과 같이 정의된다.

1968년 본질 정의 : 어떤 것의 본질이란 그것이 그것의 상대역 모두와 그리고 그들과만 공유하고 있는 속성이다. 어떤 것의 상대역이란 그것의 본질인 속성을 가지고 있는 것 모두이다.
(이때 본질은 상대역의 본질이 아니다.)

이와 같이 그는 어떤 개체의 본질적 속성이 있어서 그 개체의 상대역들이 모두 그 속성을 가지고 있는 그런 속성을 본질적 속성이라고 파악하였다. 이는 직관적으로 받아들여질 수 있는 개념이다. 그러나 상대역 관계가 모호하고 맥락 의존적 개념이라면, 과연 이러한 단순한 정의가 유지될 수 있을지 의문이다. 앞에서 봤듯이, 수정된 상대역 관계는 어떤 면에서는 매우 유사하고, 또 어떤 면에서는 전혀 유사하지 않더라도 성립될 수 있다. 어떠한 측면을 고려하느냐에 따라서 다른 측면들의 유사성은 거의 무시될 수도 있다는 것이다. 가령 다음과 같은 상황을 생각해 볼 수 있다. 철수가 4개의 중요한 속성 a,b,c,d를 가지고 있다고 해보자. 그럼 다음과 같은 4명의 철수의 상

상대역이론(COUNTERPART THEORY) - 양상의미론의 두 대안과 또 다른 대안 -, 서울대학교 대학원, 1993, p.46.

17) Lewis, D., "Counterpart Theory and Quantified Modal Logic", *The Journal of Philosophy* 1968, p.122.

대역이 존재할 수 있다.

철수1=속성 {a,b,c} 를 가짐 철수2=속성 {a,b,d} 를 가짐
 철수3=속성 {a,c,d} 를 가짐 철수4=속성 {b,c,d} 를 가짐

각각의 철수의 상대역은 철수의 중요한 속성 중 한 가지씩을 결여하고 있지만, 다른 중요한 속성 세 가지를 가지고 있기 때문에 맥락에 따라 충분히 철수의 상대역이 될 수 있다. 이제 1968년 본질적 속성 정의를 여기에 대입해보면, 철수의 본질적 속성은 ‘없다’는 결론에 이르게 된다. 철수와 철수의 상대역들 모두가 공유하는 속성이 하나도 없기 때문이다. 문제는 본질과 상대역 관계가 상호 정의될 수 있기 때문에 더 복잡해진다. 앞에서 살펴봤듯이 철수와 철수의 상대역들이 모두 공유하는 속성, 즉 본질적 속성은 존재하지 않는다. 본질적 속성이 존재하지 않는다면, 당연히 본질 역시 존재하지 않는다. 따라서 루이스의 본질 정의에 따르면 철수 1,2,3,4는 철수의 상대역이 아니다. 철수의 본질적 속성이 존재하지 않으며, 따라서 철수의 본질 역시 없고, 결과적으로 철수 1,2,3,4는 본질을 가질 수 없기 때문이다. 즉, 철수 1,2,3,4는 철수의 상대역이면서 동시에 철수의 상대역이 아니다. 이는 모순이다.

그러나 그렇다고 해서 펠드먼에 대한 답변과 1968년 본질적 속성 정의가 양립 불가능하다고 볼 수는 없다. 상대역 관계가 맥락 의존적으로 성립한다고 하더라도, 어떠한 속성은 맥락과 관계없이 어떤 개체와 그 개체의 상대역들이 모두 공유할 가능성이 배제되지는 않는다. 가령, 앞의 철수의 예에서 a가 철수의 본질적 속성이라고 해보자. 그렇다면 루이스의 정의에 따라 그냥 철수 4는 철수의 상대역이 아니라고 하면 그만이다. a는 맥락과 관계없이 결여할 수 없는 속성이기 때문에, 철수 4는 철수와 몇몇 측면에서 꽤 유사함에도 불구하고 철수의 상대역일 수 없다고 볼 수 있다.

그러나 문제는 그런 속성이 어떤 속성인지 쉽게 알 수 없다는 것이다. 루이스는 ‘모든 사람은 본질적으로 사람이고, 형체를 가진 것은(coporeal) 본질적으로 형체를 가진다’¹⁸⁾는 것을 본질의 예로 든다.

그러나 뒷장에서 자세히 논의하겠지만, 필자는 사람인 존재가 ‘사람임’이라는 속성을 결여하더라도 그 존재의 상대역이 될 수 있다고 본다. 전통적으로 본질이라고 여겨지는 속성들, 예를 들어, 종적 속성, 기원, 내부 구조와 같은 속성들도 맥락에 따라 그것을 결여한 상대역이 존재할 수 있다는 것이다. 필자가 맞다면, 우리가 생각할 수 있는 거의 모든 속성은 그것을 결여한 상대역이 존재할 수 있다. 따라서 루이스가 제시하는 ‘사람임’이란 속성이나 ‘형체를 가짐’이란 속성이 과연 본질적 속성이 될 수 있을지 의문이다. 즉, 이러한 루이스의 본질적 속성 정의가 과연 어떠한 정보력이라도 가질 수 있을지 의문이다.

