creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

문학석사 학위논문

자기지식에 대한 현상적 투명성 이론

2022 년 2 월

서울대학교 대학원

철학과 서양철학전공

신 희 재

자기지식에 대한
현상적 투명성 이론

지도 교수 한 성 일

이 논문을 문학석사 학위논문으로 제출함
2021 년 10 월

서울대학교 대학원
철학과 서양철학전공
신 희 재

신희재의 문학석사 학위논문을 인준함
2022 년 1 월

위 원 장 _____ 김 기 현 _____ (인)

부위원장 _____ 한 성 일 _____ (인)

위 원 _____ 김 현 섭 _____ (인)

초 록

이 논문의 목표는 지향적 태도에 대한 기존의 투명성 이론들의 한계를 지적하고 투명성 이론의 계열 안에서 하나의 대안적 설명인 현상적 투명성 이론을 제시하는 것이다. 여기서 '투명성'은 마음이 인식 주체와 맺는 관계적 속성으로, 인식 주체가 그것을 투과해서 세계의 대상을 바라본다는 의미를 담는다. 즉 심적 상태는 나의 의식이 바깥의 대상으로 나아갈 때 투과하는 투명한 유리와도 같다는 의미에서 투명하다. 본고는 투명성을 통한 일인칭 자기지식의 대상으로 지향적 태도 중에서도 믿음을 범례로 집중한다. 이는 스스로의 믿음을 직접적으로 아는 능력이 주체의 합리성과 긴밀하게 연결된 근본적인 능력이며 따라서 믿음에 대한 자기지식을 해명하는 것은 합리적 행위자성 전반과도 연결된 중요한 문제이기 때문이다.

본고는 투명성 이론가들을 살펴보기에 앞서 지향적 태도를 온전히 내성을 통해 발견하는 대상으로 상정하는 기존의 내성 이론이 일인칭 시점에서 직접적으로 얻어지는 자기지식에 대한 적절한 설명일 수 없다고 주장한다. 이는 기존의 내성 이론과 같은 발견 모형이 무어의 역설(Moore's paradox)이 암시하는 '일인칭 시점에서 얻어지는 자기지식이라면 사태에 대한 판단과 동기화되는 방식으로 얻어지는 것이어야 한다'는 제약 조건을 만족할 수 없기 때문이다.

반면 세계를 향하는 고려를 통해서 자기지식에 정당하게 도달할 수 있다는 투명성 이론은 무어의 역설이 주는 제약 조건을 만족한다는 점에서 자기지식에 대한 적절한 설명을 제공한다. 본고에서는 투명성에 주목한 이론으로 번의 투명성 추론주의, 모란의 합리적 행위자 이론, 보일의 반성주의 이론을 살펴본다. 하지만 세 학자는 공통적으로 투명성 절차에서 사태의 판단에 이르게 되는 판단의 메커니즘이 무엇인지에 대한 설명을 결여하고 있다는 점에서 자기지식 모형으로서 충분한 설명을 제공하지 못하고 있다.

논문의 마지막장에서 나는 투명성 절차에서 사태의 판단에 도달하는 메커니즘을 태도-특유의 현상성을 대면하는 과정으로 설명하는 현상적 투명성 이론을 제시한다. 현상적 투명성 이론에 따르면, 태도-특유의 현상성에 대한 대면은 자기지식의 근거를 제공하며, 이는 일인칭 주체에게만 허용된다는 점에서 자기지식을 얻기 위한 특유한 접근이다. 또한 이러한 자기지식은 직접적인 자기지식인데, 믿음 상태에 대한 이론적 증거에 근거하지 않고 투명성 절차를 통하여 얻어지기 때문이다. 마지막으로 주체의 심적 체계가 기능성과 현상성 사이에 내적 일관성을 이루면서 작동하는 한 주체는 투명성 절차를 통하여 자신의 심적 상태를 정당하게 귀속할 수 있는 인식적 권리를 가지며, 이러한 가정 덕분에 투명성 절차와 사태에 대한 고려에 동반하는 현상성의 대면을 통해 얻어지는 자기지식은 인식적으로 보증된다.

주요어 : 자기지식, 투명성 이론, 특권적 접근, 지향적 태도, 태도 특유의 현상성, 무어의 역설, 심리철학

학 번 : 2019-25382

목 차

제1장 서 론.....1

제2장 철학적 주제로서 "자기지식".....4

2.1 자기지식과 일인칭 특권 4

2.2 철학에서의 자기지식 8

2.3 자기지식의 방법 1. 내성 18

2.4 자기지식의 방법 2. 투명성 33

제3장 투명성과 자기지식.....39

3.1 심적 상태의 투명성 이해하기.....39

3.2 번의 투명성 추론주의 이론.....44

3.3 모란의 합리적 행위자 이론.....52

3.4 보일의 반성주의 이론.....63

3.5 세계의 대상에 대한 고려와 투명성.....69

제4장 자기지식에 대한 현상적 투명성 이론.....73

4.1 자기소외적 믿음과 일인칭 시점에서의 경험.....73

4.2 믿음 속성 특유의 현상성.....78

4.3 믿음에 대한 현상적 투명성 이론과 질문들.....87

4.4 특유한 접근과 일인칭 특권.....91

제5장 결론.....95

참고 문헌.....97

Abstract.....101

제 1 장 서 론

이 글의 목표는 심적 상태에 대한 자기귀속에서 인식 주체에게 허용되는 것처럼 보이는 일인칭 특권을 설명하는 것이다. 나는 크게는 다음의 두 질문에 대해 대답하고자 한다. 첫째, 주체에게 자신의 마음은 데카르트적 견해가 주장하는 것만큼 완전히 알려지지 않는 것 같음에도 여전히 타인의 마음과는 다른 방식으로 알려지는 것 같은데, 그 특유한 접근(peculiar access)은 무엇인가? 그리고 특유한 방식으로 얻어지는 일인칭 자기지식은 어떻게 인식적 특권성(privilege)을 가지고 있는가? 다양한 심적 상태 중에서도 나는 **지향적 태도**에 대한 자기지식을 해명 대상으로 삼으며, 지향적 태도에 대한 자기지식을 해명하기 위해서는 투과해 세계를 바라보는 투명한 유리와도 같은 마음의 **투명성**(Evans 1982, Moran 2001; 2003; 2012, Byrne 2018, Boyle 2011; 2019)에 주목해야 함을 주장할 것이다.

본격적인 논의에 앞서 1 장에서는 자기지식에 대한 기존의 철학적 논의를 개관적으로 소개한다. 전통적으로 자신의 마음은 주체에게 가장 확실하게 알려지는 대상이면서 동시에 주체의 인식적, 존재론적 근원으로 여겨졌다. 하지만 19 세기 이후 정신분석학의 등장이 보여준 무의식의 영역이나 자기지식이 오류가능하다는 심리학적 연구들은 마음에 대한 일인칭 특권성에 대해 의문을 제시했다. 또한 자연과학적 방법론으로 이루어지는 마음에 대한 뇌과학/신경과학/인지과학적 접근은 근본적으로 일인칭적 관점을 벗어나 삼인칭적 관점에서 마음을 탐구하였으며 이들의 유용성은 사람들로 하여금 마음과 주체 사이의 긴밀한 관계에 특별한 관심을 두지 않도록 만들었다. 하지만 한편으로 우리는 일인칭 주체가 타인에게는 허용되지 않는 특유한 접근을 통해서 스스로의 마음을 알 수 있다고 생각하며, 철학적으로는 자기지식의 일인칭 특권이 합리적 행위자라면 반드시 가져야하는 특성으로 주장된다. 심적 상태의 자기 귀속에 있어서 상정되는 주체의 특유한 접근과 일인칭 특권이 무엇인지를 탐구하기 앞서 우선 자기지식의 대상인 심적

상태를 구성하는 두 속성 범주인 현상성과 기능성을 개념적으로 정리하고, 자기지식의 대상과 자기지식의 방법을 대략적으로 구분함으로써 논의 대상을 지향적 태도로 좁힌다. 이어서 자기지식을 내성의 능력으로 설명하는 두 모형인 대면 이론과 내적 감각 이론을 검토해보고, 전통적인 내성 모형은 지향적 태도에 대한 일인칭 특권적 앎을 설명하는 데 문제를 가진다는 것을 무어의 역설(Moore's Paradox)을 통해 살펴볼 것이다. 그리고 지향적 태도에 대한 앎을 설명하는 비교적 새로운 방식인 투명성 절차는 무어의 역설이 암시하는 특권적 자기지식의 제한 요건에 구속되지 않음을 간단히 설명하고, 자기지식의 특권성을 설명할 수 있는 특유한 방법으로 고려될 수 있음을 주장할 것이다.

2 장에서는 자기지식의 특유한 방법과 특권성을 설명하기 위해서 심적 상태의 투명성에 주목하는 세 명의 철학자—번(Alex Byrne), 모란(Richard Moran), 보일(Matthew Boyle)의 견해를 살펴볼 것이다. 편의를 위해 자기지식의 특유한 접근과 특권성을 투명성과 추론을 통해서 설명하는 번의 입장을 '추론주의'로, 합리적 행위자성과 투명성의 긴밀한 관계를 드러냄으로써 설명하려는 모란의 입장을 '합리적 행위자 이론'으로, 세계에 대한 지향에 전제되는 의식의 방식에 대한 반성을 통해서 설명하는 보일의 입장을 '반성주의'로 부를 것이다. 이어서 투명성 이론가들의 입장들을 각각 살펴보고, 이들이 심적 상태의 일인칭적 투명성을 자기지식의 문제와 적절하게 연결했음에도 자기지식의 특유한 접근과 특권성을 설명하기에는 제각기 부족한 점이 있다고 비판할 것이다. 그리고 논의를 토대로 이들의 교훈을 수용한 자기지식 이론이라면 갖춰야 할 덕목을 제시한다.

3 장에서는 투명성에 주목하면서도 자기지식의 획득에 있어서 주체의 현상성에 주목하는 현상적 투명성 이론을 제안하고 설명한다. 현상적 투명성 이론은 주체가 스스로에게 P 에 대한 믿음을 귀속시키기 위해서 자신의 안에 있는 표상 P 를 살펴보는 것이 아니라, 다만 세계의 비-심리적 사태(혹은 명제) P 가 참인지에 대해서 생각한다는 투명성 절차를 받아들인다는 점에서 투명성 이론의 일종으로 볼 수 있다. 현상적 투명성 이론의 문제 의식은

투명성 절차를 통해서 자신이 대상에 대해 어떤 태도를 취하는지 알기 위해서는 태도-특유의 현상성이 필요하며 이는 다름 아니라 동시에 사태에 대한 판단의 근거가 되는 그러한 현상성이라는 것이다. 스스로 믿음을 귀속하기 위해 투명성 절차를 따라 세계의 사태 혹은 명제가 참인지 살펴볼 때, 주체가 만일 P 를 믿는다면 믿음-특유의 현상성이면서 동시에 참인 사태를 고려할 때 발생하는 현상성이 동반하며 이에 근거해 고려의 대상이 된 명제가 참이며 따라서 자신은 해당 명제를 믿는다고 판단하게 된다. 그리고 이어서 이처럼 투명성 이론의 계열에서 그럼에도 판단의 메커니즘에서 현상성이 자기지식의 근거가 된다는 나의 입장-현상적 투명성 이론-은 자기지식의 특유한 접근과 인식적 특권성을 설명해주는 자기지식 모형이 될 수 있다고 주장한다. 자기지식을 얻기 위해서 투명성 절차를 사용하고, 이 과정에서 주어지는 현상성을 대면하는 것은 주체에게만 허용되는 접근이기 때문에 현상적 투명성은 특유한 접근을 설명해주며, 최소한의 합리적 주체에게는 사태와 동반하는 현상성을 통해 얻게 되는 지향적 태도에 대한 판단이 현상성이 나의 심적 체계에 내적 일관성을 지키는 방식으로 주어지기 때문에 투명성 절차를 통해 얻어진 자기귀속 믿음은 지식으로서 보증되기 때문이다. 따라서 만일 현상적 투명성 이론이 옳다면, 기존의 투명성 이론들을 보완하여 일인칭 시점에서 자기지식을 얻는 하나의 가능한 모형을 제공함으로써 자기지식의 특유한 접근과 특권의 성격을 해명하는 데 기여할 것이다.

