논리연구 19-3(2016) pp. 341-368

무어의 역설과 자기-지시* **

권 홍 우

【국문요약】 “p이지만 나는 p라고 믿지 않는다”와 같은 문장은 어떤 상황에서도 자연스럽게 발화될 수 있을 것 같지 않고, 심지어는 부조리하고 모순되게 들리기까지 한다. 무어의 역설이 제기하는 문제는 왜 이런 문장이 아무런 형식적인 모순이 없음에도 불구하고 이를 발화하는 것이 부조리하게 들리는지를 설명하는 것이다. 무어의 역설에 대한 기존의 견해는 주로 믿음이나 주장(assertion)의 성격에서 그 부조리성의 근원을 찾으려 한다. 필자는 본 논문에서 기존의 견해들이 무어의 역설을 만족스럽게 설명하지 못함을 주장하고, 이에 대한 새로운 설명을 제안한다. 이 제안에 따르면 무어의 역설의 근원은 “자기-지시”에 있다. 자기-지시는 주체가 어떤 특정한 방식으로 믿음을 형성하는 성향에 의해 부분적으로 구성되는데, 무어의 역설은 주체가 자신을 “나”로 지시하는 동시에 어떤 사람을 “나”로 지시하기 위해 만족시켜야할 조건을 만족시키지 못하는 데에서 발생한다.

【주요어】 무어의 역설, 자기-지시

투고일: 2016.9.6 심사 및 수정 완료일: 2016.10.17 게재확정일: 2016.10.18

* 이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013S1A5B5A07048817).

** 이 논문을 꼼꼼히 읽고 유용한 지적을 해주신 익명의 심사위원들께 감사드린다.

1. 서론

무어(G. E. Moore)는 다음과 같은 문장에 특이한 점이 있음을 처음으로 인지하였다.[1] "국민의당이 이번 선거에서 선전하겠지만, 나는 그렇게(즉, 국민의당이 선전할 것이라고) 믿지 않는다." 이런 문장은 어떠한 상황에서도 자연스럽게 발화될 수 있을 것 같지 않고, 심지어 부조리하고(absurd) 모순되게(contradictory) 들리기까지 한다. 하지만 위의 문장은 아무런 형식적인 모순을 내포하지 않는다. **무어의 역설**(Moore's paradox)이 제기하는 중요한 문제는 이와 같은 문장이 아무런 형식적인 모순을 갖고 있지 않음에도 불구하고, 왜 이런 문장을 발화하는 것이 부조리하게 들리는가를 설명하는 문제이다.

무어의 역설이 갖는 중요성과 흥미에도 불구하고, 모든 사람이 동의할 만한 무어의 역설에 대한 설명은 아직도 나타나지 않았다고 보는 데에 이견이 있을 것 같지 않다. 필자가 본 논문에서 하고자 하는 바는 기존의 문헌에서 소홀하게 다루어져 왔던 무어의 역설의 한 가지 측면을 부각시키고, 이로부터 무어의 역설에 대한 새로운 제안을 해 보려는 데에 있다.

무어의 역설에서 부각되지 않았던 측면이란, 이 역설이 소위 "**자기-지시**"(self-reference) 현상과 밀접한 관련이 있다는 것이다.[2] 무어의 역설은 (적어도 전형적인 경우에는) "나"라는 단어를 포함하는 문장의 발화에서만 발생하며, 그런 점에서 자기-지시와 모종의 밀접한 연관이 있음이 틀림없다. 필자는 이점에 착안하여, "자기-지

---

1) Moore (1944/1983).

2) 뒤에서 명확히 하겠지만, 여기서 "자기-지시"란 단순히 자기 자신을 지시하는 것을 의미하지 않는다. 필자가 "자기-지시"라 부르는 현상은 언어철학 및 심리철학에서 소위 "지표성"(indexicality) 또는 "데 세"(de se)라는 표제 하에 다루어진 현상을 지칭한다.

시”가 바로 무어의 역설의 근원이라는 새로운 제안을 하고자 한다.[3)]

이하에서 필자의 계획은 다음과 같다. 우선 1절에서 무어의 역설과 그것이 제기하는 문제에 대해서 조금 더 자세히 서술하겠다. 2절에서는 기존에 영향력 있었던 몇 가지 견해를 개괄하고 이를 비판하겠다. 3절에서는 자기-지시가 무엇인지 설명하고, 그것이 무어의 역설과 어떤 관계가 있는지 고찰해 볼 것이다. 4절에서는 이 고찰을 바탕으로 무어의 역설을 자기-지시와 연결시킴으로서 무어의 역설에 대한 만족스러운 설명을 얻을 수 있다고 주장한다. 5절에서는 무어의 역설을 설명하기 위해서 제안한 가설을 믿을만한 독립적인 이유가 있다는 것을 논증한다.

2. 무어의 역설

우선 무어의 역설에 대해서 조금 더 정확히 기술하도록 하겠다. 무어의 역설을 일으키는 문장은 다음과 같은 두 가지 유형의 문장이다.

(1) p이지만, 나는 p를 믿지 않는다.
(2) p가 아니지만, 나는 p라고 믿는다.

이미 지적했듯이 이런 문장(앞으로 “무어-역설적 문장”이라 부르겠

3) 자기-지시 현상이 무어의 역설과 밀접한 관련이 있다는 것은 적어도 암묵적으로는 많은 철학자들에게 받아들여져 왔던 것 같다. 그럼에도 불구하고 필자가 아는 한에서, 자기-지시에서 무어의 역설에 대한 근원을 찾으려는 적극적인 시도는 있었던 것 같지 않다. 최근에 무어의 역설에서 자기-지시의 역할을 강조한 문헌으로 Chan (2010)을 볼 것. 하지만 챈은 자기-지시를 무어의 역설의 근원으로 적극적인 이론을 개진하지는 않는다.

다)을 말하는 것은 상당히 부적절하거나, 심지어는 부조리하고 모순되게 들리기까지 한다.

하지만 명백히 이런 문장은 그 내용에 있어서 아무런 형식적인 모순을 내포하고 있지 않다. 각각에 있어, 첫 번째 연언지(conjunct)는 (적어도 p가 나의 마음 상태에 대한 명제가 아닐 경우에) 나와 무관한 객관적 사실에 대한 것이다. (가령, p가 "국민의당이 이번 선거에서 선전할 것이다"라는 명제일 때, 이는 나와는 관련 없는 정치적인 사실에 대한 것이다.) 반면에 두 번째 연언지는 그와는 **완전히 독립적인** 나의 마음 상태(즉, 믿음)에 관한 문장이다. 이 둘 사이에 논리적 연결이 없음은 명백하다. 만일 (1)에 모순이 있다면, 나는 필연적으로 모든 참인 명제를 다 믿고 있어야 할 것이다. 이는 내가 인식적으로 한계가 있는 존재임을 부정하는 것에 해당하며, 물론 불합리하다. 만일 (2)에 모순이 있다면, 내가 가진 믿음은 필연적으로 모두 참이어야 할 것이다. 이는 내가 인식적으로 오류 가능한(fallible) 존재라는 것을 부정하는 것에 해당하며, 이 또한 물론 명백히 불합리하다.