5. 본질적 속성에 대한 상대역 이론의 새로운 정식화

앞 장에서 필자는 루이스의 1968년 본질적 속성 정의는 어떠한 유의미한 정보도 주지 못한다고 주장했다. 만약 그렇다면, 이러한 정의가 계속 유지되기는 힘들 것이다. 그러나 필자는 상대역 이론이 본질적 속성에 대한 철학적 논의에 상당히 유용한 도구로 쓰일 수 있다고 본다. 이에 루이스와는 독립적으로, 상대역 이론을 바탕으로 한 새로운 본질적 속성 정의를 내려 보고자 한다. 이 새로운 본질적 속성 정의는 다음과 같이 정식화될 수 있다.

- ① 개체 a 의 속성 중 논리적으로 필연적인 속성을 제외한 어떠한 속성이라도 그것을 결여한 상대역이 존재한다.
- ② 개체 a 의 속성 F 를 결여한 a 의 상대역 c 가 속한 세계 w 와 a 가 속한 aw (actual world) 간에는 일정 정도의 차이가 존재한다.
- ③ 그 차이는 속성 F 가 무엇인지에 따라 정도가 달라진다.

18) Lewis, D., “Counterpart Theory and Quantified Modal Logic”, *The Journal of Philosophy*, vol 65, 1968, p.123.

새로운 본질적 속성 정의 : 속성 F는 개체 a에게 본질적이다 iff F를 속성으로 가지지 않는 a의 상대역 c가 속한 세계 w와 현실세계 aw간의 유사도가 낮다.

어떻게 해서 이러한 정의에 이를 수 있었는지 하나하나 찬찬히 살펴해보도록 하자. 먼저, ①은 어떻게 도출될 수 있는가? ①은 사실상 새로운 본질적 속성 정의의 가장 핵심적인 단계로, 루이스의 1968년 본질적 속성 정의와 정면으로 배치된다. 개체 a의 속성 중 논리적으로 필연적인 속성을 제외한 어떠한 속성이라도 그것을 결여한 a의 상대역이 있을 수 있다는 것은, 논리적으로 필연적인 속성을 제외하고는 개체 a와 그의 모든 상대역들이 공유하는 속성이 없다는 것을 의미한다. 어떤 측면에서 유사성이 고려되는지, 어떤 측면에 더 비중치를 매기는지 등에 따라 상대역 관계가 결정된다는 것은, 한 측면에서 유사성의 정도가 낮더라도(그 측면의 속성들을 공유하지 않더라도) 다른 측면들에서 충분한 유사성이 보장된다면 상대역 관계가 성립할 수 있음을 시사한다. 가령, 우리가 보통 자기 동일성을 유지하기 위한 가장 중요한 속성이라고 여기는 기원(origin)적 속성을 결여하더라도, 다른 질적 측면에서 충분한 유사성이 담보된다면 상대역 관계가 성립될 수 있다는 말이다. 좀 더 일반적으로 말해서, (논리적으로 필연적인 속성을 제외한) 어떠한 속성을 결여하더라도, 그것을 제외한 다른 속성들을 충분히 많이 공유하고 있다면 상대역 관계가 성립할 수 있다는 것이다. 한 가지 주의해야 할 점은 그러한 속성 중에 논리적으로 필연적인 속성은 결여할 수 없다는 것이다. 즉, ‘ $\sim(P \& \sim P)$ ’과 같은 속성은 어떤 개체도 그러한 속성을 결여할 수 없다. 이는 어떠한 가능 세계도 모순을 포함할 수 없다는 상대역 이론의 기본적인 전제에 의한 필연적인 귀결이다. 아마도 루이스 자신은 ①에 동의하지 않을지도 모르지만, 상대역 관계의 모호성은 이러한 정식을 충분히 가능하게 해줄 수 있을 것 같다.

②는 상대역 이론에 의한 당연한 귀결이다. 가령 다른 모든 측면이 동일하면서, 현실 세계에서는 철수의 키가 179cm이고 한 가능 세계에서는 철수의 키가 180cm인 두 세계를 생각해보자. 이 두 세계는

적어도 ‘철수의 키가 1cm 차이난다는 것’만큼의 차이가 존재한다.