제 2 장 철학적 주제로서 "자기지식"

2.1. 자기지식과 일인칭 특권

친구는 뭣 때문에 나에게 그런 서운한 표정을 지은 걸까? 동생이 나에게 화를 내는 이유는 무엇일까? 지금 저 사람은 무엇을 보고 있는 걸까? 누구나 이처럼 종종 남이 어떤 마음에서 그런 말과 행동을 하는 건지 진지하게 고민하게 되는 경험을 할 것이다. 물론 남의 마음이 어떤지 아는 것이 굉장히 수월한 경우도 있겠지만, 때로는 아무리 고민해도 도대체 상대방이 어떤 생각을 갖고 있는지 추측조차 안되는 경험을 하게 된다.

그에 비해서 나의 마음은 나에게 너무나도 쉽고 정확하게 알려지는 것 같다. 나는 지금 맑은 하늘 하얀 구름을 보고 있다. 동시에 상쾌한 기분을 느낀다. 오늘 하루가 종일 이렇게 맑기를 바란다. 그리고 아침에 들었던 기상일보를 떠올리면서, 오늘 하루종일 맑을 것이라고 믿는다. 이상의 진술들은 내가 무엇을 보고, 기분을 느끼고, 무언가를 바라고 또한 믿는지를 보고한다. 나는 나의 마음의 상태, 즉 심적 상태에 대해서 이처럼 자연스럽게 진술할 수 있다. 나의 심적 상태에 대해서 자연스럽게 진술할 수 있는 까닭은 내가 나의 심적 상태에 대해서 알고 있기 때문이다. 나는 내가 맑은 하늘 속 하얀 구름을 보고 있다는 것을 안다. 그리고 상쾌한 기분을 느낀다는 것을 알고, 오늘 하루종일 이렇게 맑기를 바란다는 것을 안다. 그리고 오늘 하루종일 맑을 것이라고 믿는다는 것도 안다.

이처럼 나의 마음만큼 나에게 확실하게 알려지는 것은 잘 없는 것 같다. 하지만 정작 나의 마음을 내가 어떻게 아는 건지 물어보면 어떤 대답을 해야할지 난감하다. 대부분의 경우 나의 마음은 나 자신에게 내밀하게 알려지면서도, 그렇기 때문에 그것이 알려지는 방식에 대해서 굳이 고민해볼 필요가 없기 때문이다. 다만 추론적으로 얻어지며 경험적인 정당성을 통해서 뒷받침되어야 하는 마음 외적인 세계의 사태들에 대해서와는 달리 나의 마음에 대해서는 직접적이면서도 확실하게 알 수 있다고 생각해왔을 뿐이다.

이렇게 직접적이면서도 확실한 앎의 대상으로서 마음에 대한 생각은 적어도 데카르트에게까지 거슬러 올라간다. 나의 마음은 나에게 별다른 증거와 추론 과정없이 직접적으로 알려지며, 인식적으로도 그 자체로 정당하다는 생각은 전통적으로 '마음의 투명성'이라는 이름으로 전제되어 왔다.[1] 여기서 투명성은 뭉뚱그려 표현하자면 스스로에게 확실하게, 직접적으로 알려지는 마음의 속성이며, 또한 마음이 가지는 관계적인 속성이다. 즉, 나의 마음은 "나에게 있어서" 투명하다. 나의 마음은 타인에게 투명하지는 않고, 타인의 마음은 나에게 투명하지는 않다. 데카르트 이래로 자신의 심적 상태에 대한 앎은 어떤 것에 대해서보다도 인식적으로 확실한 근거를 소유하는 것처럼 여겨져왔다. 데카르트적 심신 이원론은 세계의 대상, 심지어 나의 신체에 대해서는 회의할 수 있을지언정 나의 생각의 존재에 대해서는 회의할 수 없다는 명석판명한 직관을 전제로 작동한다. 근대의 회의론자로 이해되는 흄이 세계를 향한 탐구의 가장 확실한 자료로 삼는 것도 인상과 관념이라는 나의 마음의 자료들이다. 적어도 많은 철학자들에게 나의 마음에 대한 의식은 우리에게 가장 직접적으로 접근 가능한 것, 따라서 세계에 대한 이론을 가능하게 해주는 조건으로 생각되어 왔다.

스스로의 마음에 대한 완벽할 정도의 투명성을 전제하는 견해를 '데카르트적 견해'라고 불러보자. 데카르트적 견해는 마음에 대한 다음의 몇 가지 성격을 전제한다. 첫째로 나의 마음은, 어떤 심적 상태가 있다면 그것은 나에게 알려진다는 의미에서 자기-고시적(self-intimating)이다. 둘째로, 스스로의 심적 상태에 대해서 내가 어떤 판단을 한다면, 그 판단은 틀릴 수 없다. 즉, 마음에 대한 스스로의 판단은 오류불가능(infallible)하다. 셋째, 나의 마음에 대한 판단은 타인에 의해서 교정불가능(incorrigible)하다. 나의

---

[1] 이때의 투명성은 물론 서론에서 언급한 '그것을 투과해 세상으로 나아가는 통로'로서의 투명성과는 다른 의미이다. 두 의미를 가지는 '투명성'이 혼용될 2 절부터는 두 가지 의미의 '투명성'를 서로 다른 표현을 통해 명확하게 구분할 것이다.

마음에 대해서 어떤 판단을 내렸을 때, 나 이외의 타인이 나의 판단이 틀렸음을 보이는 것은 불가능하다.

하지만 오늘날 일상적으로 그리고 철학적으로도 데카르트적 견해를 그대로 수용하는 사람은 거의 없다. 어떤 심적 상태는 무의식의 영역에 있으며 따라서 주체에게 접근가능하지 않을 수 있다는 정신분석학의 교훈은 많은 사람들을 설득했다. 스스로의 마음에 대한 판단이 많은 경우 부정확하며, 일반적으로 자신의 마음을 보고할 때 지어내는 경향이 사람들에게 있다는 심리학적 연구들은 마음에 대한 판단이 정말로 오류불가능한 것인지 의심스럽게 만든다.[2] 뇌과학, 신경과학과 같은 특수 과학적 방법론에 근거하여 진행되는 마음에 대한 연구들은 일인칭 주관적 시점이 아닌 삼인칭 관점에서 마음을 분석한다. 그리하여 자연스럽게 주체와 마음 사이에 존재한다고 생각되었던 특별한 관계적 속성인 투명성은 그들의 연구에서 별다른 관심의 대상이 아니게 되었다. 많은 사람들은 스스로에게 마음의 문제가 있다고 생각할 때 정신과에 상담을 받고, 정신과 전문의가 나의 마음에 대해서 나보다 더 잘 알 수 있는 가능성을 당연하게 생각한다.

철학적 논의에서는 퍼트넘(1975)의 쌍둥이 지구 사고실험, 그리고 버지(1979, 1988)의 '관절염' 사고실험 이래로 많은 철학자들은 심적 내용에 대한 외재주의를 진지하게 받아들이게 되었다. 자연스럽게 심적 내용 외재주의가 자신의 마음이 직접적이면서도 특권적인 방식으로 스스로에게 알려진다는 데카르트적 견해와 상충하는지에 대해서 철학적 논의가 활발하게 이루어졌다.[3] 심적 내용 외재주의가 보여주는 바는 하나가 다른 하나의 엄밀한 물리화학적 복제자이더라도 그들에게 정당하게 귀속되는 심적 내용이 다를 수 있다는 것이다. 퍼트넘과 버지의 사고실험 이후, 심적 내용 외재론은 주체가 자신의 내면을 살펴보는 것만으로 어떻게 심적 내용에 대해

---

[2] Nisbett and Wilson (1977)

[3] McKinsey (1991; 1994), Boghossian (1997), Brueckner (1990; 1992), Davidson (1987), Burge (1988)

특권적으로 알 수 있는지에 대한 의문을 던지며 따라서 자기지식의 직접성에 대한 심각한 문제를 제기하는 것으로 이해되었다.

하지만 이러한 반발들에도 불구하고 우리는 여전히 자기지식의 일인칭 특권에 대한 강한 직관을 갖는 것처럼 보인다. 설령 무의식의 영역을 인정하여 의식되지 않는 심적 상태가 있음을 받아들이더라도, 직접적으로 나에게 의식되는 심적 상태들이 많이 있다는 생각은 상식적이다. 어떤 경우 나의 심적 상태에 대해서 나도 모르게 부풀리거나 지어내서 보고한다고 하더라도, 이로부터 나에게 조심스러운 반성을 통해서 나의 마음을 정확하게 보고할 능력이 없다는 것이 따라나오지는 않는다. 어떤 철학자들은 비록 내 심적 내용이 가지는 의미론적 요소를 내성으로 완전히 알 수는 없더라도, 의미론적 요소와 긴밀한 연관을 갖는 것처럼 보이는 "좁은 내용"을 알 수 있다고 이야기한다. 무엇보다 자기지식과 타인지식 사이에는 분명히 불균형(disparity)이 존재하는 것 같다. 남이 무엇을 믿고 욕구하고 또 느끼는지는 그 사람의 행동을 통해서 간접적으로 알게 되는 반면, 내가 무엇을 믿고 욕구하고 느끼는지를 알기 위해서 나는 대부분의 경우 마음 외적 증거에 기초해서 추론하지 않는 것처럼 보인다.

이 글의 목적은 데카르트적 견해에 대한 반발에도 불구하고 일인칭 주체에게 여전히 직관적으로 주어지는 것처럼 보이는 일인칭 특권을 설명하는 것이다. 나의 마음은 비록 데카르트적 견해가 주장하는 바처럼 완전히 투명하지는 않으면서도 여전히 타인의 마음과 알려지는 방식이 다른 것 같은데, 이 **특유한 방식**(peculiar access)이 과연 무엇인가? 그리고 특유한 방식으로 얻어지는 자기지식이 어떤 점에서 **인식적 특권성**(epistemic privilege)을 가지고 있는가? 이어지는 2 절에서는 앞으로의 논의를 위해서 자기지식의 대상과 방법을 구분하면서 논의에 필요한 개념을 명료하게 할 것이다.

2.2. 철학에서의 자기지식

2.2.1. 심적 상태의 현상성과 기능성

자기지식의 대상은 자신의 심적 상태이다. 일반적으로 철학자들은 심적 상태의 중요한 두 측면으로 그 심적 상태가 주체에 의해서 의식될 때 주체가 경험하는 질적인 느낌인 현상성과 심적 상태에 놓인 주체가 특정 행동적/심리적 성향을 갖도록 만드는 기능성에 주목한다. 심적 상태의 현상성과 기능성은 심적 상태의 본질을 결정하는 구성요소이면서 동시에 특정 개체에 심적 상태를 귀속시킬 때 고려되는 인식적 근거가 된다.