애초에 무어가 이 현상을 지적할 당시에는 이 문제는 특정 종류의 문장(즉, 무어-역설적 문장)의 발화에서 생기는 **언어적인** 기현상으로 생각되었지만, 이후의 철학자들은 무어-역설적인 문장으로 표현되는 되는 바를 **믿거나**(believe) **판단**(judge)할 때도 똑같은 종류의 역설이 발생함을 깨닫게 되었다.[4] 다시 말해, 어떤 사람이 (1)과 (2)의 형식을 가진 문장을 발화하지 않고, 이것들로 적절히 표현될 만한 것을 마음속으로 믿거나 판단하는 경우에도 그 사람에게 심각한 문제가 있어 보인다는 것이다.

무어의 역설이 제기하는 문제는 (1), (2)와 같은 문장이 형식적인 모순을 내포하지 않는데도 불구하고, 왜 이를 주장하는(assert) 것이

4) 가령, Heal (1994), p. 6을 볼 것.

(그리고 이로 표현되는 바를 믿거나 판단하는 것이) 심각하게 잘못된 것으로 보이는지를 설명하는 문제이다. "무어의 역설"이라는 말은 이 역설적인 **현상** 자체를 가리키는 말로 주로 사용되지만, 필자는 이 **문제**를 일컫는 말로도 사용하도록 하겠다.

다음 절에서는 기존에 영향력 있었던 무어의 역설에 대한 몇 가지 해결책은 간단히 소개하고, 왜 이 이론들이 만족스럽지 않은지에 대해 설명하겠다. 이를 통해서 새로운 이론을 시도할 만한 좋은 이유가 있음이 드러나기 바란다.

3. 기존의 견해들

필자가 살펴보려는 첫 번째 견해는 무어 자신에 의해서 제시되었던 견해로, 화용론적(pragmatic) 견해라 불리는 견해이다.[5] 우선 다음과 같은 문장을 보자.

(3) 만일 **국민의당이 선거에서 선전하고 있지만 내가 그렇게 믿지 않는다**면, 나는 정치 감각이 떨어지는 사람일 것이다.

여기에서는 무어-역설적인 문장("국민의당이 선거에서 선전하고 있지만 내가 그렇게 믿지 않는다")이 **조건문 안에 포함되어 가정**되고 있고, 이런 경우에 전체 문장은 전혀 부적절하게 들리지 않는다. 즉, 무어의 역설은 (1), (2)와 같은 유형의 문장을 **주장**(assert)할 때에만 발생하는 문제인 것 같다. 이런 점에 착안해서, 화용론적 견해는 무어의 역설은 주장의 언화 행위(speech act)가 갖는 특별한 성질 때문에 나타나는 현상이라고 주장한다. 이 견해는 무어의 역

5) Moore (1944), 그 이후 많은 철학자들이 이에 동조하였다. 가령, Searle (1969), 보다 최근에는 DeRose (1991)을 볼 것.

설에 연루된 부조리성을 다음과 같이 설명한다. 첫 번째 유형의 무어-역설적인 문장의 경우, 발화자가 p라고 주장할 때, 발화자는 이미 자신이 p를 믿고 있다는 것을 "암묵적으로 표상"하고 있다. 하지만 두 번째 연언지에서, 이를 명시적으로 부정한다. 암묵적으로 표현된 바와 명시적으로 표현된 바가 서로 모순을 일으키며, 이런 점에서 "화용론적 모순"이 일어난다는 것이다.

하지만 이 견해에는 심각한 문제가 있다는 것이 잘 알려져 있다.[6] 만일 이 견해가 옳다면, 무어-역설적인 문장을 명시적으로 발화하고 주장할 때에만 문제가 나타나야 할 것 같다. 하지만, 앞서 보았듯이 무어의 역설은 단순히 언어적인 현상이라고 보기 힘들다. 어떤 사람이 무어-역설적인 문장을 발화하지 않고, 그것이 표현하는 바를 마음속으로 믿고만 있다고 해보자. 여전히 직관적으로 그런 사람의 믿음은 무언가 심각하게 잘못된 것 같다. 화용론적 견해는 언화 행위에 초점을 맞춤으로써, 왜 무어-역설적인 문장으로 표현되는 바를 믿거나 판단하는 것이 잘못되었는지를 설명할 수 없는 것으로 보인다.

두 번째 견해는 "나는 p라고 믿는다"와 같은 주장이 사실은 심적 상태, 즉 나의 믿음 상태에 대한 주장이 아니고, 단순히 "p이다"를 완곡하게 표현하는 방식이라는 아이디어에 기반을 둔다. 가령, 일상 맥락에서, "나는 국민의당이 선거에서 선전할 것이라고 **생각해 (또는 믿어)**"라는 말은 "국민의당이 선거에서 선전할거야"라는 말을 완곡하게 표현하기 위해서 사용될 수 있다. 이 견해에 따르면, 일반적으로 "나는 p를 믿는다"는 발화자의 심적 상태를 보고(report)하는 것이 아니라, "p이다"를 표현하는 다른 방식에 지나지 않으며, "p를 믿지 않는다" 역시 "p가 아니다"를 표현하는 다른 방식에 불과하다. 무어-역설적 문장 "p이지만, 나는 p를 믿지 않는다"

---

6) 가령, Heal (1994), Shoemaker (1995) 및 Moran (2001), 2장을 볼 것.

에서의 두 번째 연언지 또한 사실은 “p가 아니다”를 말하는 완곡한 방식에 지나지 않는다. 따라서 처음 보는 인상과는 달리, 그 의미를 따져보면 무어-역설적인 문장을 발화할 때 발화자는 “p이지만 p가 아니다”를 의미하는 셈이 되며, 따라서 무어-역설적인 문장의 부조리성은 다름 아닌 의미상의 모순에 불과하다는 것이다.[7]

하지만 이 견해 역시 심각한 결함이 있다. 물론 어떤 맥락에서 “나는 p를 믿는다”가 “p이다”에 대한 완곡한 표현으로 사용되는 것이 사실이다. 하지만 모든 경우에 그러한가? 예를 들어보자. 어떤 여론조사원이 나에게 “국민의당이 선거에서 선전할 것이라고 믿으십니까?”라고 묻는다. 이 맥락에서 여론조사원의 관심은 실제로 선거에서 누가 이길 것이냐가 아니고, 사람들의 의견, 즉 심리 상태인 것이 분명하다. 나 역시 이를 잘 인지하고 있다. 이런 경우에 내가, “아 나는 그렇게 믿고 있지 않습니다”라고 말한다면, 이는 나의 심리 상태에 대한 진술임이 분명한 것으로 보인다. 문제는 이런 경우에 조차도 뒤돌아서서 “아, 그렇지만, 국민의당이 선거에서 선전할거야”라고 말한다면, 나의 믿음 상태에는 무언가 심각한 문제가 있어 보인다는 것이다.[8]

일반적으로 말해서, “p이지만 나는 p라고 믿지 않는다”에서의 두 번째 연언지가 화자의 심리 상태에 대한 보고인 것이 명확한 경우가 있지만, 이런 경우에 조차 무어-역설적인 문장은 부조리하게 들리며, 위의 견해는 이를 설명하지 못하는 것으로 보인다.