이에 ③ 역시 어렵지 않게 추론될 수 있다. 위의 경우에 철수의 키 1cm 정도의 차이만 존재했지만, 철수의 어떤 속성들의 경우 그것보다 더 큰 정도의 차이가 요구될 수 있다. 다음의 예를 통해서 살펴보자. 현 미국 대통령인 오바마가 현실 세계에서 ‘흑인임’인 속성과 ‘단정한 헤어스타일을 가지고 있음’이란 속성을 가지고 있다고 해보자. 이제 다음과 같은 두 양상 문장에서 현실의 오바마와 각 가능세계의 오바마의 상대역 간에는 일정 정도의 차이가 존재한다.

- (4) 오바마는 아프로 헤어스타일을 할 수도 있었다.
- (5) 오바마는 흑인이 아닐 수도 있었다.

(4)의 경우, 오바마의 머리가 부풀어 오르는 것 이외에 다른 것들이 현실 세계와 다를 필요가 없다. 그러나 (5)의 경우는 조금 다르다. 흑인이 아닌 오바마의 상대역이 존재하기 위해서는, 그 개체가 속한 가능 세계와 현실 세계 간의 상당한 정도의 차이가 요구된다. 가령, 오바마가 흑인이 아니면서 그 외의 속성에서 충분한 정도의 유사성을 유지하려면, 현실 세계의 백인과 흑인의 상황이나 역사 등이 뒤바뀌는 것 등이 요구될 수 있다.¹⁹⁾ 이제 (4)의 세계와 (5)의 세계를 비교해보자. (4)의 세계와 현실세계는 상당한 정도의 유사성 관계가 성립하지만, (5)의 세계, 즉 오바마의 상대역이 흑인이 아닐 수 있는 세계는 현실세계와 상당히 다른 세계가 될 것이다. 즉, 개체의 어떤 속성들은 그것이 달라지더라도 그 상대역이 속한 세계에 별다른 변화를 요구하지 않지만, 다른 어떤 속성들은 그 상대역이 속한 세계에 상당한 정도의 변화를 요구한다.²⁰⁾ 따라서 우리는 위와 같은 본질적

19) 이는 오바마의 ‘소수 인종 출신의 미국 대통령’이라는 측면을 척도로 한 예라고 할 수 있다. 그 밖에도 다른 측면을 척도로 한 ‘흑인이 아닌 오바마’의 예는 얼마든지 가능할 수 있다.

20) 사실 속성이 세계에 변화를 요구한다는 표현은 옳지 않다. 현실 세계가 a 상태라고 하고 어떤 속성을 결여한 상대역이 존재하는 가능 세계를 b 상태라고 했을 때, 세계가 a 상태에서 b 상태로 ‘변화’하는 것이 아니라, a 상태인 현실 세계와는 다른 b 상태의 가능 세계가 ‘존재’하는 것이기 때문

속성 정의를 할 수 있는 것이다.

독자들의 이해를 돕기 위해 추가적인 설명이 필요할 것 같다. 아래의 그림을 살펴보자.



우리가 속해 있는 현실 세계(AW)를 정 가운데 두고, 현실 세계와 유사한 정도를 기준으로 가능 세계들을 배치시켜보자. 그렇다면 현실 세계와 근접해 있는 세계들은, 우리의 현실세계와 매우 유사한 세계들일 것이고, 멀리 떨어진 세계들은 현실세계와 거의 유사하지 않은 세계일 것이다. 가까이 있는 세계들은 물리적 법칙, 생물학적 법칙 등이 매우 유사할 것이고, 심지어 역사적 사실들 역시 매우 유사할 것이다. 그러나 충분히 멀리 있는 세계들은 우리와 물리적 법칙마저도 다를 수 있다.²¹⁾

그런데 이러한 본질적 속성 정의는 결국 본질이 정도의 문제(a matter of degree)라는 것을 함축하는 것 같다. F를 속성으로 가지지 않는 상대역이 속한 세계가 멀면 멀수록, 즉 그 세계와 현실세계의 유사성이 떨어질수록, 속성 F는 그 개체에게 더욱더 본질적이다. 어떤 개체에게 있어서 어떤 속성이 다른 속성보다 더 본질적이라거나, 혹은 덜 본질적이라는 식의 말이 가능한 것이다. 따라서 우리는 단순히 ‘~가 본질적이다’라고 말할 수 없다. ‘~가 본질적이다’라는 술어

이다. 그러나 직관적인 이해를 돕기 위해 ‘변화’라는 용어를 사용했다.

21) 이러한 도식은 필자가 새롭게 제시한 것이 아니라, 루이스가 직접 제시한 도식으로써, 그의 이론에 중요한 역할을 한다. Lewis, D., *On the Plurality of Worlds*, Blakwell, 1986, pp.20-27 참조.