차머스(1996: 16-24)에 따르면, 모든 심적 속성은 현상적 속성이거나 심리적 속성이거나 혹은 둘의 조합이며, 이 구분 기준은 모든 심적 속성류를 포괄한다. 여기서 차머스가 '심리적 속성'이라고 부르는 것은 일반적으로 심적 상태의 현상적 요소를 제외한 기능적 요소를 의미하는 것으로 보아도 무방하다. 일상적으로 사용되는 심적 개념들 중 상당수는 현상적, 기능적 측면을 모두 동시에 가지고 있는 것으로 여겨진다. 예컨대 개념 〈고통〉같은 경우는 특정한 불쾌한 현상적 느낌을 가리키기 위해서 사용된다.[4] 하지만 동시에 우리는 고통이 유기체에 특정한 손상으로 인해 야기되고, 피하기 반응을 동반하는 어떤 기능적 속성을 가리킨다고도 생각한다. 즉 일반적으로 〈고통〉은 주체가 느끼는 특유한 질적 경험과 그가 보이는 특정 유형의 성향을 이중적으로 가리킨다. 대부분의 심적 개념들은 이처럼 현상적/기능적 속성을 이중적으로 가리키는 것으로 사용된다.

핵심은 모든 심적 개념은 현상적 속성이나 기능적 속성, 혹은 둘의 조합을 가리키며 이 구분은 심적 속성의 영역을 **포괄**한다는 것이다. 다만 차머스는 여기서 〈고통〉, 〈느낌〉, 〈믿음〉, 〈욕구〉같은 개별 심적 개념들이 현상적 속성과 기능적 속성 중 무엇을 핵심적 개념의 요소로 가지는지는 우리의 일상적인 직관에 맡길 뿐 구체적으로 논하고 있지는 않다. 다만 그는 어떤

---

[4] 편의를 위해, 앞으로 개념을 가리키기 위해 사용되는 용어는 '〈 〉'로 표시하도록 할 것이다.

개념이 주로 심리적(기능적) 개념인지를 보기 위한 다음의 테스트를 제안한다: 심적 속성 M 이 어떠한 특유한 현상적 느낌없이 예화되는 것이 가능한지를 상상해보고, 만일 그렇다면 〈M〉은 심리적(기능적) 개념이며, 아니라면 〈M〉은 현상적 개념이거나 혹은 복합적 개념이다. 이 제안을 따른다면 예컨대 학습(learning)은 어떤 특유한 현상적 속성의 예화없이도 예화될 수 있어 보인다는 점에서 〈학습〉은 심리적 개념이고, 현상적 느낌없이 예화되는 감각(sensation)을 상상하기 어렵다는 점에서 〈감각〉은 현상적 개념이거나 최소한 복합적 개념이다. 또한 지각(perception)은 심적 속성에 주체와 세계의 대상 사이의 인과적 관계를 개념적으로 요구하는 것처럼 보이는데, 따라서 〈감각〉과는 달리 〈지각〉은 심리적 개념에 가까워 보인다.

〈믿음〉의 경우는 〈학습〉, 〈감각〉과는 달리 현상적 개념 혹은 기능적 개념인지를 확정하는 것이 직관적으로 보다 어렵고 복잡해 보인다. 일반적으로 믿음은 인간의 인지 체계에서 다른 심적 상태들 혹은 행위와 맺는 다양한 관계를 통해 인과적 기능을 수행하는 기능적 속성을 가리키는 것으로 생각된다. 하지만 어떤 철학자들은 믿음이 현상성과 밀접한 연관성을 가진다는 입장들도 제기하기도 한다.[5]

각각의 심적 개념이 현상적 개념인지, 심리적 개념인지, 혹은 복합적 개념인지를 결정하는 것은 이 글의 목표에서 벗어난다. 다만 믿음의 경우 일반적으로 믿음을 현재 의식되지 않는 상설적 믿음(standing belief)와 지금 의식되고 있는 발생적 믿음(occurrent belief)으로 구분하는 기준을 받아들인다면, 상설적 믿음은 의식되지 않는 동안에도 존재하는, 따라서 어떠한 현상성도 결여된 속성이고, 발생적 믿음은 의식되는, 따라서 어떠한

[5] 예를 들어 Schwitzgel (2001), 혹은 현상적 지향성론자들; Horgan & Tienson (2002), Pitt (2004), Kriegel (2011), Mendelovici (2010, 2018)

현상성이 동반되는 속성이다.[6] 그리고 만일 두 속성을 포괄적으로 가리키는 〈믿음〉이 중의적인 개념이 아닌 이상 복합적 개념이라고 가정하는 것은 자연스러운 한 가지 방식일 것이다.

어떤 심적 상태와 그 상태의 현상적 측면과 기능적 측면 사이의 형이상학적 관계는 어떻게 이해되어야 할까? 자기지식의 논의에서, 일반적으로 자기지식의 대상으로서 심적 상태라는 대상이 실제로 있음을 가정하는 것처럼 보인다. 심적 상태가 속성을 가질 수 있는 어떤 대상이라면 현상성과 기능성은 각각의 심적 상태에 예화되는 속성으로 이해할 수 있을 것이다. 하지만 주체가 심적 상태를 가진다는 것은 일반적으로 주체에게 특정 심적 속성이 귀속되는 것으로 이해되기도 한다. 이 경우 현상성과 기능성은 어떤 심적 속성이 가지는 두 가지 측면으로 이해될 수 있겠다. 논의의 편의를 위해서 이 글에서 심적 속성 M 의 예화 M_1 은 어떤 현상성과 기능성을 속성으로 가질 수 있는 존재자라고 가정하겠다. 물론 현상성과 기능성은 심적 세계를 빠짐없이 포괄하는 두 속성 범주이기 때문에, M_1 이 심적 속성의 예화가 되기 위해서는 현상적 속성이나 기능적 속성 중 어떤 속성은 반드시 예화를 해야한다. 또한 앞선 논의에 따라서 어떤 심적 상태 M_1 이 순수한 현상적 개념 P 의 외연에 속하는 심적 상태라면 M_1 은 개념적으로 기능적 속성을 결여할 수밖에 없을 것이고, 만일 순수한 기능적 개념 F 의 외연에 속하는 심적 상태라면 개념적으로 현상적 속성을 결여할 수밖에 없을 것이다. 하지만 심적 상태 M_1이 대부분의 심적 개념처럼 복합적

---

[6] 물론 의식되지 않더라도 현상성을 가질 수 있거나, 의식되더라도 현상성을 갖지 않을 수 있다는 입장이 있을 수 있다. 전자의 예시로 의식에 대한 고차 이론에서의 고차 상태를 갖지 않는 일차 심적 상태가, 후자의 예시로 블록(Block, N. 1995)의 현상성이 결여된 접근적 의식(access-consciousness)를 들 수 있다. 하지만 일반적으로 현상성은 곧 주체에게 의식적으로 경험되는 방식이며, 의식은 곧 주체의 입장에서 무언가를 경험함으로 이해되며, 현재의 논의에서는 이에 따라서 의식되는 심적 상태는 현상성을 가지고, 현상성을 갖는 심적 상태는 의식된다는 것을 전제로 가정할 것이다.

개념이 적용되는 속성의 예화라면, M_1 은 특정한 현상성과 기능성을 모두 예화하는 심적 상태일 수 있다.

일상적으로는 복합적 심적 개념의 현상적/기능적 측면을 구분할 일은 거의 없고 따라서 현상성과 기능성을 가리키는 데에 별다른 개념 구분을 두지 않지만, 철학적 논의에서는 때로는 해당 개념을 통해서 심적 상태의 어떤 속성을 가리키는지 분명히 하는 것이 중요하다. 일상에서 심적 개념의 두 측면을 구분할 일이 거의 없는 것은, 한 개념에 의해서 포괄되는 현상적 속성과 기능적 속성이 일반적으로 공외연적이기 때문이다.[7] 일반적으로 나타나는 한 개념에 의해서 포괄되는 현상적 속성과 기능적 속성의 공외연성은 해당 현상성과 기능성이 같은 상태에 동시예화되는 것으로 이해될 수 있다. 고통을 느끼는 일반적인 상황을 생각해보자. 고통의 주체는 고통의 원인을 피하려고 하고, 병원에 가거나, 구급약품함을 찾거나, 혹은 신음소리를 내면서 고통의 기능적 속성을 예화한다. 동시에 그는 고통 특유의 현상적 경험을 하면서 현상성을 예화한다. 이 경우처럼 일반적으로 고통 상태를 가지는 주체는 고통의 현상성과 기능성을 동시에 가지는 그러한 고통 상태를 겪고 있다.[8] 이처럼 복합적 심적 개념의 예화 M_1 이 그 복합적 심적 개념이 포괄하는 현상적 속성과 기능적 속성을 모두 예화하고 있다면, 이때 심적 상태의 소유자 S 는 현상적 속성과 기능적 속성을 동시에 가지는 심적 상태 M_1 을 가지는 것을 가리켜 'S 에게 현상성과 기능성이 동시 예화된 M_1 이 있다'고 말할 수 있다. 그리고 이 상황에서 특별히 S 의 현상성 혹은

---

[7] 차머스가 보기에 둘의 공외연적 관계가 결코 개념적 관계가 아니며, 다만 주어진 세계의 사실이다. (1996: Chapter 1) 물론 심리 세계에 대한 연구의 성과로 현상적 속성이 물리적 속성에 형이상학적으로 의존하며, 특정 기능적 속성이 물리적 속성에 형이상학적으로 의존한다는 것이 것이 밝혀진다면, 기능성과 현상성 사이에 공외연적 관계 사이에 형이상학적 관계가 성립함을 우리가 알게 될 수도 있을 것이다.

[8] 물론 〈고통〉의 개념적 분석에 따라서 〈고〉통이 순수한 현상적 개념이라면 고통의 주체는 고통 경험을 하는 것만으로 고통 상태에 있다고 이야기할 수 있을 것이다. 이는 직관과 개념적 분석에 달린 것으로, 여기서는 우선 〈고통〉이 복합적 개념이라고 가정하겠다.

기능성을 가리키고자 할 때, 같은 상황을 'S 에게 현상적 상태 M_1 이 있다 / S 는 현상적 상태 M_1 에 있다'나 'S 에게 기능적 상태 M_1 이 있다/S 는 기능적 상태 M_1 에 있다'고 나타낼 수 있다.

2.2.2. 자기지식의 대상적 구분

나의 마음에 관한 많은 사실들이 있지만, 이 글의 목표는 자기지식의 특유한 방식과 인식적 특권을 해명하는 것이다. 일인칭 특권이 주장될 수 있는 자기지식의 대상은 일차적으로 나의 심적 상태이다. 하지만 어떤 심적 상태는 일인칭 특권이 강하게 보장되는 것처럼 보이는 반면에, 일인칭 특권이 있는지 분명하지 않거나 혹은 일인칭 특권이 개념적으로 허용되지 않는 심적 상태도 있다.