세 번째로 살펴보고자 하는 견해는 무어의 역설에 나타나는 부조리성의 원천이 모종의 **비합리성**(irrationality)에 있다고 보는 입장이다. 합리적 주체의 이차 믿음(second-order belief), 즉 자신의 믿

---

7) 이런 입장을 취하는 대표적인 논문으로 Wright (1998)가 있다. 종종 비트겐슈타인 역시 이런 견해를 취한 것으로 여겨지지만 이런 해석이 맞는지는 논쟁거리이다.

8) 이런 방향의 비판에 대해서는 Moran (2001)과 Stalnaker (2000)를 볼 것.

음에 대한 믿음은 모종의 **규범**(norm)의 지배를 받는데, 무어-역설적 문장을 발화할 준비가 된 사람은 이 규범을 어기고 있으며, 이에 기인하는 비합리성(irrationality)에 무어의 역설의 근원이 있다고 보는 견해이다. 여기에 연루된 규범의 종류가 어떤 규범이냐에 따라 몇 가지 다른 견해가 있을 수 있다. 여기에서는 그 규범을 **인식적 규범**(epistemic norm)으로 보는 견해를 간략히 고려해 보도록 하겠다.[9)]

이 견해에 따르면, 어떤 사람이 p를 믿는다면, 인식적으로 합리적이기 위해서 그 사람은 반드시 "나는 p를 믿는다"라는 것 또한 믿어야 한다. 다시 말해, p를 받아들이면서, "나는 p를 믿는다"를 받아들이지 않는 것은 인식적으로 불합리하다는 것이다. "p이지만 나는 p를 믿지 않는다"고 주장하는 사람은, 그가 p를 받아들인다는 점에서, "나는 p를 믿는다" 또한 받아들여야 한다. 하지만 오히려 이의 부정을 받아들이고 있으니, 이 사람은 인식적 불합리성을 드러내고 있다는 것이다. 이 견해는 이러한 인식적 불합리성이 무어의 역설에 연루된 부조리성을 설명한다고 주장한다.

필자는 이런 견해가 무어의 역설에 이차 믿음이 연루됨을 분명히 인지한다는 점에서 앞에서 살펴본 두 견해에 비해 진일보한 이론으로 평가한다. 그럼에도 불구하고 필자는 이런 시도가 만족스럽지 못하다고 생각한다. 우선, p를 믿으면 "나는 p를 믿는다" 또한 믿어야 한다는 것이 왜 합리적인 사람이 따라야 하는 인식적 규범인지를 설명할 필요가 있다. 앞서 강조했듯이 p는 "나는 p를 믿는다"와 완전히 독립적이다. 그렇다면 왜 합리적 주체는 전자가 참이라고 판단했을 때, 후자가 참이라고 판단해야하는가?[10)] 여기서 애초에 무어의 역설이 제기하는 당혹스러운 물음이 다시 그대로 제기

9) 이런 입장을 지지하는 대표적인 문헌으로 Williams (2004), Fernandez (2013) 등등이 있다.

10) 이를 설명하기 위한 몇 가지 시도로 Williams (2004)와 Byrne (2005)를 볼 것.

되는 것 같다.

하지만 이런 인식적 규범이 있다는 것을 인정한다고 하더라도, 필자는 이런 설명이 무어의 역설을 만족스럽게 설명하는지에 대해서 의심을 갖는다. 핵심적인 이유는 다음과 같다. 일반적으로 규범은 (인식적 규범을 포함하여) 그 본성 상 지켜지지 않는 경우가 있기 마련이다. 하지만 규범이 지켜지지 않는 경우에 무어의 역설에서 나타나는 바와 같은 부조리성이 나타나는 선례를 찾아보기 힘들다. 가령, "지금까지 매일 아침에 해가 떠올랐다"로부터 "내일 아침에도 해가 떠오를 것이다"를 추론하는 것은 인식적 규범이라 할만하다. 어떤 사람이 (특별한 이유 없이) 이 규범을 따르지 않는다고 해보자. 이 사람은 "지금까지 매일 아침에 해가 떠올랐지만, 내일은 떠오르지 않을 것이다"라고 말할 위치에 놓이게 될지 모른다. 어떤 사람이 이런 발언을 하는 것은 상당히 괴상하게 들리기는 하겠지만, 부조리하게까지 들리지는 않는다. 간단히 말해, 인식적 비합리성은 무어의 역설에 연루된 부조리성에 대한 진단으로서는 지나치게 약하다는 것이다. 혹자는 무어의 역설에 연루된 규범은 상당히 특별한 종류의 규범이어서, 그것을 어기는 것만으로도 부조리성을 일으킬 수 있다고 주장할지도 모른다.[11] 하지만 이 규범이 그러한 특별한 지위를 가진다는 것을 보이는 것이 쉬운 일 같이 보이지 않는다. 이것을 보이지 않는 한, 이런 주장은 억측에 불과하다.

넷째로 살펴 볼 견해는, 무어의 역설은 우리가 가진 "믿음"의 개념의 **논리적** 성격에 기인한다는 견해이다. 이런 견해에 따르면, 다음과 같은 문장은 "믿음"의 개념에 의해서 **논리적 참**이다.[12]

(4) $B_x p \rightarrow B_x B_y p$ (x=y일 경우).

---

11) 가령, 그린과 윌리엄스는 다음과 같이 말한다. "극단적인(extreme) 이론적 합리성의 실패는 부조리할 수 있다" (Green and Williams 2007, 9).

12) 이런 입장을 취하는 대표적인 저서로 Hintikka (1962)가 있다.

이것이 옳다면, X가 p라고 믿는 동시에 자신이 p라고 믿지 않는다고 믿는 것은 논리적으로 불가능하며, 따라서 (믿음이 연언에 대해 닫혀 있다는 가정 하에) S가 “p이며, X는 p라고 믿지 않는다”라고 믿는 것 역시 논리적으로 불가능하다. 무어의 역설에 연루된 부조리성은 무어-역설적 문장으로 표현된 바를 믿는 것이 논리적으로 불가능하다는 것에 기인한다는 것이다.

이 견해의 가장 근본적인 문제는 과연 (4)를 논리적인 참으로 여기는 것이 합당한가이다. (4)가 논리적 참이라면, 그것이 거짓인 상황은 없어야 하지만, 그런 상황을 어렵지 않게 생각해 볼 수 있다. 가령, 대근이는 그가 평소에 여성을 대하는 방식을 보았을 때 남존여비 사상을 믿고 있다고 생각하는 것이 합당하다고 해보자. (즉, (4)의 전건을 만족시킨다.) 그럼에도 불구하고 대근이 자신은 (이런 믿음을 가진 사람이 흔히 그렇듯이) 자신이 그러한 사람이라는 것을 인정하지 않으려고 하며, “너는 남존여비 사상이 옳다고 믿느냐?”라고 물었을 때, 그렇지 않다고 대답할지 모른다. (즉, 위의 조건문에 후건은 거짓이다.)

이런 식의 반론이 결정적인지에 대해서는 논란의 여지가 있다. 어떤 철학자들은 위의 대근이의 사례와 같은 경우는 그 사람의 믿음이 “파편화”(fragmented)된 것으로 보는 것이 합당하며, (4)의 반례가 되지 못한다고 주장한다.[13] 한 가지 견해에 따르면, 대근이의 남존여비 사상에 대한 믿음은 암묵적(implicit) 믿음이며, 그가 자기 자신에 대해서 갖고 있는 명시적(explicit) 믿음과는 구분된다. (4)는 “믿음”의 의미를 단일하게 해석할 때만 성립한다는 것이다.