로는 정도의 문제를 포착할 수 없기 때문이다. 따라서 새로운 본질적 속성 정의에서 전건은 수정될 필요가 있다. 가령,

새로운 본질적 속성 정의' : 속성 F와 G를 가진 개체 a에게 속성 F가 G보다 더 본질적이다 iff F를 속성으로 가지지 않는 a의 상대역 c1가 속한 세계 w1와 현실세계 aw간의 유사도가, G를 속성으로 가지지 않는 a의 상대역 c2가 속한 세계 w2와 현실세계 aw간의 유사도보다 낮다.

이와 같이 비교급을 사용해서 표현해야 할 것이다. 그러나 이러한 정의는 다소 복잡해서 그 의미가 제대로 전달되지 못할 위험이 있다. 따라서 필자는 “본질도”(degree of *essence*)라는 새로운 개념을 도입하고자 한다. 이 개념을 도입한 정의는 다음과 같다.

새로운 본질적 속성 정의' : 속성 F는 개체 a에게 있어서 본질도가 높다 iff F를 속성으로 가지지 않는 a의 상대역 c가 속한 세계 w와 현실세계 aw간의 유사도가 낮다.

이와 같이 “본질도”라는 표현을 통해, 보다 간단하게 본질의 정도 차 개념을 확실히 전달해 줄 수 있을 것이다.

6. 새로운 본질적 속성 정의의 장점

이제 이와 같은 우리의 새로운 본질적 속성 정의를 본질에 관한 기존의 논의에 적용시켜보자.²²⁾ 아리스토텔레스 식의 본질주의가 영향력을 잃은 이후, 어떠한 속성이 본질적인지 아닌지에 대한 합의는

22) 본질주의에 관한 논의는 A.C. 그렐링, 이윤일 역, 『철학적 논리학』 3rd Edition, 선학사, 2005, pp.121-133 의 논의를 중심으로 한다.

잘 이루어지지 않았다. 이에 크립키는 나름대로 설득력 있는 두 가지 방안을 제시한다. 그의 제안은 많은 이들에게 공감을 얻었지만 논쟁의 여지는 여전히 남아있다. 먼저 살펴볼 것은 한 개체의 기원이나 혹은 그 개체의 재료가 그 개별자에게 본질적이라는 견해이다. 크립키의 논증은 다음과 같다. 만약 한 책상이 나무로 되어 있다면, 나중에 그 책상을 이루는 구성성분이 나무가 아니더라도, 혹은 금과 같은 것으로 바뀌더라도, 그 책상은 본질적으로 나무로부터 유래한 것이라는 주장이다.²³⁾ 이에 대해 슬로트(Slote, M.)은 다음과 같은 반론을 제시한다.²⁴⁾ 앞서 크립키의 논증에서 책상의 재료로 쓰였던 나무가 사실은 어떤 마법과 같은 법칙으로 인해 금에서 나무로 변한 것이라고 해보자. 그렇다면 그 책상은 본질적으로 나무로부터 유래했다고 볼 수 없다. 고작해야 우연적으로만 나무로부터 유래했다고 할 수 있다.

크립키의 두번째 방안은 내부 구조에 관한 것이다. 크립키는 다음과 같이 말한다.

호랑이는 그 외양만을 가지고는 정의될 수 없다. 호랑이와 그 외양은 완전히 같으면서도 다른 내부 구조를 가진, 따라서 호랑이의 종이 아닌, 전혀 다른 종이 있을 수도 있다는 것이 가능하다.²⁵⁾

예를 들어, 코끼리가 그것들을 동물로 만드는 어떤 내적 구조를 가지는 동물이라면, 외양이 코끼리와 완전히 동일하더라도 코끼리와 같은 내적 구조를 가진 개체는 결코 코끼리일 수 없다는 것이다. 그런 존재가 있다면 그것은 코끼리가 아니라 전혀 새로운 종이라고 봐야한다.

이처럼 크립키와 일군의 학자들은 본질에 대해 나름대로 설득력 있는 두 가지 방안을 제시했지만, 이러한 방안에 대해서도 여전히 논란이 일고 있다. 이처럼 본질 개념은 “결코 명료한 개념이 아니며, 그렇다면 어떻게 본질이 인식될 것인지도 분명하지 않을 것이다.”²⁶⁾

23) Kripke, *Naming and Necessity*, Harvard, 1972, p.13.

24) Slote, M., *Metaphysics and Essence*, Oxford, 1974, p.8.

25) Kripke, *Naming and necessity*, Harvard, 1972, p.56.