순수한 현상적/감각적 상태

순수한 현상적/감각적 상태에 가까운 심적 상태일수록 주체는 그 심적 상태를 직접적인 방식으로 알 수 있으며, 따라서 일반적으로 주체에게 특유한 접근과 인식적 특권이 주어진다고 생각된다. 둥둥 울리는 듯한 청각 경험, 빨강을 보는 듯한 시각 경험, 찌르는 듯한 아픈 느낌과 같은 심적 상태가 바로 그러한 예시다. 순수한 현상적/감각적 상태에 대해 알기 위해서 주체는 현상적/감각적 판단의 근거가 되는 현상성(혹은 감각질)을 대면하면 되는 것처럼 보인다.[9]

지각 상태

지각 상태의 경우에는 순수한 현상적/감각적 상태보다는 주체에게 직접적으로 알려지지 않는 요소가 있는 것처럼 보인다. 예를 들어서 내가 스스로 눈 앞의 하얀 튤립을 보는 지각 상태에 있다고 믿더라도, 만일 눈

---

[9] 차머스는 현상성과 현상적 판단에 대한 이러한 근거 관계를 일종의 구성적 관계로 정식화한 바 있다. (2003: 234-241)

앞의 꽃이 하얀 튤립이 아니라면 이 지각 상태에 대한 믿음은 거짓 믿음이다. 이는 〈지각〉이 개념적으로 지각 상태가 성립되기 위해서 내용이 되는 외부 대상의 사물과 실제로 인과적 관계를 맺고 있음이 요구하기 때문이다. 하지만 여전히 주체는 지각 상태의 중요한 근거가 되는 감각적 경험에 대해 일인칭 특권을 갖고 있으며, 대상에 대해서 잘 작동하는 지각 능력과 인과적 체계를 갖춘 주체라면 자신이 무엇을 지각하는지에 대해서 신빙성있는 판단을 내릴 것이므로, 지각에 대한 판단에 대해서는 일반적으로 주체에게 특유한 접근과 인식적 특권이 있다고 여겨진다.

지향적 태도

믿음과 욕구 같은 지향적 태도에 대해, 많은 철학자들은 주체가 자신의 지향적 태도에 대해서 직접적으로 알 수 있으며 따라서 완전함에 가까운 인식적 특권이 성립하는 것을 당연하게 생각해왔다. 지향적 태도의 내용이 주체의 내적 상태에 수반하지 않으며 주체 외재적으로 결정된다는 심적 내용 외재주의가 많은 동의를 얻은 이후에도, 철학자들은 지향적 태도에 대한 직접적인 자기지식과 내용 외재주의를 둘 모두 포기할 수 없는 일종의 딜레마로 파악하였다.[10] 그리고 이후에도 많은 철학자들은 여전히 지향적 태도에 대해서도 직접적이고 특권적인 자기지식이 가능함을 설명하고자 하였다.[11]

성격

개념적으로 일인칭 특권이 허용되지 않는 심적 상태들도 있다. 나의 성격은 나의 심리가 갖는 성향에 대한 사실이며 심적 상태로 볼 수도 있지만 특별히 일인칭 특권이 보장되는 대상은 아니다. 예컨대 내가 가족에게 무신경한 성격을 가지고 있다거나, 무언가에 쉽게 몰두하는 성격을 가지고 있다고

---

[10] 예를 들어, Boghossian (1989)

[11] Burge (1988), Shoemaker (1994), Moran (2001)

해서 이 사실을 내가 남들과는 다른 방식으로, 더 잘 알고 있다고 생각할 수 있어 보이지는 않는다. 오히려 나의 성격에 대해서는 나를 보다 객관적으로 바라봐줄 수 있는 사람이 나보다 더 잘 알 수도 있다. 또는 나의 심적 상태 중에서는 개념적으로 복잡한 해석이나 주체 외부의 사실을 요구하는 것들도 있다. 예컨대 내가 누군가에게 배신감을 느낀다 것을 아는 것은 예컨대 어떤 화남의 느낌에 더해 누군가가 나를 배신했고 그 누군가가 누구인지에 대한 앎으로부터 일종의 추론과 해석을 필요로 한다. 자기기만적 상태 또한 주체의 다양한 심리적/행동적 총체에서 어떤 불일치가 일어나는지에 대한 고려를 통해서 포착될 수 있는 것이며, 따라서 자기기만도 그 자체로 일인칭 특권적으로, 직접적으로 알기 어려운 유형의 심적 상태인 것처럼 보인다.

이 글에서 고려되는 자기지식의 주된 대상은 세계의 대상을 내용으로 가지는 **지향적 태도**이다. 지향적 태도가 이론의 주된 대상이 되는 까닭은, 감각적/현상적 상태에 대해서는 일인칭 특유의 직접성과 비대칭적 특권은 비교적 분명해보이는 반면,[12] 내용을 갖는 지향적 태도의 경우 심적 내용 외재주의의 함축에 따라서 지향적 태도에 대한 일인칭 특권이 불분명해지기 때문이다. 또한 지향적 태도 중에서도 대부분의 철학자들이 지향적 태도에 대한 자기지식을 논하는 데 있어서 가장 중요한 범례로 삼는 **믿음**을 이 글에서도 가장 중심적인 제재로 삼는다. 이는 자신의 믿음을 직접적으로 아는 능력이 지향적 태도 중에서도 합리적 주체가 합리적일 수 있도록 만들어주는 가장 근본적인 능력이며 따라서 우리의 합리적 행위자성 전반과도 연결된 중요한 문제로 생각되기 때문이다.[13]

---

[12] 이러한 가정에 반대해서 윌리엄슨은 현상적 경험에 대해서조차도 우리가 선명하게 알 수 없다는 논변을 제시한 바 있다. Williams, T. (2000) 참고.

[13] Shoemaker (1994), Burge (1996), Moran (2001), Boyle (2009)

2.2.3. 자기지식의 방법적 구분

3 절과 4 절에서 자기지식을 획득하는 방법으로 내성과 투명성을 살펴보기에 앞서 이 절에서는 명료화를 위해서 자기지식의 방법적 구분을 제시하겠다. 주체가 자기지식을 얻는 방법으로 크게 다음의 네 가지의 구분을 생각해볼 수 있다.

행동적 증거로부터 추론함

타인에게 심적 태도를 귀속하기 위해서 사용할 수 있는 가장 보편적인 방법은 그 사람의 표정, 발화를 포함하는 행동적 증거로부터 추론하는 것이다. 행동적 증거로부터 추론하는 방법에는 일인칭/삼인칭 질적인 비대칭이 성립하지 않는다. 물론 나의 행동을 내가 가장 많이 지켜볼 수 있다는 점에서 일종의 양적인 비대칭이 성립할 수 있다고 볼 여지는 있으나, 정보의 풍부함 또한 나보다 나의 행동을 더 면밀히 관찰할 수 있는 타인이 현실적으로도 충분히 있을 수 있다는 점에서 이는 특유한 접근과 인식적 특권성의 근거가 될 수 없다.

한편 몸의 움직임으로서의 행동과 의도적 행위를 구분하고, 내가 어떤 행위를 하는지에 대해서는 근본적인 일인칭 특권을 가진다고 말할 수도 있다. 예를 들어 내가 차의 창문 밖으로 손을 흔드는 행동이 뒷차에 신호를 주는 행위인지 혹은 저린 팔을 푸는 행위인지의 여부를 행위자는 직접적으로 알 수 있다는 것이다. 이 경우 행위자에게 일종의 인식적 특권이 주어진다는 것은 맞지만, 이 사례 자체가 1 이 포함하고자 하는 행동적 증거로부터 추론되는 예시가 아니라고 답변할 수 있다. 이 경우 행위자가 자신의 행위를 구분할 수 있게 해주는 것은 몸의 움직임으로서의 행동에 대한 관찰적 앎 덕분이 아니라 행위를 할 의도와 동기적 이유에 대한 앎이며, 의도와 이유에 대한 앎은 여기서 말하는 행동적 증거가 아니기 때문이다. 결론적으로, 행동적 증거를 통해 심적 상태를 이론적으로 귀속시키는 것은 자기지식을 얻는 특유한 접근이거나 인식적으로 특권적인 방법이 아니다.

심리적 증거로부터 추론함

심리적 증거로부터 추론함이 행동적 증거로부터 추론함과 가지는 차이는, 일반적으로 심리적 증거에 대한 접근성은 주체에게만 주어지며 따라서 특유한 방법이라는 것이다. 추론을 위한 심리적 증거로서 가장 대표적으로 고려되는 자료는 개인의 내적 발화를 들 수 있다.[14] 내적 발화에 주목하는 철학자들에 따르면, 내적 발화는 내가 어떤 심적 태도를 갖고 있는지를 직접적으로 알려주지는 않지만, 내가 어떤 내적 발화를 하고 있는지 혹은 어떤 내적 발화를 하는 경향이 있는지에 대한 자료는 나의 심적 태도를 추론하는 근거가 될 수 있다.

심리적 증거로부터의 추론은 넓은 의미에서 일종의 내성 기능을 필요로하는 것처럼 보인다. 예컨대 내적 발화를 "듣기" 위해서 나는 내면의 소리에 "귀"기울이지만, 이때 귀를 기울여 듣는다는 것은 어디까지나 청각 능력에 대한 비유일 뿐이다. 하지만 철학적 논의에서 "내성"은 좁은 의미로는 마음에 대한 특정한 이론적 추론이 필요하지 않다는 의미에서 심적 상태를 **직접적으로** 아는 능력을 가리키며, 따라서 뒤에 이을 마음을 직접 들여다보는 방법과 구분될 수 있다.[15]

내성 (마음을 직접 들여다봄)

마음을 **직접** 들여다본다는 의미에서의 내성 능력은 2 의 방법을 제외하는, 어떠한 이론적 추론 없이 심적 상태에 대해서 알 수 있는 능력을 가리킨다. 전통적으로 많은 학자들은 나의 마음이 나에게 내성을 통해서 직접적으로 알려진다고 믿어 왔다. 한편 거틀러(2003)의 구분에 따르면, 내성은 크게

---

[14] Ryle (1949), Carruthers (2011), Cassam (2014)

[15] "내적 발화"에 주목한 라일, 카삼과 카루더스는 일반적으로 내성 이론가로 묶이지 않는다. 라일은 자기지식에 대한 행동주의자로 평가되며, 카삼(2014)은 스스로를 추론주의자로 표방한다. 카루더스(2011)는 자신을 해석이론가로 여기며 내적 감각 이론가들과 명확하게 구분한다.

"대면"과 "내적 감각"으로 구분할 수 있는데 이어지는 3 절에서 이 구분에 따라 대면과 내적 감각 이론, 그리고 각각에 대한 비판을 보다 상세히 살펴볼 것이다.

투명성 (세계를 내다 봄)

에반스는 우리가 스스로의 믿음을 알기 위해서 우리는 내면의 믿음을 찾아본다기보다는 믿음의 대상이 되는 세계의 대상에 직접 주목한다는 현상을 드러낸 바 있다.

> "믿음을 스스로에게 귀속하기 위해서, 그의 눈은, 말하자면, 혹은 때로는 문자 그대로, 세계를 향해 바깥을 향한다. 만일 누군가 나에게 "당신은 세계 삼차 대전이 일어날 것이라고 생각하나요?"라고 묻는다면, 나는 그에게 대답하면서 "세계 삼차 대전이 일어날까요?"라는 질문을 받았을 때 내가 주목할 바로 그 바깥의 현상(phenoemena)에 주목해야만 한다." (1982:225)[16]

에반스의 언급은 이후 여러 학자들에 의해서 심적 상태의 투명성(transparency)을 가리키는 것으로 이해되었다. [17] 주의할 것은 여기서의 투명성은 1 절에서 이야기했던 데카르트적 투명성과는 구분되는 개념이라는 것이다. 데카르트적 투명성은 오류불가능성, 자기-고시성, 교정불가능성으로 이해되는 완벽에 가까운 일인칭 특권을 의미했다. 즉, 데카르트적 투명성은 심적 상태의 인식을 가로막는 어떠한 불투명한 장애물이 존재하지 않는다는 의미를 가진다. 반면 에반스의 투명성은 심적 상태는 **인식 주체가** 그것을 투과해서 세계의 대상을 바라본다는 비유적

---

[16] "[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?'"