필자는 이런 대응을 논박하기 보다는, 잘 알려지지 않은 다른 한 가지 사례를 통해 (4)가 논리적 참이 아니라는 것을 보이겠다. 도

13) 이런 견해에 대한 논의로는 Greco (2015)를 볼 것.

균이는 갑자기 내리기 시작한 비를 피해서 잠시 백화점에 들어간다. 에스컬레이터를 타고 돌아다니는 중에 천장 거울에 비춰진 남루한 옷차림의 한 사람을 발견한다. 도균이는 그 남루한 모습의 사람이 자기 자신일 것이라고는 미처 생각하지 못한다. 그는 거울에 비춰진 사람의 옷차림이 비 오는 날씨에 걸맞지 않는다고 생각하여, "저 사람은 비가 온다고 믿지 않나보다"라고 생각하게 된다. 이 경우 도균이는 비가 온다고 믿는 동시에 (즉, 위 (4)의 전건을 만족시키는 동시에), 거울에 비춰진 그 사람은 비가 온다는 것을 믿지 않는다고 믿게 되는 것이다 (즉, (4)의 후건을 만족시키지 않는다). 도균=거울에 비춰진 그 사람임에도 불구하고 말이다! 이 경우 하나의 믿음 명시적 믿음이고, 다른 믿음은 암묵적 믿음이라고 피해가는 것은 옳지 않아 보인다. 따라서 이는 (4)에 대한 명백한 반례가 된다. (이런 종류의 예는 다음 절에서 제시할 필자의 견해에서 중요한 역할을 하게 될 것이다.)

이상에서 무어의 역설에 대해서 영향력 있었던 몇 가지 견해를 소개하고, 각 견해에 심각한 문제가 있음을 보았다. 물론 각 견해에 대한 이러한 문제점이 과연 그 견해의 핵심 통찰을 폐기해야 할 정도로 치명적인지에 대해서는 이론의 여지가 있을 수 있다. 하지만 이상의 논의가 기존의 견해와는 다른 새로운 이론을 추구해 볼 만한 한 가지 이유를 제공한다는 것을 부인하게 어렵다.

4. 자기-지시와 무어의 역설

필자는 무어의 역설에 대한 만족스러운 설명을 얻기 위해서는 기존의 견해들이 크게 관심을 두지 않았던 무어의 역설의 한 가지 측면으로 관심을 돌릴 필요가 있다고 생각한다. "**자기-지시**" (self-reference) 현상이 그것이다.

자기-지시의 문제는 1970년대 이래에 하나의 독립적인 연구 주제로 여겨져 왔는데, 이는 존 페리(John Perry)와 데이빗 루이스(David Lewis)의 저작에 힘입은 바 크다.[14] 직관적으로, 자기-지시란 물론 언어와 생각에서 자기 자신을 지시하는 것을 의미한다. 그러나 자기-지시는 단순히 **자기 자신을 지시함**과는 구별되어야 한다. ("자기-자신"이란 표현에서의 줄표("-")는 이 점을 부각시키기 위한 것이다.) 가령, 철수가 가벼운 뇌진탕으로 잠시 기억상실증에 걸려 자신이 철수라는 사실을 잊어버렸다고 하자. 철수는 오늘 아침 신문에서 "철수, 당 대표에 선출"이라는 기사 제목을 보고, "아, 철수라는 사람이 당 대표에 선출되었구나"라고 말한다. 여기서의 "철수"는 철수 자신을 지시하는 것이 틀림없다. 그럼에도 불구하고, 철수가 그 자신이 철수임을 인지하지 못하고 있다면, "철수"란 표현은 자기-지시를 위해 사용된 것이 아니다. 또 다른 사례를 보자. 도균이가 백화점 천장에 설치된 큰 거울에 비추어진 어떤 사람의 뒷모습을 보게 되었다고 하자. 도균이는 그 사람이 자기 자신임을 깨닫지 못한 채, "저 사람은 탈모 초기임에 틀림없다"라고 말한다. 이 경우 역시 도균이는 어떤 의미에서 자기 자신을 지시하고 있지만, 거울 속에 비친 사람이 자기 자신임을 인지하지 못한다는 점에서 "저 사람"이라는 표현을 자기-지시를 위해 사용하고 있는 것이 아니다.

자기-지시는 한국어에서 일반적으로 "나" 및 이와 동족어(cognate)로 행해진다. 하지만 반드시 이런 표현을 통해서만 자기-지시가 이루어지는 것은 아니라는 점에 주목할 필요가 있다. 가령, 재용이는 "나"를 사용해서 할 만한 말들을 "재용"이라는 고유명사로 표현하는 기이한 언어 습관을 가지고 있다고 하자. (가령, 재용이는 배가 고플 때, "엄마, 나 밥 좀 줘!"라고 말하는 대신에 "엄

14) Perry (1977, 1979)와 Lewis (1979).

마, 재용이 밥 좀 줘!"라고 말한다.) 게다가 주변 사람들 역시 재용이의 이런 언어적 습관을 잘 인지하고 있을 수 있다. 이런 경우에는 재용이는 "재용"이라는 이름을 자기-지시를 위해 사용하고 있다고 보는 것이 합당하다.

자기-지시 현상이 제기하는 철학적 문제는 물론 자기-지시가 정확히 무엇인지를 규명하는 문제이다. 직관적으로 말해서 어떤 언어적 표현을 자기-지시를 위해 사용한다는 것은, 그 표현이 지칭하는 사람을 **자기 자신으로 여긴다**는 것이다. 따라서 자기-지시가 제기하는 문제는, 어떤 사람을 자기 자신으로 여긴다는 것이 무엇인지에 답하는 문제이기도 하다. 프레게주의자(Fregeans)들은 자기-지시는 어떤 사람을 특별한 일인칭적 "제시 양식"(mode of presentation) 하에서 지시하는 것에 다름 아니라고 주장할 유혹을 느낄 것이다. 하지만 자기-지시의 문제는 기존의 프레게적인 틀 내에서 쉽게 다루어질 수 없음이 잘 알려져 있다.[15] 자기-지시의 본성이 무엇인지에 대한 문제는 여전히 활발히 논의되고 있는 문제이다.

이제 다시 무어의 역설로 돌아와 보자. 흥미로운 것은 무어의 역설은 자기-지시를 위해 사용되는 표현을 포함할 경우, 오직 그런 경우에만 발생한다는 점이다. 가령, 미경이가 철수에게 다음과 같이 말한다고 해보자.

(5) 국민의당이 선거에서 참패할 것이지만, **당신은** 그렇게 믿지 않는다.

---

15) 프레게 자신은 자기-지시의 문제를 해결하기 위해서 "각자의 사람이 자기 자신에게는, 다른 사람에게 제시되지 않는, 특별하고 원초적인 방식으로 제시된다(presented)"(Frege, 1919/1997)고 주장한다. 하지만 페리(1977)가 성공적으로 보였듯이, 적어도 프레게의 "의미"(Sinn) 또는 "제시 방식"(mode of presentation)을 표준적인 방식으로 (즉 기술적(descriptive) 의미로) 해석했을 때, 프레게의 발언을 일관되게 해석할 방법은 없어 보인다.