이러한 상황에서 우리의 새로운 본질적 속성 정의는 어떻게 도움이 될 수 있을 것인가? 일단 기원에 관한 논의부터 살펴보자. 주의해서 살펴봐야 할 것은 슬로트의 반론이다. 슬로트는 어떤 마법과 같은 법칙으로 인해 금이 나무가 되는 것과 같은 상황을 상정하고 있다. 이는 논리적인 면만 따져 봤을 때 불가능한 일이 아니다. 위와 같은 상황은 상상가능하다. 그러나 우리의 현실 세계에서는 그런 일은 일어나지 않는다. 어떻게 금이 갑자기 나무가 되고, 은이 되고 그렇게 될 수 있겠는가. 이는 현실세계에서는 불가능하다. 물론 이러한 일이 일어나는 가능세계는 얼마든지 있을 수 있으며, 루이스에 따르면 실제로 존재한다. 그러나 이러한 가능세계들은 우리의 세계와는 매우 상이한 세계일 것이다. 즉, 이러한 가능세계들은 현실 세계와 매우 먼 세계들이다. 우리의 새로운 본질적 속성 정의에 따르면, 슬로트가 제시한 마법과 같은 반론은 오히려 기원이란 속성이 한 개체에게 얼마나 본질도가 높은가를 보여준다고 할 수 있다. 앞서 살펴본 “오바마가 흑인이 아닐 수도 있다”라는 문장도 마찬가지이다. 흑인이 아닌 오바마가 현실의 오바마의 상대역이 되기 위해서는 그 세계의 여러 가지 모습이 변해야만 한다.

내부 구조에 대한 논의도 마찬가지이다. 내부 구조가 전혀 다르면서 외양이 같은 어떤 것이 존재하는 것은 매우 힘든 일이다. 물론 논리적으로 불가능하지는 않지만, 현실 세계에서는 거의 불가능한 일이다. 예를 들어 어떤 로봇 코끼리를 상상해보자. 이 코끼리가 일반 코끼리의 모든 속성을 가지고 있는 것이 가능할까? 로봇 코끼리는 코를 손처럼 쓸 수는 있어도, 다른 코끼리와 사이에서 코끼리 새끼를 잉태하는 것은 불가능할 것이다. 필자가 말하고자 하는 바는, 내부 구조가 완전히 다르면서 속성은 완전히 같은 그런 것을 우리의 가까운 세계들에서 발견할 수 없을 것이라는 점이다. H2O와 모든 속성이 같으면서도 내부 구조는 전혀 다른 XYZ를 상상하는 것은 가능하지만, 적어도 우리의 현실 세계와 가까운 세계들에서는 그런 XYZ를 발견할 수 없을 것이다. 즉, 그런 일이 일어날 수 있는 가능 세계

26) A.C.그렐링, 이윤일 역, 『철학적 논리학』 3rd Edition, 선학사, 2005, p.133.

는 우리의 현실 세계와는 매우 먼 세계라고 할 수 있다.

이에 대해서는 추가적인 설명이 필요할 것 같다. 루이스는 “흠적 수반”(Humean supervenience)이라는 자신만의 형이상학 체계를 지니고 있다.²⁷⁾ 그에 따르면, 이 세상의 모든 것은 더 이상 분리될 수 없는 시공간 입자로 이루어져 있다. 따라서 궁극적으로 세상의 모든 것은 시공간 입자들의 배열로 설명될 수 있으며, 이는 모든 가능 세계들에서도 마찬가지이다. 그렇다면, 루이스에게 있어서 내부 구조란 결국 이 ‘시공간 입자들의 배열’에 다름 아니다. 그리고 이 ‘시공간 입자들의 배열’이 세상의 모든 것을 설명해준다. 즉, 어떤 개체의 속성들은 결국 시공간 입자들의 배열 상태로 설명될 수 있다. 그런데 이런 내부 구조가 전혀 다르면서 그 개체의 모든 속성을 공유하는 다른 개체가 존재하기는 힘들다.²⁸⁾ 굳이 “흠적 수반”까지 들먹이지 않더라도, 물리적 환원을 진지하게 생각한다면 내부 구조에 대한 논의는 어려워질 수밖에 없을 것이다. 우리가 일상적으로 경험할 수 있는 속성들이 모두다 더 근본적인 내적 상태로 환원될 수 있으며, 그것들은 다시 그보다 더 근본적인 내적 상태로, 결국엔 시공간 입자들의 배열 상태까지 환원될 수 있다. 그렇다면 내부 구조가 갖는 특징은 어쨌든 일상적으로 경험할 수 있는 특징으로 나타나게 마련이다. 따라서 우리는 우리 세계의 물리적 법칙에 따라, 내부 구조가 다르면 일상적 특징도 다를 것이라 기대할 수 있고, 또한 그 반대 역시 기대할 수 있을 것이다.

그러나 내부구조가 완전히 다르면서 일상적 수준에서는 완전히 같은 특징을 보이는 것이 원천적으로 불가능하다고 볼 수는 없다. 즉, 우리는 XYZ를 물로 인정하는 상황이 원천적으로 봉쇄되는 것은 아니다. 예컨대 우리는 고통이 다양한 물리적 상태로 구현될 수 있음을

27) 흠적 수반(Humean supervenience)에 관한 내용은 Ned Hall, David Lewis's Metaphysics, Stanford Encyclopedia of Philosophy (<http://plato.stanford.edu/entries/lewis-metaphysics>) 참조.