[17] Moran (2001), Gertler (2011), Byrne (2018), Boyle (2019)

의미에서 비롯하는 개념이다. 심적 상태는 나의 의식이 대상으로 나아갈 때 투과하는 투명한 유리와도 같으며 나는 나의 심적 상태를 심적 상태 자체를 바라보기보다는 인식의 대상이 되는 세계의 대상 자체를 바라 봄으로써 알게 된다. 이 글에서는 앞으로 '투명성'을 후자의 에반스의 투명성을 가리키는 것으로 사용할 것이며, 전자의 의미로 사용할 때는 '데카르트적 투명성'이라는 표현을 사용할 것이다.

자기지식을 얻는 이상의 네 가지 방법 중에서 나는 이 글의 초점을 내성과 투명성에 맞출 것이다. 이는 이 글의 주된 관심사가 지향적 태도에 대한 일인칭 귀속이 특유한 접근인지 그리고 어떻게 증거에 기반하지 않은 채 직접적으로 얻어질 수 있으며 따라서 특권적일 수 있는지를 설명하는 것이기 때문이다. 이제 이어지는 3 절에서는 내성 능력에 대한 소개와 비판을, 4 절에서는 투명성에 대한 소개와 비판을 다룬 후 2 장에서는 본격적으로 투명성 이론가들–번, 모란, 보일–의 이론을 다루겠다.

2.3. 자기지식의 방법 1. 내성

'내성(內省)'은 그 한자의 어원으로 알 수 있듯이 그 자체로 "안을 살펴본다"는 의미를 가진다. '내성'의 번역어 'introspection' 또한 어원적으로 "안을 바라본다"는 의미를 갖고 있다. 하지만 스스로의 심적 상태를 읽는 방식을 의미하는 것으로 일상적으로 사용되기에 전혀 문제 없어 보이는 이 표현이 가리키는 바가 무엇인지, 즉 내성의 본성이 무엇인지 그리고 구체적으로 어떤 과정으로 구성되는지를 해명하는 것은 생각만큼 쉽지 않다. 철학적으로도 내성이 정확히 어떤 능력인지 합의된 바는 없다. 하지만 일반적으로 철학적인 논의에서 내성은 자신의 심적 상태를 비-추론적인, 직접적인 방식으로 감지하는 능력이라고 이해된다.[18]

---

[18] Schwitzgebel, E. (2010)

내성의 능력을 설명하는 다양한 이론들이 있는데 크게는 대면 이론과 내적 감각 이론으로 나눌 수 있다.[19] 대면 이론은 직접적으로 의식하는 현상적 경험을 근거로 내재적으로 정당화되는 심적 상태에 대한 자기지식을 얻을 수 있다는 이론이다. 직접적으로 마주함이란 나의 의식과 현상성 사이에 어떠한 인과적 과정도 결여되었다는 것을 뜻한다. 대면 이론에 따르면 주체는 자신이 대면하고 있는 현상성에 주의를 기울이고 이에 근거하여 자신의 현상성에 대한 인식적으로 보증된 자기지식을 얻을 수 있다. 이때 이 믿음에 대한 일종의 인식적 보증은 단지 판단의 외재적인 신빙성에서 비롯하는 것이 아니라 주체가 의식하는 근거임에 비롯한다는 점에서 대면 이론은 자기지식에 대한 내재적 이론과 어울린다.[20]

반면 내적 감각 이론은 두뇌 안에 내적 감각을 위한 인과적 기능이 갖추어져 있고, 이 기능이 잘 작동만 한다면 심적 상태에 대한 앎을 신빙성 있는 방식으로 산출하므로 정당한 자기지식의 방법이 된다는 것을 이론의 핵심으로 한다.[21] 내적 감각 이론에 따르면, 마치 외부 대상에 대해 인과적

---

[19] 물론 이 구분에 대해서 카루더스와 같은 고차 이론가들은 반대할 것이다. 내적 감각 이론은 그 자체로는 고차 이론의 종류 중 하나로 여겨지며, 고차 이론가 중에서 내적 감각 이론보다는 고차 사고 이론을 선호하는 학자들이 있기 때문이다. 하지만 이 글은 내적 감각 이론, 혹은 소위 신빙성 있는 내성 이론 내에서의 논쟁을 다루지는 않을 것이다. 이는 내적 감각 이론에 대한 나의 비판은 그것이 함축하는 자기지식에 대한 신빙성 외재주의에 초점이 맞춰져 있으며, 이 비판은 내적 감각 이론가와 마찬가지로 고차 이론가들에게도 적용될 수 있는 비판이기 때문이다.

[20] 버지(1996: 68-87)는 일반적으로 '정당화(justification)'로 불리는 지식의 인식적 요건을 보다 넓은 인식적 보증(epistemic warrant) 개념으로 대체할 것을 제안한다. 이는 정당화가 일반적으로 명제 사이의 관계로 여겨지는데 반해, 버지가 보기에 어떤 논리적 추론이나 혹은 지각적 믿음은 이를 믿을 명제적 이유가 없음에도 인식적으로 그것을 믿는 것이 바람직할 수 있기 때문이다. 버지는 명제적 이유가 없음에도 인식적 바람직함이 보장되는 것을 '권리부여(entitlement)'로 부르고, 인식적 보증 개념을 정당화와 권리부여가 구성하는 것으로 규정한다.

[21] 하지만 많은 학자들은 이러한 특별한 "기관"이 있다는 주장에 개입하지는 않는다. 예컨대, 골드만(2006: 246)에 따르면 내적 감각은 우리의 주의(attention) 능력을 그 기관으로 한다.

과정을 거쳐 지각하는 것처럼 특정 내적 감각 인과적 메커니즘이 작동하여 인식 주체의 의식과 독립적으로 존재하는 심적 상태를 감각하고 자기지식을 내용으로 하는 새로운 표상을 산출한다. 이때 자기귀속 믿음이 지식이 되는 것에 내적 감각의 신빙성이 핵심이라는 점에서 내적 감각 이론은 자기지식에 대한 외재적 이론이라고 할 수 있다.

2.3.1. 대면 이론

대면은 주체가 일인칭 주체에게만 고유하게 주어지는 의식적 경험을 통해서 현상성을 마주하는 능력을 가리킨다. 주체는 일인칭의 시점에서 느껴지는 경험에 직접적으로 개념을 적용하고 현상적 판단을 내린다.[22] 어떤 철학자들은 현상성이 그 자체로 주체의 시점에서 경험하는 방식을 가리키는 속성이기 때문에 많은 경우 공적 언어로 형언할 수 없는(ineffable) 특성을 가지며 따라서 정당한 지식의 대상이 될 수 없다는 입장을 견지한다. 우리는 일상적으로 다양한 현상적 경험들을 유형적으로 구분하고 있으며, 만일 현상적 경험에 대한 개념 적용이 가능하지 않다면 우리가 당연하게 여기는 것들—예컨대 다양한 색조를 색상에 따라 구분하거나, 시간의 차이를 두고 경험된 후각 경험-토큰을 특정 대상에 대한 후각 경험으로 분류하는 능력 등—을 설명하기 어려워진다. 따라서 나는 여기서 현상적 경험에 대한 유형적 구분과 개념적 판단이 인식적으로 정당하게 이루어질 수 있다고 가정한다.

현상성을 대면적으로 경험하고 이를 근거로 내리는 판단이 해당 심적 상태의 본질을 포착하는 그러한 심적 개념들이 있다. 고통이나 감각질과 같은 현상적/감각적 상태들이 바로 그러한데, 고통은 그것을 인식하는 근거가 되는 현상성이 곧 고통 상태 자체의 본질이라는 점은 크립키의 기념비적인 강연 이후 이미 다수의 철학자들에 의해서 받아들여지고 있다.[23] 그리고 고통처럼 심적 속성의 본질이 그것의 현상성에 있는 현상적 개념들의

---

[22] Chalmers (2003), Gertler (2001)

[23] Kripke (1980, 152-3)

경우 현상성에 대한 대면이 곧 그 심적 상태의 본질을 마주하는 것이기 때문에 겉모습으로 인해 잘못된 식별을 하는 데서 발생하는 오류가 발생할 수 없으며, 따라서 강한 인식적 특권성이 보장되는 것처럼 보인다.

물론 이로부터 현상적/감각적 상태에 대한 판단이 오류불가능하다는 것이 따라나오지는 않는다. 예컨대 뜨거운 것이 살갗에 닿을 것을 예상하는 사람에게 차가운 얼음을 갖다 댔을 때 그는 이것이 뜨거운 느낌이라고 착각할 수 있고, 특정 현상적 경험의 경계적 사례들에 대해서 명확한 판단을 내리기 어려운 경우도 많이 있다. 하지만 여전히 현상적 경험에 대한 판단의 인식적 특권이 그것을 경험하는 주체의 주관에게 있다는 것은 우리의 강한 직관이다. 일상적으로 현상적 판단은 타인에 의해서 교정불가능한 것으로 여겨진다. 예를 들어서 타인의 입장에서는 이 사람이 전혀 고통을 느낄 법한 증거를 찾을 수 없음에도 불구하고 만일 주체가 고통을 호소한다면, 그때 일반적인 반응은 그의 고통을 인정하지 않는 것이 아니라, 무언가 그 원인을 찾지 못한 문제가 있다는 판단일 것이다. 또는 객관적 시점에서 소리가 들리지 않는 상황에서 환청을 호소하는 환자에 대해 우리는 그가 거짓말을 하고 있다고 생각하지 않고, 다만 그가 듣는 청각적 감각질은 그 자체로 인정하면서 무언가 이상이 있다고 판단할 것이다.

현상성에 대한 대면과 이에 근거한 자기지식은 현상적/감각적 상태에 대한 특유한 접근이면서 동시에 일인칭 특권적 방식이 될 수 있는 것처럼 보인다. 그렇다면 과연 대면이 내용을 가지는 지향적 태도에 대해서도 특유한 접근이면서도 일인칭 특권적인 방법이 될 수 있는가? 지향적 태도 중에 지각적인 믿음에 대해서는 대면이 일인칭 특권적인 자기지식 획득 방식이 될 수 있는 것처럼 보인다. 이는 지각적 믿음이 내용을 가지고 있지만 그 내용이 지각적 믿음이 근거하고 있는 현상성과 밀접한 관계를 갖고 있기 때문이다. 하늘 위의 하얀 구름이 떠있는 것처럼 느껴지는 그러한 감각 경험은 그 자체로 내가 하늘 위의 하얀 구름을 지각한다는 판단의 근거가 된다. 하지만 2 절에서 말했듯 지각 경험은 내용이 되는 세계의 대상과의 관계적 속성인 지향성을 만족할 때에야 그러한 내용을 가진 지각 경험이

된다. 즉, 지각 경험은 감각 경험의 내재적 성질인 현상성에 더하여 지각되는 대상과의 관계적 속성으로 구성되므로 감각 경험보다는 훨씬 오류가 발생할 여지가 크다. 그렇지만 여전히 지각 경험을 구성하는 감각 경험에 대한 접근이 일인칭 비대칭적이고, 또한 감각 경험에 대한 직접적인 의식으로부터 지각 경험이 얻어지므로 특유한 접근과 일인칭 특권이 성립한다.