미경이의 말은 철수의 정치 감각이 좋지 않음을 지적하는 것일 수 있으며, 물론 이 말에는 아무런 부조리함이 없다. 이제 정치 문제에서 뿐만 아니라 심리적인 문제에서도 미경이의 조언이 틀린 적이 없다고 굳게 믿는 철수는 이 말을 듣고 미경이의 말을 전적으로 신뢰하기로 한다. 철수는 미경이가 (5)를 말하는 것을 듣고 다음과 같이 말할 위치에 놓이게 될 것이다.

(6) 국민의당이 선거에서 참패할 것이지만, **나는** 그렇게 믿지 않는다.

철수는 미경이의 (5)를 듣고 (6)을 믿게 되었으며, 그런 의미에서 (5)와 (6)는 정확히 같은 내용을 갖는다고 할만하다. 그렇지만 (5)에는 아무 문제가 없는 반면에 (6)은 심각한 문제가 있어 보인다. 즉 "나"라는 표현이 연루된 것이 결정적인 차이를 만드는 것이 분명하다.

여기까지는 비교적 잘 알려진 사실이다. 하지만 무어의 역설과 자기-지시가 본질적으로 연결되어 있다는 것을 뒷받침하는 조금 더 강력한 증거가 있다. 앞에서 보았던 백화점 거울에 비친 자신의 모습을 보고 있는 도균이의 경우를 생각해 보자. 도균이가 거울 속의 사람이 자기 자신이라는 것을 인지하고 있지 못한다면, 다음과 같이 말하는 것이 문제가 있어 보이지 않는다.

(7) 비가 오지만, **저 사람**은 그렇게 믿지 않는다.

이는 "저 사람"이 도균이 자신을 지시함에도 불구하고, 그것이 자기-지시를 위해 사용되지 않고 있기 때문이다. 이는 자기-지시가 무

어의 역설을 일으키는 필요조건임을 강력히 시사한다. 게다가 필자는 자기-지시가 무어의 역설을 일으키는 충분조건이기도 하다고 주장한다. 다음과 같은 예가 이를 보여줄 것이다. 습관적으로 자신의 이름을 자기-지시를 위해 사용하는 재용이가 다음과 같이 말한다고 하자.

(8) 엄마는 외출 중이지만, 재용이는 그렇게 믿지 않는다.

재용이가 "재용"이라는 이름을 자기-지시를 위해 사용한다는 것을 아는 사람들에게는, 이는 무어-역설적 문장을 주장하는 것만큼이나 모순적으로 들릴 것이다.

이런 관찰을 요약해서, 다음과 같이 일반화를 하는 것이 적절할 것 같다.

어떤 지시어 τ가 자기-지시를 위해 사용되는 경우, 오직 그 경우에만 "*p*이지만 τ는 *p*라고 믿지 않는다" (또는 "*p*가 아니지만 τ는 *p*라고 믿는다")라고 주장하는 것이 부조리하다.

무어의 역설에 대한 어떤 이론도 이 일반화가 왜 참인지를 설명할 수 있어야 할 것이다. 필자는 여기서 한 걸음 더 나아가, 자기-지시야 말로 무어의 역설에 연루된 부조리성의 근원이라고 주장하고자 한다.

4. 새로운 이론

필자가 제안하고자 하는 무어의 역설에 대한 새로운 가설은 무어의 역설은 자기-지시의 특수성에 의해서 발생하는 현상이라고 주

장한다. 핵심적인 가설을 다음과 같이 표현하겠다.

<핵심 가설> 어떤 언어 표현 τ를 자기-지시를 위해 사용하는 사람은 반드시(necessarily) 명제 p를 "τ는 p를 믿는다"의 참, 거짓과 직접적으로(directly) 대등한(equivalent) 것으로 여긴다.[16]

여기에 포함된 몇 가지 개념을 설명할 필요가 있겠다. 어떤 사람이 명제 ϕ를 명제 ψ에 **대등한 것으로 여긴다**는 것은 다음과 같은 성향(disposition)이 있음을 말하는 것이다: 그 사람이 ϕ를 긍정하면 ψ 역시 긍정하는 성향이 있으며, 반대로 ϕ를 부정하면 ψ 또한 부정하는 성향이 있다. 적절한 배경 믿음이 주어지면, 합리적인 사람은 임의의 명제 ϕ를 임의의 명제 ψ와 대등한 것으로 여길 수 있다. (가령, "ϕ↔ψ"가 참이라고 믿는 사람은 당연히 위의 의미에서 ϕ와 ψ를 대등한 것으로 여길 것이다.) 어떤 사람이 ϕ와 ψ을 "**직접적으로**"(directly) 대등한 것으로 여긴다는 것은, 이런 배경지식에 상관없이 ϕ와 ψ를 대등한 것으로 여기는 성향이 있다는 것을 의미한다.

이 가설이 말하는 바를 예를 들어 설명해 보겠다. 방금 전에 내가 "재용"이라 불리는 인물을 소개 받았다고 하자. 재용이와 대화를 이어나가는 가운데, 재용이는 계속 "재용"이라는 이름으로 누군가를 지칭하며 그에 대한 이야기를 들려주고 있다. 이야기를 듣는 가운데 나는 두 가지 가능성을 품게 된다. 한 가지 가능성은 재용이가 아마도 그와 동명이인인 누군가에 대해서 이야기 하고 있을 가능성이다. 다른 가능성은 재용이가 자기-지시를 위해서 "나"라는 표현 대신에 자신의 이름을 사용하는 기이한 습관을 가지고 있다는

16) 잘 알려져 있지 않지만, 필자는 이러한 주장이 가렛 에반스의 유명한 저서 *The Varieties of Reference* (Evans, 1982)에서 이미 제시되었다고 본다. 하지만 문헌 해석에 대한 논쟁을 피하기 위해 이에 대한 언급을 자제하겠다.

것이다. 어떤 것이 맞는지 어떻게 테스트할 수 있을까? (물론 가장 쉬운 방법은 "네가 말하고 있는 재용이가 너 자신이야?"라고 묻는 것이다. 하지만 무례함을 피하기 위해 다른 방법을 찾고 있다고 하자.) <핵심 가설>이 옳다면, 한 가지 결정적인 테스트가 있을 수 있다. 다음과 같은 대화를 고려해 보자.

나: (창밖을 보면서) 오랜만에 비가 오네!
재용: 어. 진짜.
나: 재용이도 여기에 비가 온다는 걸 믿을까?
재용: 글쎄, 잘 모르겠는데 …….

대화가 이 지점에 이르렀을 때, 나는 재용이가 "재용"이라는 이름을 자기-지시를 위해 사용하는 것이 아니라는 결정적인 힌트를 얻게 되는 것 같다. 왜 그런가? 만일 재용이가 "재용"이라는 이름을 자기-지시로 사용한다면, 물론 그는 "아, 당연하지, 지금 보고 있는 걸!"이라고 대답했을 것이다. <핵심 가설>은 이와 같은 직관을 정식화한 것이다. 위의 대화에서 재용이는 "여기에 비가 온다"와 "재용이는 여기에 비가 온다는 것을 믿는다"를 대등한 것으로 여기지 않으며, 이것은 곧 재용이가 "재용"이로 지칭하는 사람을 자기 자신으로 여기지 않는다는 것을 함축한다는 것이다.