28) 불가능하다고 볼 수는 없다. 서로 다른 물리 상태에 의한 복수 실현은 가능하다고 볼 수 있기 때문이다. 예컨대, 고통과 같은 감각질이 서로 다른 물리 상태에 의해 실현될 수 있다는 것은 널리 받아들여지고 있다.

인정하지 않는가? 우리가 ‘물=H2O’는 받아들이면서, ‘고통=Cff’를 받아들이지 않는 것은 단지 우리와 우리의 가까운 세계에서 물의 상대역이 될 수 있는 것들이 모두 H2O뿐이기 때문이지, 물과 H2O 사이에 불가분의 필연성이 있어서는 아니다. 우리와 충분히 먼 세계에서는 우리의 H2O의 상대역이 XYZ인 그런 상황이 발생할 수 있을 것이다. 우리의 본질적 속성 정의는 이러한 내부 구조에 대한 논의를 간단하게 설명할 수 있다. 내부 구조란 우리와 우리의 가까운 세계, 그리고 꽤 먼 세계에서까지 공유되는 속성이다. 즉 그러한 점에 있어서 내부 구조란 속성은 본질도가 매우 높다고 할 수 있다. 그러나 내부 구조가 전혀 다르면서 모든 속성을 공유하는 그런 어떤 것을 생각할 수 있는 한에서, 그러한 존재가 존재하는 세계가 원천적으로 봉쇄되는 것은 아니다. 그러나 그러한 세계는 우리의 세계와 상당히 먼, 즉 상당히 이질적인 세계일 것이다.

지금까지 살펴본 것처럼 '기원'이나 '내부 구조'와 같은 속성들은 한 개체의 본질로서 진지하게 논의되었다. 그러나 그에 대한 반론 역시 상당히 설득력이 있어서, 이들을 개체의 본질로 받아들이기는 힘들어 보인다. 하지만 적어도 이러한 속성들이 '헤어스타일'과 같은 속성들보다는 훨씬 본질적으로 보인다는 것은 자명하다. 우리의 새로운 본질적 속성 정의는 이러한 현상을 잘 설명해 줄 수 있다. 새로운 정의에 따르면, 논리적으로 필연적인 속성을 제외한 어떠한 속성도 그것을 결여하는 것이 불가능하다고 볼 수는 없다. 그러나 적어도 우리가 직관적으로 어떠한 속성들은 다른 속성들보다 본질적이라고 느끼는 것 역시 사실이다. 이러한 우리의 직관은, '그러한 속성을 결여한 상대역이 속한 세계가 현실 세계와 상이하다'는 우리의 본질적 속성 정의를 통해 잘 설명될 수 있다. 즉, 전통적 의미의 본질은 아니라도, 각 속성이 그 개체에게 얼마나 본질적인지의 정도차는 말해줄 수 있다는 것이다.

7. 형이상학적 필연성에 관하여

이상의 논의가 함축하는 바는 다음과 같다. 논리적으로 필연적인 속성들을 제외하고는, 어떠한 속성도 그만의 특별한 지위를 갖지 못한다는 것이다. (논리적으로 필연적인 속성을 제외한) 모든 속성은 어디까지나 비교적 우위를 지닐 뿐이다. 이는 결국 크립키 식의 형이상학적 필연성을 부정하는 일일 것이다. 앞서 살펴봤듯이, 크립키는 ‘물=H₂O’에 어떤 형이상학적 필연성이 있다고 보았다. 즉, 그는

$$(6) (a=b) \rightarrow \Box(a=b)$$

라고 본 것이다. 그러나 필자는 그런 형이상학적 필연성을 인정하지 않는다. 우리의 논의에 있어서, 형이상학적 필연성이란 상당히 높은 필연성이지만 결코 논리적 필연성과 동등한 지위에 놓일 수는 없다. 즉, 형이상학적 필연성이라는 구분되는 필연성이 따로 존재하는 것이 아니라, 상당히 높은 정도의 필연성을 형이상학적 필연성이라고 부를 뿐이다.²⁹⁾

그러나 그렇다면 우리는 춘원과 이광수가 같다는 것이 필연적이지 않다고 말하는 것이 아닌가? ‘춘원’과 ‘이광수’가 지시하는 인물이 동일한 한에 있어서, 이 둘이 같다는 것이 필연적이지 않다는 것은 결국 대상 동일성이 필연적이지 않다는 것을 함축하지 않는가? 이러한 입장은 받아들여질 수 있는가? 상대역 이론, 적어도 수정된 상대역 이론에 따르면 그럴 수 있다. 다음과 같은 상황을 생각해보자. 어떤 가능 세계 c에 거주하는 두 개체 a와 b가 있는데, a는 소설 “무정”을 썼고, b는 현실 세계의 이광수의 부모님의 상대역인 개체들에 의해 태어났다. 그리고 a는 현실의 이광수의 부모님의 상대역인 개체들에 의해 태어나지 않았고, b는 소설 “무정”을 쓰지 않았다. 그리고 다른 부분들에 있어서 a와 b는 모두 이광수와 충분히 닮았다. 즉, 상

29) ‘본질’이란 개념이 ‘그 개체의 필연적 속성(들)’인 한에 있어서, 본질이 정도의 개념이라면 필연성 역시 정도의 개념이다.