하지만 지각적 상태 이외의 다른 지향적 태도에 대해서 어떻게 대면을 통해 자기지식을 얻는지는 분명하지 않다. 현상적 상태에 대해 알기 위해서는 그 상태의 현상성을 의식하는 것만으로 충분하지만, 심적 내용을 가지는 지향적 태도에 대해서 주체가 그것의 현상성을 통해서 지향적 태도를 충분히 정당하게 알 수 있는지는 명확하지 않기 때문이다. 지각적 내용이 아닌 심적 내용을 즉각적으로 알 수 있게 해주는 그러한 현상성이 있는지는 논란의 여지가 있지만, 21 세기 초 일군의 철학자들은 심적 내용의 지향성이 현상성에 형이상학적으로 의존한다는 현상적 지향성 논제를 제기하였다.[24] 현상적 지향성 이론에 따르면, **모든** 지향적 상태는 각각의 지향성에 특유한 현상성으로부터 구성되며, 따라서 나의 지향적 태도에 대해서도 우리는 지향적 태도를 의식하고 이로부터 특유한 현상성을 대면해서 해당 태도를 식별할 수 있다. 이처럼 현상적 지향성 이론은 인과적 관계항으로 포함될 수 있는 재현(representation) 개념을 통해서 현상성을 지향성으로 환원하고자 했던 기존의 재현주의자들에 대한 이론적 반발로서 야심차게 출발하였지만, 아직까지는 이제야 활발히 논의되기 시작하는 연구 프로그램이다.

현상적 지향성 이론이 구체적으로 어떤 근거로 지향성을 현상성으로 환원한다고 주장하는 것인지, 그리고 그러한 프로젝트가 얼마나 가망성이 있는지에 대해서 다루는 것은 이 글의 범위를 넘어선다. 다만 현상적 지향성 이론가들이 지어야 하는 한 가지 부담은 현상적 지향성은 어디까지나 나의 두뇌의 물리화학적 상태에 수반하는, 대면을 통해서 직접 마주할 수 있는 "좁은 내용"의 존재론을 받아들이기를 함축한다는 것이다. 이는 현상적

---

[24] Horgan and Tienson (2002), Pitt (2004), Kriegel (2013), Mendelovici (2018) 등.

지향성 이론가들에 따르면 현상성은 세계의 대상을 지향하는 역할을 맡고 있으며, 세계의 대상을 지향하는 것은 곧 세계의 대상을 내용으로 가진다는 것이므로, 결과적으로 현상성으로 구성되는 일명 "좁은 내용"이라는 개념이 따라나오기 때문이다. 하지만 철학자들은 많은 이유로 좁은 내용이라는 개념을 거부하며, 이는 현상적 지향성 이론에 대한 한 가지 중요한 문제점으로 제기된다.

한편 좁은 내용에 대한 개입보다도 지향적 태도에 대한 자기지식 모형에서 현상적 지향성 이론이 맞닥뜨리는 문제는 이 입장이 이어서 볼 내적 감각 이론과 마찬가지로 지향적 태도를 "발견"하는 자기지식 모형이라는 것이다. 현상적 지향성 이론가들이 일반적으로 주목하는 현상은 자신에게 떠오르는 생각을 어떻게 스스로 즉각적으로 알 수 있는가에 집중되어 있다. 하지만 자기지식에 대한 발견 모형을 선제하지 않는다면 머릿속에서 떠오르는 생각, 혹은 내적 발화에 대해서 아는 것과 자신의 지향적 태도를 아는 것은 분명히 구분할 수 있다. 다음 절에서 자세하게 살펴보겠지만, 합리적인 주체에게 스스로의 심적 상태는 단순히 대상으로서 발견되는 것이 아니라 주체가 사태 자체에 대해 어떤 판단을 내리는지와 긴밀하게 연관되어 있으며, 이를 고려한다면 발견 모형에 속하는 현상적 지향성 이론은 그 지향성 자체가 심적 상태 자체로 향해있다는 점에서 합리적 주체의 특권적 자기지식 모형으로 한계를 가진다. 이와 같은 발견 모형에 대한 비판은 이어지는 내적 감각 이론에 대한 비판에서 더 자세하게 다룰 것이다.

2.3.2. 내적 감각 이론

내적 감각 이론은 이론적으로는 의식에 대한 고차 의식 이론 중 하나로 분류될 수 있다.[25] 내적 감각 이론에 따르면, 나의 심적 상태는 나에게 마치 감각 대상에 대한 감각 기관의 작용을 통해서 얻어지는 것과 유사한

---

[25] Carruthers (2001)

방식으로 알려진다. 하지만 내적 감각 이론가들이 모두 눈이나 코처럼 특정한 감각을 수용하는 고유한 기관으로서 독립적인 내적 감각 기관이 있다는 주장을 하는 것은 아니다. 내적 감각 이론 내에서도 제각기 두뇌 속에 어떤 인과적 메커니즘(Nicols and Stich 2003)이 있다거나, 혹은 "주의(attention)를 기울임"이 내적 감각을 수행하는 역할을 한다(Goldman 2006)는 식의 다양한 설명이 존재한다.

내적 감각 이론 내에서 주장되는 다양한 인과적 메커니즘 사이의 차이를 비교하는 것은 이 글의 목적에서 벗어난다. 다만 내적 감각 이론에서 공통적으로 전제되고 있는 바는 다음과 같다. 첫째, 내적 감각 이론은 주체적 의식과 독립적으로 존재하는 심적 상태를 상정하고 있다. 둘째, 내적 감각 이론은 자기지식이 지식이 될 수 있는 인식론적 근거를 신빙성있는 인과적 메커니즘에서 찾는다. 믿음을 예로 들어보자. 내가 세계 삼차 대전이 벌어질 것이라는 믿음을 가지고 있음을 아는 것을 내적 감각 이론은 다음과 같이 설명한다. 나의 마음에는 지금 의식되고 있지 않은 〈세계 삼차 대전이 벌어짐〉에 대한 표상이 있다. 내적 감각은 이러한 〈세계 삼차 대전이 벌어짐〉에 대한 표상으로부터 특정 인과적 과정을 통해서 〈나는 세계 삼차 대전이 벌어질 것을 믿는다〉는 내용을 갖는 새로운 믿음을 형성한다. 그리고 이러한 인과적 메커니즘은 〈세계 삼차 대전이 벌어짐〉에 대한 표상이 있을 때에만 그러한 결과물을 산출하기 때문에 신빙성있는 메커니즘이며, 따라서 새로 만들어진 믿음은 지식으로서 충분하다.

슈메이커(1994)는 내적 감각 이론이 가정하는 위의 첫 번째 전제에 대한 비판을 제기한 바 있다. 슈메이커의 내적 감각 이론 비판의 핵심을 다음과 같이 정리해볼 수 있다:

(i) 내적 감각 이론은 주체적 의식과 독립적으로 존재하는 심적 상태를 함축한다.

(ii) 심적 상태가 주체적 의식과 독립적으로 존재한다면, 주체는 심적 상태에 대해서 의식하지 못하는 상태인 "자기맹(self-blindness)"의 상태에 놓이는 것이 가능하다.

(iii) 합리성을 가진 주체라면 자기맹의 상태에 놓이는 것이 가능하지 않다는 것은 개념적으로 참이다.

(iv) 만일 내적 감각 이론이 합리성을 가진 주체가 자기지식을 얻는 올바른 설명이라면, 합리성을 가진 주체도 자기맹의 상태에 놓일 수 있다.

(v) (iii)과 (iv)는 모순이다. 따라서 내적 감각 이론은 주체에 대한 적절한 내성 이론이 아니다.

논변을 위해서 슈메이커는 현상적 상태와 지향적 상태에 대하여 자기맹에 놓여 있는 행위자를 가정한다. 그리고 이러한 가상적 상황에서 행위자는 해당 심리적 상태를 가진 사람이라고 볼 수 없거나, 혹은 합리성을 갖지 못하였거나 둘 중의 하나로 귀결된다. 고통에 대해서 자기맹을 가진 자는 고통 상태를 의식하지 못하면서도 예컨대 구급함으로 달려가는 고통 반응을 보인다. 하지만 그는 고통 상태를 의식하지 못하기 때문에 구급함으로 달려가는 자신의 행동을 도무지 이해할 수 없다. 이 경우 우리가 주체에 대해서 내릴 수 있는 가능한 판단은 다음의 두 가지이다. 이 고통은 실은 그 행위자의 고통이 아닌 그의 몸을 점유한 다른 주체의 고통이며 따라서 이 사례는 자기맹의 사례가 아니다. 혹은, 만일 이것이 자기맹의 사례라면 행위자는 스스로의 행동을 도무지 이해할 수 없을만큼 일반적인 합리성을 결여하고 있다.

슈메이커에 따르면 지향적 태도에 대해서도 자기맹과 합리성의 양립불가능성이 성립한다. 합리성을 가지는 주체가 자신의 믿음, 욕구 전반에 대해서 자기맹의 상황에 놓여있다고 가정하자. 만일 그가 정말로 합리적인 사람이라면, 자신의 믿음들을 자신의 추론의 전제로 삼는 성향을 갖고, 그것이 참이라는 가정하에 행위할 이해 관계를 가지며, 타인에게도 일반적으로 자신이 그러한 믿음들을 믿는다고 믿게 만드는 방식으로 행위할 이해 관계를 가져야만 한다. 그러나 이러한 성향은 자신이 어떤 명제를 믿는다는 자기지식의 능력이 전제되어야만 가능하다. 하지만 그는 믿음, 욕구 전반에 대해서 자기맹의 상황에 놓여 있는 것으로 가정되므로 모순이

발생하며 따라서 지향적 태도에 대한 전반적인 자기맹과 합리성은 양립불가능하다.

슈메이커 논변의 핵심은 내적 감각이 자기맹의 가능성을 개념적으로 함축하기 때문에 합리적 행위자가 자기지식을 얻는 적절한 방법일 수 없다는 것이다. 하지만 나는 슈메이커의 반박이 내적 감각에 대한 결정적인 비판이 될 수는 없다고 생각한다. 이는 내적 감각이 자기맹의 가능성을 개념적으로 함축한다고 하더라도, 만일 내적 감각이 잘 작동하기 때문에 자기맹이 아닌 것이라면 슈메이커의 논변은 작동하지 않기 때문이다. 즉, 그의 논변은 잘 작동하는 내적 감각을 통해 주체가 자기지식을 얻는 덕분에 합리적 주체가 되는 가능성을 배제하지 못한다. 위의 논변에서 전제 3 은 개념적 참에 대한 진술이기 때문에, 여기서 '합리성을 가진 주체'는 언표 양상(de dicto)적으로 이해되어야 한다. 즉, 3 이 나타내는 바는 합리성을 가진 개별 주체는 합리성을 가진 주체이면서 *동시에* 자기맹일 수는 없다는 것이다. 하지만 만일 4 에서 '합리성을 가진 주체'는 사물 양상(de re)적으로 이해한다면 3 과 4 로부터 5 가 따라나오지는 않는다. 왜냐하면 이 진술에서 '합리성을 가진 주체'라는 표현으로 지시되는 주체가 다른 상황에서 합리성을 잃을 수 있는 가능성이 있으며, 이때 주체는 자기맹에 빠져있을 수 있기 때문이다. 만일 그렇다면, 인간은 필연적으로 혹은 개념적으로 언제나 합리적인 존재라는 추가적인 전제가 가정되고 있지 않은 이상, 슈메이커의 논변은 내적 감각이 잘 작동할 때 주체는 그 덕분에 자기맹에 빠지지 않고 따라서 합리적인 주체일 수 있을 가능성을 배제하지 못한다. 내적 감각을 통해 자기지식을 얻는 덕분에 합리성을 갖춘 어떤 주체가 내적 감각이 제대로 작동하지 않아 자기맹에 빠졌기 때문에 합리성을 잃는 상황에는 어떠한 개념적 모순도 존재하지 않기 때문이다.