앞 단락에서 말했던 것과 같은 이유로, 필자는 <핵심 가설>이 직관적으로 상당히 그럴 듯하다고 생각한다. 필자는 이 가설을 받아들일 만한 다른 이론적 이유가 있다고 생각하지만, 이에 대한 논의는 다음 절로 미루겠다. 여기서는 이 가설이 어떻게 무어의 역설을 설명하는지 먼저 보이겠다.

우선 다음과 같이 비교적 논란의 여지가 없는 가설 하나를 덧붙이겠다.

<보조 가설> 한국어에 능통한 화자는 "나"라는 표현을 항상 자기-지시를 위한 표현으로 사용한다.

다시 말해, 한국어에 능통한 사람이 어떤 사람을 "나"라고 지칭하면서, 그 사람을 자기 자신으로 여기지 않는 경우는 없다는 것이다. 왜 이 가설이 참인지에 대해서는 이론의 여지가 있을 수 있다. 어떤 철학자들은 "나"라는 말의 언어적(linguistic) 의미에 의해 참이라고 주장할 것이며, 다른 철학자들은 비언어적인 모종의 규약에 의해 참이라고 볼지 모르겠다.[17] 그러나 그 이유가 어떻든 간에 <보조 가설>이 참이라는 것에 대해서는 논란의 여지가 없는 것으로 보아도 무방할 것 같다.

이제 철수가 "국민의당이 이번 선거에서 선전할 것이지만, 나는 그렇게 믿지 않는다"고 말했다고 해보자. 철수는 "국민의당이 선전할 것이다"라는 명제를 "나는 국민의 당이 선전할 것이라고 믿는다"의 참, 거짓과 대등한 것으로 여기지 않고 있는 셈이다. **<핵심 가설>에 따르면**, 이는 철수가 "나"라는 표현을 자기-지시를 위해 사용하고 있지 않음을 의미한다. 하지만 **<보조 가설>에 따라**, 철수가 한국어에 능통한 화자라는 가정 하에 "나"라는 표현을 자기-지시를 위해 사용하고 있는 것이 틀림없다. 그렇다면, 철수는 "나"라는 표현을 자기-지시를 위해 사용하는 동시에, 자기-지시를 위해 사용하지 않고 있는 셈이 된다. 철수가 "나"라는 표현을 사용하는 방식에 명백한 "비일관성"이 존재한다. 필자는 이 비일관성이 바로 무어의 역설에 연루된 부조리성의 근원이라고 주장한다.

언어적 현상으로서의 무어의 역설에 대한 이와 같은 설명이 주

17) 데이빗 카플란(David Kaplan)은 그의 기념비적인 논문 "Demonstratives" (1989)에서 "나"의 언어적 의미가 그것의 "캐릭터"(character)에 의해서 포착된다고 주장한다. 만약 이것이 옳다면, <보조 가설>이 "나"의 언어적 의미에 의해서 참이라는 주장은 상당히 위태로워질 것이다.

어지면, 이를 믿음(또는 판단) 차원의 무어의 역설로 확장하는 것은 어려운 일이 아니다. 자기-지시는 언어적인 차원에서 뿐만 아니라 생각의 차원에서도 일어난다. 철수가 "국민의당이 이번 선거에서 선전할 것이지만, 나는 그렇게 믿지 않는다"라고 적절하게 표현될 바를 마음속으로만 믿는다고 하자. 그는 자기 자신에 대해서 생각하는 것일까? 물론 그렇다. "나"를 사용해 적절히 표현될 수 있는 믿음이므로. 물론 그렇지 않다. 철수의 믿음은 자기-지시를 위해서 만족시켜야할 조건을 만족시키지 못하므로. 여기서도 철수의 마음 상태에 모종의 비일관성이 있음이 드러난다.

앞에서 무어의 역설이 "나"를 포함하지 않는 문장의 발화에서 나타날 수 있음을 보았다. 가령, 재용이의 "엄마는 외출 중이지만, 재용이는 그렇게 믿지 않는다"는 적어도 재용이가 그 이름을 자기-지시를 위해 사용한다는 것을 인지한다는 사람들에게는 부조리하게 들린다. 재용이의 발화는 "나"를 포함하지 않으며 이런 점에서 위에서처럼 <보조 가설>에 의존할 수 없다. 하지만 재용이의 주장이 그가 "재용"이라는 이름을 자기-지시를 위해 사용한다는 것을 전제하고 있는 사람들에게만 모순적으로 들림에 유의할 필요가 있다. <핵심 가설>에 따르면, 재용이는 "재용"이라는 이름을 자기-지시를 위해 사용하고 있지 않는 셈이 되므로, 이는 사람들이 이미 전제하고 있었던바(즉, 재용이가 "재용"이라는 이름을 자기-지시로 사용한다는 것)와 직접적으로 모순된다.

무어의 역설에 대한 이러한 설명의 큰 장점 중에 하나는, 이 견해가 무어의 역설에 연루된 특별한 종류의 부조리성을 잘 설명해 주는 것 같다는 점이다. 많은 철학자들이 무어-역설적인 문장을 말하는 사람은 자기 자신으로부터 "소외"(alienation) 또는 "분리"(dissociation)되며, "분열된 자아"(divided self)를 갖는다는 직관을 피력했는데,[18] 새로운 이론은 이런 직관을 잘 설명한다. 과연 어떤

사람이 자기 자신으로부터 “소외”된다는 무엇을 의미할 수 있을까? 필자의 견해에 따르면, 무어-역설적인 문장을 발화하는 사람은 어떤 사람을 자기 자신으로 여기는 동시에 (즉, 그 사람을 “나”라는 말로 지시한다는 점에서), 그 사람을 자기 자신이 아닌 다른 사람으로 여기고 있는 셈이다 (자기-지시를 위한 필요조건을 따르지 않는다는 점에서). 이런 점에서 이런 사람은 자기 자신으로부터 “소외”되고 “분리”된 것이다.

5. 핵심 가설에 대한 추가 논증

앞 절에서 <핵심 가설>을 가정했을 때, 무어의 역설이 만족스럽게 설명된다고 논증하였다. 물론 <핵심 가설>에 동조하지 않는 사람에게는 이런 설명은 그다지 인상적으로 보이지 않을 것이다. 본 절에서는 무어의 역설을 잘 설명한다는 것 외에 <핵심 가설>을 받아들일만한 다른 독립적인 이유 몇 가지를 제시하고자 한다.

<핵심 가설>은 자기-지시에 대한 (양상적) 필요조건을 진술한다는 점에서 자기-지시에 대한 부분적인 설명(partial account)을 제공하도록 의도된 것이다. 따라서 <핵심 가설>이 참이라는 것을 설득하기 위해서는 왜 이 가설이 자기-지시에 대한 그럴듯한 부분적인 이론이 될 수 있는지를 설득해야할 것이다. 앞서 말했듯이 자기-지시에 대한 만족스러운 이론은 여전히 존재하지 않지만, 많은 철학자들이 동의할 만한 접근 방법이 있는데, 그것은 자기-지시의 **기능적 역할**(functional role)을 규명함으로써, 그 본성이 밝혀질 수 있다는 생각이다. 다시 말해, 자기-지시가 어떤 합리적 주체에게 어떠한 행동 또는 심리적 차이(difference)를 만들어 내는지를 규명함으

18) 가령, Moran(2001), p.68을 볼 것. 비트겐슈타인 역시 유사한 직관을 피력한 바 있다. Wittgenstein(1974), p.192을 볼 것.