대역이 되기에 충분할 정도의 속성을 공유하고 있다. 그렇다면 우리가 보통 ‘춘원’을 ‘소설 “무정”의 저자’라고 생각했을 때, 이 세계 c에서 춘원의 상대역은 a가 될 것이다. 또한 우리가 ‘이광수’를 ‘그 부모님의 아들’로 생각했을 때, 세계 c에서 이광수의 상대역은 b가 된다. 즉, 같은 세계에서 춘원과 이광수의 상대역이 각각 존재할 수 있다.

좀 더 정확한 예로 ‘지킬’과 ‘하이드’를 가지고 살펴보자. 역시 한 세계에 두 개체 a,b가 존재한다. 이 중 a는 오직 하이드로 변하지 않는다는 것만 빼고는 지킬 박사와 완전히 똑같다. 또한 b는 오직 지킬 박사로 변하지 않는다는 것만 빼고 하이드와 완전히 똑같다. 그렇다면, ‘지킬’의 맥락을 고려할 때에는 a를 상대역으로, ‘하이드’의 맥락을 고려할 때에는 b를 상대역으로 간주할 수 있을 것이다. 이제 a와 b가 같은 세계에 존재한다는 점에 있어서 크립키 식의 대상 동일성은 성립될 수 없다. (6)을 상대역 이론으로 번역하면 다음과 같다.

$$(6') (a=b) \rightarrow (x)(Wx \rightarrow (y)(z)((Iyx \& Ixz \& Cya \& Czb) \rightarrow (y=z)))$$

(6')는 '(a=b)라면, 모든 가능 세계에서 그 세계에 속한 두 개체가 각각 a와 b의 상대역이라고 했을 때, 그 두 개체는 동일하다'라는 의미이다. 따라서 (6')는 (a=b)일 때, 모든 세계에서 그 세계에 속한 a와 b의 상대역이 동일하다면 참이다. 그러나 위에서 살펴본 것처럼, ‘지킬’과 ‘하이드’가 대부분의 세계들에서 동일한 인물이라고 하더라도, 어떤 세계에서는 각각 다른 인물을 상대역으로 가질 수 있다. 즉, 모든 세계에서 ‘지킬’의 상대역과 ‘하이드’의 상대역이 같은 것은 아니다. 따라서 (6)은 상대역 이론에서 받아들일 수 없다.

그러나 필자는 이러한 대상 동일성이 상당히 높은 수준의 필연성을 가지는 것을 부정하지는 않는다. 예컨대 셋별과 개밥바라기는 같은 행성이지만, 그 둘이 다른 개체로 존재하는 세계는 가능하다. 그러나 그러한 세계는 물리적인 면에서 우리의 세계와 매우 상이한 세계일 것이다. 그렇다면 우리의 본질적 속성 정의를 응용하여, ‘셋별=개밥바라기’라는 것은 상당히 높은 정도의 필연성을 가지는 명제

임을 인정할 수 있다. 그러나 ‘셋별=개밥바라기’가 성립하지 않는 세계가 있는 한에서, 이 명제가 크립키가 주장하는 정도의 필연성을 갖는다고 인정할 수는 없을 것이다.