나는 자기지식 능력이 주체의 합리성과 긴밀한 연관을 갖고 있다는 슈메이커의 통찰에는 동의한다. 슈메이커의 지적처럼, 자기지식은 이를 갖지 못한다는 사실로부터 주체의 합리성에 심각한 문제가 있음이 따라나온다는 점에서 단순히 주체 외부의 대상에 대한 지식과는 구별된다. 하지만 내적

감각 이론에 대한 슈메이커의 비판은 내적 감각 모형이 합리적 주체의 자기지식 모형으로서 정확히 어떤 측면에서 어긋나는지를 정확히 설명하고 있지는 않으며, 다만 내적 감각이 잘못 작동하는 상황의 경우가 곧 외연적으로 합리적 주체로 보기 어려운 경우임을 사고실험을 통해 보여주는 데 그치고 있다.

자기지식의 특권성을 설명하는 이론으로서 내적 감각 이론이 갖는 가장 큰 문제점은 내적 감각이 매우 잘 작동하는 어떤 상황에서도 내적 감각을 통해 갖게 된 믿음이 합리적 주체가 일인칭의 시점에서 얻은 자기지식에 이르지 못하는 경우가 생기기 때문이다. 내적 감각 이론의 기본적인 아이디어는, 우리의 심적 세계 내에 특정 인과적 메커니즘의 투입값이 될 수 있는 요소들로 구성된 심적 상태가 있으며, 이로부터 새로운 심적 상태인 이차 믿음이 산출되며 이 이차 믿음, 즉 자기지식은 바로 이 신빙성있는 과정 덕분에 지식의 지위를 얻는다는 것이다. 이 과정에서 주체의 시점에서 자기지식에 대한 내적 정당화는 요구되지 않기 때문에, 내적 감각 이론은 그러한 메커니즘이 산출한 믿음이 지식임에도 불구하고 그것이 왜 자신의 심적 상태에 대한 정당한 지식인지 이유를 전혀 알지 못하는 경우가 발생할 가능성을 함축한다.

이러한 가능성을 고려하기 위해 인식론적 외재론에 대해 제기되는 일명 "신통력(clairvoyance)" 반론을 자기지식에 적용해보자. 신빙성주의에 대한 신통력 반론은, 갖게 되는 믿음에 대한 어떠한 내적 정당화도 결여되어 있으면서 동시에 신통력을 갖는 사람은 과연 정당하게 지식을 가지고 있다고 여겨질 수 있는지에 대하여 직관적으로 그렇지 않다는 반론이다. 예컨대 미국의 대통령이 뉴욕에 있을 때마다 일종의 신통력을 통해서 미국의 대통령이 뉴욕에 있다는 사실을 알게 되는 노먼이라는 사람을 상상해보자. 그의 신통력은 매우 정확해서 외재주의자들이 이야기하는 신빙성 조건을 충족하지만, 우리가 그에게 미국의 대통령이 뉴욕에 있는지 어떻게 아는지 물어본다면 노먼은 어떠한 이유도 제시할 수 없을 것이다. 그리고 비록 그의

믿음이 신빙성있는 메커니즘의 산출이라고 해도 이 경우 우리는 그가 정말로 정당한 지식을 가지고 있는 것인지 의문스럽다. (Bonjour 1980) [26]

하지만 보통의 경우 신통력 반론이 정말로 지식에 대한 우리의 직관에 근본적으로 거스르는 것인지는 불분명하다. 여기서 우리의 직관에 거스르는 것은 노먼이 자신이 그렇게 믿을 정당화 이유를 아예 결여하고 있다는 부분이다. 노먼이 뉴욕에 대통령이 있다는 것을 믿을 이유를 아예 대지 못함에도 불구하고, 단지 그런 믿음이 신빙성있는 과정을 통해서 산출되었다는 사실로부터 그 믿음이 지식이 된다는 것은 이상하다는 것이다. 하지만 만일 노먼이 자신이 대통령이 뉴욕을 방문할 때마다 해당 믿음을 가지게 된다는 사실을 알게 되었고, 이를 이유로 자신의 믿음을 내적으로 정당화한다면 우리의 직관은 변하지 않는가? 귀납적으로 정당화되는 많은 믿음들은, 그것이 우리에게 일종의 "정당화를 해주는 관계 리스트"에 들어가기 위해서 결국에는 그것의 신빙성에 대한 확인이 선행되어야 한다는 것은 그럴 듯하다. 만일 그렇다면 같은 이유에서 노먼 또한 자신의 신통력이 신빙성있는 능력이라는 것을 확인하는 과정을 통해서 정당하게 신통력을 정당화 리스트에 넣고, 따라서 내적인 정당화 이유를 가진다면 결과적으로 노먼의 믿음이 지식일 수 없다는 직관은 흐려지게 된다. 설령 그러한 직관이 남아 있다고 하더라도, 이는 신통력의 존재에 대한 의문에 기초하고 있는 것일 뿐이다.[27]

하지만 신통력 반론에 대한 이러한 정당화 리스트 대응이 자기지식의 경우에도 성공적으로 적용될 수는 없어 보인다. 일차 심적 상태를 가질 때마다 신빙성있는 내적 인과 과정을 통해서 그에 대한 이차 믿음을 갖게

---

[26] 카삼(2014: 134-136) 또한 신통력 반론이 신빙성주의를 함축하는 내적 감각 이론에도 그대로 적용될 수 있다고 생각한다. 하지만 카삼이 생각하기에 내적 감각 이론가들이 "이를 악물고" 신빙성주의의 반직관적 함축을 받아들이는 것도 충분히 가능하다. 그보다 카삼이 주목하는 것은, 인식론적 외재주의 자체가 내적 감각이 "감각"과 유사한 메커니즘이라는 것에 너무나도 어울리지 않는다는 것이다.

[27] Fricker (2016)도 이와 유사한 방식을 신통력 반박에 대한 대응으로 제시한 바 있다.

되는 주체 S 를 가정해보자. 신통력 반론의 경우처럼, 그는 자신의 이차 믿음의 근거가 되는 심적 상태에 대한 의식 없이도 그러한 심적 상태를 자신이 갖고 있다는 믿음을 갖게 된다. 그리고 또한 그는 지금껏 자신의 경험에 비추어봤을 때 자신은 신빙성있는 내적 감각을 갖고 있다는 것을 이유로 자신의 이차 믿음을 정당화한다고 하자. 하지만 신빙성있는 능력을 정당하게 정당화의 리스트에 포함한다고 하더라도, 이렇게 얻은 자기지식은 적어도 합리적 주체가 직접적인 방식으로 얻는, 따라서 인식적 특권성을 가지는 자기지식이 될 수는 없어 보인다.

다음의 사례를 보자: 자신의 무의식적인 믿음에 대해서 신통력을 통해 자기지식을 형성하는 어떤 주체에게 예컨대 〈소금을 뿌리면 나쁜 일이 생기지 않는다〉는 무의식적인 일차 믿음이 있다고 해보자. 그는 자신이 그러한 믿음을 가지고 있다는 것을 의식하지는 못하고, 다만 신통력을 통해서 자신이 그러한 믿음을 가지고 있다는 것을 알게 될 뿐이다. 또한 그는 자신이 다른 일차 믿음들을 가질 때마다 그리고 그럴 때에만 이에 신빙성있게 반응하는 이차 믿음들을 가진다는 사실도 다른 경험들을 통해 알고 있다. 즉, 자신의 내적 감각이 자신의 믿음을 신빙성있게 발견한다는 것도 알고 있다. 따라서 그는 자신에게 신빙성있는 내적 감각이 있다는 사실을 자신의 정당화 리스트 목록에 넣어서 정당화 근거로 사용한다고 하자. 이 경우 이 주체에게 "소금을 뿌리면 나쁜 일이 생기지 않는다고 믿니?"라고 묻는다면, 그는 이러한 이유를 근거로 "응. 왜냐하면 나에게는 나의 믿음을 신빙성있게 감각해서 알게 되는 능력이 있는데, 나는 내가 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿는다는 믿음을 지금 갖고 있거든"과 같은 이유를 제시할 것이다.

이 주체의 사례는 주체가 자신의 일차 믿음을 일인칭 시점에서 의식하지 못해도 신빙성있는 메커니즘을 이유로 자기지식을 정당화하는 사례다. 하지만 다음의 논의를 고려하면 내적 감각이 아무리 신빙성있는 자기지식을 산출하고 또한 새로운 정당화 리스트에 따라 내적 정당화까지 성공하더라도, 그가 얻은 이차 믿음은 우리가 내성으로 얻기를 기대하는 합리적 주체의

일인칭 특권적 자기지식은 아닌 것처럼 보인다. 그의 심적 상태 자기 귀속 진술 앞에 해당하는 상태의 대상이 되는 사태에 대한 판단 진술을 연언으로 연결한 다음의 진술을 생각해보자: "소금을 뿌리면 나쁜 일이 생기지 않는다는 것은 거짓이야. 그리고(하지만) 나는 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿어." 이 진술 연언에서 앞의 진술은 세계의 사태에 대한 판단을, 뒤의 진술은 나의 일차 믿음에 대한 귀속을 진술하고 있다. 그리고 세계의 사태와, 세계의 사태에 대한 나의 믿음 사이에는 어떠한 논리적 관계도 없기 때문에 두 진술 사이에는 어떠한 논리적 모순도 없다. 그러나 세계의 사태에 대한 보고와 이 사태에 대한 나의 믿음에 대한 보고를 붙여놓았을 때, 우리는 서로 붙어있는 두 진술의 연언이 어떻게 합리적인 주체로부터 발화되었는지, 혹은 두 판단이 어떻게 동시에 합리적 주체로부터 내려지고 있는지를 정합적으로 이해하기 어려워지는 상황에 놓인다.

"지금 비가 오고 있어. 하지만 나는 비가 오지 않는다고 믿어", "지금 비가 오지 않고 있어. 하지만 나는 비가 온다고 믿어"와 같은 문장에서 느껴지는 이러한 모순에 가까운 긴장감은 무어가 처음 발견했다고 하여 소위 무어의 역설(Moore's Paradox)로 불린다. 무어의 역설은 'P 이다. 그리고 나는 P 를 믿지 않는다' 혹은 '-P 이다. 그리고 나는 P 를 믿는다'의 형식을 갖춘 진술에서 공통적으로 나타나는데, 무어의 역설을 어떻게 해석해야 하는지에 대해서는 다양한 견해가 있어 왔지만 지금의 논의에서 중요한 사실은 무어의 역설이 **일인칭 믿음 보고**의 경우에만 발생한다는 사실이다. 즉, 뒤의 진술이 믿음을 삼인칭 귀속하는 "소금을 뿌리면 나쁜 일이 생기지 않는다는 것은 거짓이야. 하지만 **철수**는 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿어"라는 진술에서는 일인칭 진술에서 발견된 유사-모순적인 긴장이 사라지는 것이다.