로써, 그것의 본성을 파악할 수 있다는 것이다. 이런 접근 하에서 특별히 많은 관심을 받은 것은, 자기-지시가 **행동**에 대해 만드는 차이이다.

예를 들어 보자. 미경이는 "나는 곧 강의가 있다"라고 말하고, 옆에 있던 철수가 이것을 듣고, "아, 당신은 곧 강의가 있구나"라고 말한다고 하자. 어떤 의미에서 명백히 미경이와 철수는 동일한 믿음을 가지고 있다. (미경이는 그가 믿는 바를 철수에게 말하였고, 그것을 들은 철수는 미경이가 믿는 바를 자신도 믿게 되었다.) 게다가 두 사람 모두 미경이가 강의를 놓쳐서 난처한 상황에 놓이기를 원하지 않는다고 가정하자. 즉, 두 사람이 동일한 믿음과 욕구를 공유하고 있는 것이다. 하지만 물론 두 사람은 다르게 행동할 것이다. 미경이는 곧 강의실로 발걸음을 옮길 것이고, 철수는 가만히 있거나 아니면 미경이에게 빨리 가라고 재촉할지 모른다. 자기-지시에 대해서 연구하는 많은 철학자들은 이런 두 사람이 믿음과 욕구를 공유함에도 불구하고 왜 다르게 행동하는 성향을 갖는지에 대해서 규명하는 것을 자기-지시가 제기하는 핵심 문제로 보았다.[19] 즉, 두 사람의 행동의 차이를 만들어내는 어떤 것이 곧 자기-지시를 구성(constitute)한다는 생각이다.

하지만 행동의 차이가 자기-지시가 만들어 내는 유일한 차이일까? 어떤 점에서 <핵심 가설>은 자기-지시가 사람의 행동에 대해서 뿐만 아니라, **믿음**에 대해서 만드는 차이를 포착하고자 하는 입론이다. 위의 철수와 미경이의 사례와 유사한 사례를 구성해 보겠다. 철수와 미경이는 (이전의 사례에서처럼) 모든 유관한 믿음과 욕구를 공유한다고 해보자. 이제 둘은 각자의 방에서 창밖을 보고 있

19) 자기-지시 문제의 선구자라고 할 수 있는 페리는 이와 같은 예를 통해서 자기-지시 현상이 기존의 지시 이론과 명제 이론에서 설명될 수 없음을 논증하였다. Perry(1977, 1979, 2006)를 볼 것. 위의 미경이와 철수의 예는 Perry(2006)의 예로부터 각색한 것이다.

다. 빗방울이 떨어지기 시작하는 것을 보고, 각각은 "비가 내린다"라고 믿게 된다. (각자는 상대방 역시 같은 경험을 하고 있다는 것을 모른다고 가정하자.) 이 시점에서 철수와 미경이의 마음 상태는 동일할까? 그렇지 않다. 철수는 비가 온다는 믿음을 갖게 되는 동시에, **자기 자신이(즉, 철수가) 비가 온다고 믿는다는 이차 믿음**을 갖게 될 것이며, 미경이는 비가 온다는 믿음을 갖게 되는 동시에 **자기 자신이(즉, 미경이가) 비가 온다**고 믿는다는 이차 믿음을 갖게 될 것이다. 철수와 미경이는 같은 믿음 상태에서 출발했고, 같은 경험을 했음에도 불구하고, 왜 이런 차이가 벌어질까?[20] 명백한 대답은 철수는 철수를 자기 자신으로 여기고, 미경이는 미경이를 자기 자신으로 여기기 때문이라는 것이다. <핵심 가설>은 자기-지시가 만들어 내는 바로 이런 차이가 자기-지시를 구성한다고 주장하는 것이다.

앞서 말했듯이, 자기-지시가 행동에 대해 만드는 차이가 자기-지시를 구성하는 것이 합당하다면, 자기-지시가 믿음에 대해서 만드는 차이 역시 자기-지시를 구성하는 것으로 보는 것 역시 똑같이 합당하다고 보는 것이 합리적이다. 그렇다면 이는 <핵심 가설>이 참이라고 받아들일 만한 좋은 이유가 된다.

여전히 <핵심 가설>에 대해서 의구심을 품는 독자가 있다면, 아마도 이보다 조금 더 직관적인 이유 때문이 아닐까 한다. 이런 독자는 아마도 다음과 같이 물을 것이다. "도대체 어떤 사람을 자기 자신으로 여긴다는 것과, p를 '나는 p를 믿는다'와 대등한 것으로 여긴다는 것이 직관적으로 무슨 관계가 있다는 말인가?" 필자는 이에 대한 개략적이고 다소 사변적인 답변을 제시하면서 본 논문을

20) 물론 두 사람의 이차 믿음이 어떤 점에서는 유사한데, 이는 두 사람의 믿음 모두 "나는 비가 온다고 믿는다"라는 문장으로 적절히 표현된다는 점에서이다. 그러나 두 사람의 발화에서 "나"는 각각 철수와 미경이를 지시하고, 따라서 믿음의 전통적인 견해에 따르면 두 사람의 믿음의 내용은 상이하다.

마무리하겠다.

필자는 이 질문에 대해서 최근에 영향력 있었던 리차드 모란(Richard Moran)의 무어의 역설에 대한 연구에서 중요한 통찰을 얻을 수 있다고 생각한다.[21] 모란은 합리적 주체가 p의 참과 "나는 p를 믿는다"의 참을 대등하게 여기는 성향을, 주체가 자기 자신의 믿음에 대해서 취하는 "숙고적인 태도" (deliberative stance)에 기인한 것으로 본다. "숙고적인 태도"가 무엇인지 설명하기 위해, 우선 행동의 경우를 생각해 보는 것이 좋겠다. 나는 내가 오늘 무슨 옷을 입을지에 대해서 어떻게 판단하는가? 다른 사람은 평소에 나의 옷 입는 습관과 기호 등등을 통해서 내가 무엇을 입을지 **예측**하려 할 것이다. 하지만 나 자신의 경우는 이런 식으로 내가 입을지에 대한 판단에 이르는 것 같지 않다. 나 자신은 이 문제에 대해서 단순히 **방관자** 또는 **예측자**가 아니며, 무엇을 입을지에 대해 결정해야 하는 "**참여자**"이기 때문이다. 내가 묻게 되는 질문은, "내가 무슨 옷을 입는 것이 **적절한가**?"이다.

모란에 따르면, 인간은 근본적으로 자신의 행위를 결정하는 존재인데, 이런 점 때문에 일인칭적인 관점에서는 다음 두 물음이 대등할 수밖에 없다는 것이다.

(9) 내가 A를 할까?
(10) 내가 A를 하는 것이 적절한가?

이는 내가 나 자신의 행위에 대해서 방관자적인 태도를 취하는 것이라 능동적이고 숙고적인 태도를 취하고 있음을 의미한다.