8. 나오며

지금까지의 논의를 요약해보면 다음과 같다. 필자는 가장 먼저 상대역 이론이 나오게 된 배경인 양화양상논리 체계와 QML 의미론을 간략하게 설명하고, 상대역 이론을 살펴보았다. 그러나 펠드먼이 상대역 이론에 결정적인 반론을 제기했고, 그에 따라 상대역 이론의 핵심 개념 중 하나인 상대역 관계가 매우 모호하고 불분명한 개념이 되고 말았다. 루이스는 대 레(*de re*) 표상을 위한 조건으로 유사성을 이용했는데, 펠드먼은 바로 이 유사성으로 대 레 표상을 하는 것의 난점을 지적한 것이다. 루이스는 이에 상대역 관계에 몇몇 조건을 추가함으로써 이러한 펠드먼의 반론을 피할 수 있었다. 그러나 이러한 조건들이 추가된 상대역 이론은 그가 1968년에 내린 본질적 속성 정의에 다소간의 문제를 발생시켰다. 결과적으로 필자는 이렇게 수정된 상대역 이론 하에서 그의 본질적 속성 정의는 그 자체만으로는 어떠한 유용한 정보도 주지 못한다고 보았다. 이에 필자는 기존의 본질적 속성 정의를 대신할 새로운 본질적 속성 정의를 내려 보고자 시도했다. 이러한 새로운 정의는, 어떠한 속성을 결여한 상대역이 속한 세계와 현실 세계 간의 상이성을 통하여 그 속성의 본질도를 평가하는 것이다. 이 새로운 본질적 속성 정의에 따르면, 본질이란 정도의 문제가 된다. 즉, ‘어떤 개체에게 어떤 속성이 본질적이다’라는 문장은, ‘어떤 개체에게 어떤 속성이 본질도가 높다’라는 식으로 번역되어야 한다. 그 다음으로 우리는 이 정의를 기존의 철학자들, 특히 크립키 등이 전개한 본질 논의에 맞추어 살펴보았다. 우리의 새로운 정의는, 논리적으로 필연적인 속성을 제외한 어떠한 속성도 그것을 결여할 수 있다는 것을 받아들인다. 따라서 기존에 본질적 속성이라고 주장

된 속성들 역시 결여될 수 있음을 보였다. 그러나 동시에 그러한 속성들의 본질도가 매우 높음을 밝힘으로써, 그러한 속성들이 다른 속성들보다는 훨씬 본질적으로 느껴지는 우리의 직관 역시 잘 설명될 수 있음을 보였다. 우리의 새로운 정의는 이와 같이 본질에 관한 우리의 상반된 직관을 동시에 설명할 수 있다는 데에 강점이 있다. 마지막으로 우리는 대상 동일성에 대한 문제를 살펴봤다. 현실세계에서 ‘a’와 ‘b’가 같은 개체를 지칭하더라도, 어떤 가능 세계에서는 각각 다른 개체를 지칭할 가능성이 있다. 그러한 점에 있어서 크립키 식의 대상 동일성은 상대역 이론 하에서는 주장될 수 없으며, 따라서 형이상학적 필연성이라는 특권적 지위 역시 주장될 수 없다.

크립키 등이 제시한 후험적으로 필연적인(necessary a posteriori) 본질 개념은 인간의 인식과 독립된 “자연종”(natural kind)을 객관적으로 탐구할 수 있는 단초를 보여주었다. 그러나 이 논문에서 제시한 본질 개념은, 자연종이 크립키가 주장하는 것처럼 형이상학적으로 필연적인 개념이 아님을 보여주는 것 같다. 자연종 개념은 모든 가능 세계에서 성립될 수 있는 개념은 아니다. 그러나 그것은 우리와 비슷한 법칙 체계를 가지고 있는 근접 세계에서는 충분히 유용할 수 있는 개념이다. 즉, 크립키 식의 본질주의를 받아들이지 않더라도 자연종은 탐구될 수 있다. 본 논문의 본질적 속성 정의는 이렇게 상반된 견해들을 수렴하고 종합할 수 있다는 점에서 그 의미가 있다 하겠다.

참고문헌

- 김영배, 「D.루이스의 양상실재론 옹호」, 서울대학교 대학원, 1994.
- A.C. 그렐링, 이윤일 역, 『철학적 논리학』 3rd Edition, 선학사, 2005.
- 선우환, 「양화양상논리(QUANTIFIED MODAL LOGIC) 의미론과 루이스의 상대역이론(COUNTERPART THEORY) - 양상의 미론의 두 대안과 또 다른 대안 -」, 서울대학교 대학원, 1993.
- 손병홍, 『가능세계의 철학』, 소피아, 2004.
- Chisholm, R., "Identity through Possible Worlds", in Loux, 1979.
- Feldman, F., "Counterparts", *Journal of Philosophy* 68. 1971.
- Ned Hall, David Lewis's Metaphysics, Stanford Encyclopedia of Philosophy(<http://plato.stanford.edu/entries/lewis-metaphysics>)
- P. Johnson, "Origin and Necessity", *Philosophical Studies* 32, 1977.
- Kripke, *Naming and Necessity*, Harvard, 1972.
- C. I. Lewis, *A Survey of Symbolic Logic*, California Univ. Press, 1918
- Lewis, D., "Counterpart Theory and Quantified Modal Logic", *The Journal of Philosophy*, vol 65, 1968.
- _____, *Counterfactuals*, Blackwell, 1973.
- _____, *Philosophical Papers, vol.1*, Oxford University Press, 1983.
- _____, *On the Plurality of Worlds*, Blackwell, 1986.
- C. McGinn, C., "On the Necessity of Origin", *Journal of philosophy*, 73, 1976.
- Slote, M., *Metaphysics and Essence*, Oxford, 1974.