비슷한 아이디어를 믿음에 대해서 적용해 보자. 누군가 나에게

---

21) Moran (1997, 2001, 2012). 이하에서 설명할 모란의 통찰를 자기-지시와 연결시키는 것은 모란의 저서에는 나타나지 않으며, 전적으로 필자의 아이디어이다.

“너는 국민의당이 선거에서 선전할 것이라고 믿니?”라고 묻는다면, 나 자신은 이 문제에 대해 방관자가 아닌, 무엇을 믿어야 할지를 결정하는 참여자가 된다. 이런 점에서 다음과 같은 두 물음이 대등하게 취급될 수밖에 없다.

(11) 내가 국민의당이 선거에서 선전할 것이라고 믿는가?
(12) 국민의당이 선거에서 선전할 것이라고 믿는 것이 적절한가?

그렇다면 (12) 물음에 대한 답은 어떻게 내리게 되는가? 흔히 받아들여지기로 믿음의 규범은 참이라고 한다. 즉, 어떤 것을 믿는 것이 적절한지에 대한 질문은 항상 그것이 참인지에 따라 결정되어야 한다는 것이다. 따라서 나에게 (12)는 다음과 대등한 질문이 될 수밖에 없다.

(13) 국민의당이 선거에서 선전할 것이라는 것은 참인가?

이런 식으로 나는 왜 p의 참 거짓이 “내가 p를 믿는다”의 참 거짓과 직접적으로 대등하게 여기는지에 대한 설명을 얻게 된다.

필자가 보기에 이것이 옳다면, <핵심 가설>에 대한 조금 더 직관적인 설명을 얻게 된다. 자기-지시는 이 세계에 존재하는 어떤 객관적인 사람을 나 자신으로 여김의 문제이다. 내가 세계에 존재하는 수많은 사람들 중에 특정한 한 사람을 특별히 나 자신으로 여긴다는 것이 무엇인가? 이는 그 사람의 행위와 믿음에 대해서 방관자적인 태도를 취하는 것이 아니라, 그 사람의 행위와 믿음을 내가 결정할 문제로, 즉 이에 대해 숙고적이고 참여적인 태도를 취한다는 것에 다름 아니라는 생각이 상당히 그럴듯하게 들린다.

6. 결론

이상에서 무어의 역설에 대한 기존의 견해들을 비판하고, 새로운 이론을 제안하였다. 새로운 견해의 핵심은 p와 "나는 p를 믿는다"를 대등하게 여기는 성향이 자기-지시를 구성한다는 아이디어이다. 본문에서 <핵심 가설>이라 불렀던 이 가설이 주어지면, 이로부터 무어의 역설에 대한 설명은 비교적 명쾌하게 이루어진다고 주장하였다. 그렇다면 필자가 제시한 무어의 역설에 대한 설명이 얼마나 만족스러운지는 <핵심 가설>이 얼마나 그럴듯하냐에 달려 있다고 할 수 있다. 이에 바로 앞 절에서 필자는 이 가설을 받아들일 만한 독립적인 이유가 있다고 논증하였다. 필자는 이 논증들이 완결된 논증이라고는 생각하지는 않는다. 이를 더 발전시키는 것은 차후의 연구를 위해서 남겨두기로 하겠다.

참고문헌

Byrne, A. (2005) “Introspection”, *Philosophical Topics,* 33, pp. 79-104.

Chan, T. (2010) “Moore's Paradox Is Not Just Another Pragmatic Paradox”, *Synthese,* 173, pp. 211-229.

DeRose, K. (1991) “Epistemic Possibilities”, *Philosophical Review,* 100, pp. 581-605.

Evans, G. (1982) *The Varieties of Reference,* New York: Oxford University Press.

Fernandez, J. (2013) *Transparent Minds: A Study of Self-Knowledge*, New York: Oxford University Press.

Greco, D. (2015) “Iteration and Fragmentation”, *Philosophy and Phenomenological Research,* 91, pp. 656-673.

Heal, J. (1994) “Moore’s Paradox: A Wittgensteinian Approach”, *Mind,* 103, pp. 5-24.

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press.

Green, M. and Williams, J. N. (2007) “Introduction”, in M. Green, et al. (eds) *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*. New York: Oxford University Press, pp. 3-36.

Kaplan, D. (1989) “Demonstratives,” in J. Almog, et al. (eds) *Themes from Kaplan*. New York: Oxford University Press, pp. 481-563.

Lewis, D. K. (1979) “Attitudes *De Dicto* and *De Se*”, *The Philosophical Review,* 88, pp. 513-45.

Moore, G. E. (1944/1993) “Moore’s Paradox”, in T. Baldwin (ed.)

G. E. Moore: Selected Writings, London: Routledge, pp. 207-212.

Moran, R. (1997) "Self-Knowledge: Discovery, Resolution, and Undoing", *European Journal of Philosophy,* 5, pp. 141-161.

Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.

Moran, R. (2012) "Self-knowledge, 'Transparency,' and the Forms of Activity", in D. Smithies, and D. Stoljar (eds.) *Introspection and Consciousness*. Oxford University Press, pp. 212-235.

Perry, J. (1977) "Frege on Demonstratives". *Philosophical Review,* 86, pp. 474-497.

Perry, J. (1979) "The Problem of the Essential Indexical", *Noûs,* 13, pp. 3-21.

Perry, J. (2006) "Stalnaker and Indexical Belief", in J. Thomson, and A. Byrne (eds.) *Content and Modality: Themes from the Philosophy of Robert Stalnaker*, New York: Oxford University Press, pp. 204-221.

Searle, J. R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, New York: Cambridge University Press.

Shoemaker, S. (1995) "Moore's Paradox and Self-Knowledge", *Philosophical Studies* 77, pp. 211-228.

Stalnaker, R. (2000) "On Moore's Paradox", in P. Engel (ed.) *Believing and Accepting,* Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 93-100.

Williams, J. N. (2004) "Moore's Paradoxes, Evans's Principle, and Self-Knowledge", *Analysis,* 64, pp. 348-353.

Wittgenstein, L. (1974) *Philosophical Investigations*, Trans. G. E.

M. Anscombe, Oxford: Blackwell.
Wright, C. (1998) “Self-Knowledge: The Wittgensteinian Legacy”, in C. Wright et al. (eds), *Knowing Our Own Minds,* New York: Oxford University Press, pp. 13-46.

연세대학교
Yonsei University
hongwoo@gmail.com

ARTICLE ABSTRACTS

Moore's Paradox and Self-Reference

Hongwoo Kwon

Asserting a sentence of the form “p but I do not believe that p” sounds inappropriate, and even absurd or contradictory. The problem that Moore's paradox raises is to explain why asserting such a sentence is absurd despite the fact that there is apparently no logical contradiction in it. Many of the influential accounts of Moore's paradox try to locate its source in the nature of belief or in the nature of assertion. In this paper, I argue that these accounts are not satisfactory, and develop and defend a novel account. According to this account, the source of Moore's paradox should be located in self-reference. Self-reference is constituted by a certain disposition to form second-order beliefs. A subject who is ready to assert “p but I do not believe p” fails to conform to the disposition that is constitutive of self-reference, while at the same time referring to the relevant individual with “I.”

Key Words: Moore's Paradox, Self-Reference