그렇다면 일인칭 사태 보고와 그에 대한 믿음 보고 사이에 어떤 관계가 무어의 역설이 가지는 유사-모순적 긴장을 만드는 것인가? 이 질문에 대한 힌트는 무어의 역설을 일인칭 주어를 유지하면서도 일관적으로 이해하고자 하는 한 가지 방식에서 찾아볼 수 있다. 무어의 역설에서의 두 번째 진술, 즉

나의 일차 믿음에 대한 보고가 만일 삼인칭적 증거로부터 뒷받침되는 진술이라고 생각해본다면 무어의 역설이 가지는 모순성을 어느정도 해소할 수 있는 것처럼 보인다. 즉, 세계의 사태를 고려할 때, 나는 소금을 뿌리면 나쁜 일이 생기지 않는다고 생각하지 않는다. 하지만 나쁜 일이 생기면 갑자기 소금을 뿌리고 싶은 갑작스러운 욕구를 느낀다거나, 혹은 실제로 소금을 뿌리는 행동을 하는 나 자신을 보면서 나는 삼인칭 시점에서의 이론 추론을 통해 나 스스로에게 소금을 뿌리면 나쁜 일이 생기지 않는다는 믿음을 귀속시켰다고 하자. 상황을 이와 같이 이해했을 때, 주어진 진술의 연접에서 느꼈던 유사-모순감은 상당히 사라지는 것 같다. 이 상황에서는 나는 내가 세계의 사태에 대해 의식하는 방식('소금을 뿌리면 나쁜 일이 생기지 않는다는 것은 거짓이야')과 별개로 나의 믿음에 대한 판단(나는 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿어')을 내리는 것이 충분히 개연적이기 때문이다. 만일 그렇다면 무어의 역설에서 유사-모순적 긴장을 만드는 것은 단지 진술의 주어가 일인칭 주어 '나'라는 사실에 있다기 보다는, 나의 일차 믿음에 대한 보고가 **일인칭 시점**에서 이루어질 때 자연스럽게 갖게 되는 주체의 태도의 초기값이 무엇인지에 달려있을 것이다.

그렇다면 우리는 도대체 어떤 방식으로 **일인칭 시점**에서 믿음에 대한 자기 귀속을 판단하기에 무어의 역설을 자연스럽게 역설*로서* 읽게 되는가? 이에 대한 나의 생각은 다음과 같다: 무어의 역설을 역설이 아닌 방식으로 이해할 수 있음에도 불구하고 자연스럽게 역설로서 이해하는 까닭은, 일인칭 시점에서 스스로의 믿음 귀속을 판단할 때 우리는 자연스럽게 믿음의 내용이 되는 세계의 사태가 성립하는지 여부, 믿음의 내용으로 고려되고 있는 명제가 참인지의 여부와 동기화되는 방식으로 스스로에게 믿음을 귀속하기 때문이다. 또한 이는 나의 믿음이 나의 시점에서 세계의 사태가 어떠한지에 대한 판단과 일치하는 방식으로 나 스스로에게 귀속되며, 세계의 사태의 성립에 대한 개입(혹은 명제의 참에 대한 개입)이 곧 내가 그러한 사태(혹은 명제)를 내용으로 가지는 믿음을 가짐에 다름 아니라는 것이 일인칭 시점에서 너무나도 당연하기 때문이다.

이제 내적 감각 이론의 신통력의 사고 실험으로 다시 되돌아가보자. 신통력있는 내적 감각을 가지고 있는 주체는 진심으로 "소금을 뿌린다고 나쁜 일이 안 생기지는 않아. 하지만 나는 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿어."라고 진술한다. 여기서 보통의 경우 "정당화 리스트" 재반론을 인정하여 신통력에 근거한 자기지식이 어찌됐건 믿음 귀속의 이유가 될 수 있다는 것도 받아들이자. 이 경우 그렇다면 우리는 그의 무어식 진술을 모순이 아닌 방식으로 이해해볼 수 있다. 즉, 그는 나쁜 일이 생기면 소금을 뿌려야 한다는 일차 믿음에 대해서 의식하지 못하기 때문에, 세계의 사태에 대해서 고려할 때 나쁜 일이 생기면 소금을 뿌려야 되는지에 대해서는 의문을 가진다. 그리고 신통력 덕분에, 그리고 신통력을 포함하는 정당화 리스트 덕분에 두 번째 진술 '나는 소금을 뿌리면 나쁜 일이 생기지 않는다고 믿어'는 자신의 믿음에 대한 정당한 지식을 나타낸다. 하지만 무어의 역설이 더이상 역설이 아니게끔 이해될 때, 그렇게 얻어진 이차 믿음은 더이상 우리가 원하는 "일인칭 시점에서" 얻은 믿음일 수 없다.

이는 이렇게 신통력을 통해서 얻어진 자기지식은 직접적인 방식으로 얻어졌다는 사실로부터 보증되는 일인칭 특권성을 부여받지 못하기 때문이다. 이 경우 이차 믿음은 다만 내가 신통력을 가지고 있다는 또 다른 증거로부터 추론을 통해서만 스스로에게 자기지식으로 여겨지는데, 이는 2 절의 구분에서 우리가 관심을 갖는 직접적인 방식으로서의 내성 조건을 충족하지 못하게 된다. 한편 외재적 신빙성 조건을 만족하는 방식으로 산출되는 자기지식이라는 점에서 내적 감각 이론의 핵심 요건을 만족하는 해당 주체의 사례가 일인칭 특권성을 갖지 못하는 자기지식만을 산출한다면, 일반적인 내적 감각 이론도 일인칭 특권적 자기지식을 설명하는 데에는 적절한 이론일 수 없다.

지금까지의 내적 감각 이론에 대한 비판에서 얻을 수 있는 핵심적인 교훈은 다음과 같다. 삼인칭 시점이 아닌, 일인칭 시점에서 얻어지는 자기지식의 특징은 그것이 주체의 시점에서 명제의 참 여부와 동기화되는 방식으로 얻어진다는 점에 있다. 자기지식의 본성이 이처럼 일인칭 시점에서

대상에 대한 판단에 긴밀하게 연결되는 것, 즉 동기화됨에 있음은 무어의 진술을 당연히 역설로서 이해하는 현상으로부터 알 수 있다. 하지만 내적 감각 이론은 본질적으로 나의 심적 상태를 대상으로서 발견하는 모형이기 때문에 언제든지 주체의 시점에서 보기에 해당 명제가 참인지의 여부와 불화하는 자기지식을 만들어낼 수 있다. 이는 자기지식을 발견 모형으로 설명하는 앞선 내성적 대면 이론 또한 마찬가지로 피하기 어려운 문제점이다. 따라서 내성적 대면이론, 그리고 내적 감각 이론은 합리적 주체가 자기지식을 일인칭 특권적으로 얻는 방식을 나타내는 적절한 자기지식 모형일 수 없다. 그리고 더 나아가 이러한 한계는 이미 마음에 존재하는 심적 상태를 그 자체로 "발견"하는 것으로 설명하는 어떠한 모형에 있어서도 적용될 수 있을 것이다.

2.4. 자기지식의 방법 2. 투명성

2 절에서 잠시 인용한 에반스의 언급으로 돌아가보자.

> "믿음을 스스로에게 귀속하기 위해서, 그의 눈은, 말하자면, 혹은 때로는 문자 그대로, 세계를 향해 바깥을 향한다. 만일 누군가 나에게 "당신은 세계 삼차 대전이 일어날 것이라고 생각하나요?"라고 묻는다면, 나는 그에게 대답하면서 "세계 삼차 대전이 일어날까요?"라는 질문을 받았을 때 내가 주목할 바로 그 바깥의 현상에 주목해야만 한다." (1982:225)

만일 에반스의 언급이 자기지식에 대해서 무언가 새로운 것을 말해준다면, 그것은 자기지식을 얻기 위해서 우리에게 필요한 것은 단지 나의 바깥 대상을 향할때 사용되는 일반적인 인지 능력만 있으면 된다는 것이다. 하지만 에반스는 이러한 진술을 통해 어떠한 새로운 자기지식 모형을 제시하고 있는 것은 아니며, 엄밀히 말해서 일인칭 믿음 귀속 질문에 주체는 "P 를 믿는가(P 라고 생각하는가)?"라는 질문을 "P 인가?"로 자연스럽게

이해한다는 현상을 기술하는 것 뿐이다. 다만 만일 세계를 내다 보는 방식이 나의 믿음을 아는 적절한 방식이라면, 세계를 내다 보는 것으로는 타인의 믿음을 알 수 없다는 점에서 이는 일인칭 특유의 방식이며, 동시에 일차 심적 상태 자체에 대한 증거에 기반하여 자기지식을 얻는 것이 아니라는 점에서 직접적으로 얻어지는, 따라서 특권적인 자기지식을 설명하는 새로운 방식이 될 수 있다.

심적 상태의 투명성을 통해서 자기지식을 얻는 과정을 '투명성 절차'라고 부르자. 투명성 절차란 자신이 무엇을 믿는지를 알기 위해서 그 믿음의 대상 자체를 생각해보는 절차이다. 간단히 말해서, 에반스의 언급에서 세계 삼차 대전이 발생할지를 내가 믿는지를 알기 위해서 세계 삼차 대전이 발생할지를 생각해보는 것이 투명성 절차다.

거틀러(2011)는 투명성 절차가 믿음 일반에 대해서 어떻게 자기지식을 얻는지에 대한 실질적인 설명을 줄 수 없다고 비판한다. 우선 그녀는 믿음 개념을 발생적 믿음, 암묵적 성향적 믿음, 일반적 성향적 믿음으로 세분하고 이들 모두가 투명성 절차로부터 얻어질 수 없다고 주장한다. 그녀의 주장을 요약하면 다음과 같다.[28]

첫째로, P 라는 믿음을 가지고 있는지 확인하기 위해서 "P 인가?"를 묻는 것을 투명성 절차라고 해보자. 투명성 절차는 믿음에 대한 일인칭 자기-귀속 질문이 주어지기 이전에 이미 갖고 있는 발생적인 자기지식 믿음을 설명해주는 것이 아니라, 해당 질문이 주어진 이후에 자기지식이 어떻게

---

[28] 논의를 위해서 거틀러의 믿음의 구분을 간단히 요약하자면 다음과 같다. 발생적 믿음이란 현재 나의 의식 앞에 발생하는 믿음이며 따라서 그 믿음에 대해 의식하고 있음을 함축한다. 반면 우리는 발생적 믿음 이외에도 우리의 심적 상태 내부에 잠재되어 있는 성향적 믿음을 갖고 있는데, 그 중 명제 p를 명시적으로 인지함으로써 가지고 있는 믿음은 명시적 성향적 믿음, 그리고 명시적으로 믿고 있지는 않지만 이미 갖고 있는 믿음들로부터 빠르게 동의할 수 있는 믿음을 암묵적 성향적 믿음으로 구분한다. 거틀러의 구분은 스탠포드 철학백과의 "Belief" 항목에서 슈비츠게벨이 소개하는 믿음에 대한 〈발생적/성향적〉, 〈명시적/암묵적〉 구분과 거의 같다.

형성되는지를 설명해준다. "당신은 세계 3 차대전이 벌어질 것이라고 믿는가?"에 대한 질문에 세계 3 차대전이 벌어질 이유들을 고려하고 답하는 에반스의 사례는 엄밀히 말해서 새로운 믿음을 형성한 것이지 이미 있는 믿음 귀속 질문에 답하는 것은 아니라는 것이다.

둘째로, 암묵적 성향적 믿음에 대해 주체는 해당 믿음을 가지고 있는지를 판단하기 위해서 대상이 되는 바깥 세계를 바라보는 것이 아니라 그 증거를 찾기 위해서 자신의 내부를 바라본다. 투명성 절차가 시작되는 시점 이후에 추가적으로 증거를 획득하고 대답하는 경우를 믿음의 적절한 자기귀속 예시로 인정하는 것은 성향적 믿음 p 를 가지는 것과 p 를 믿을 성향을 가지는 것을 혼동하는 것이다(2011: 8).

셋째로, 성향적 믿음 일반에 대하여 투명성 절차로 적절하게 믿음을 자기-귀속 시킬 수 없는 사례가 많이 있다. 거틀러는 우선 각각이 주체 S 에게 성향적인 믿음 p 를 귀속시키는 충분한 조건으로 생각되는 것들을 다음과 같이 대략적으로 제시한다:

(조건 A) S는 명제 p를 옹호하는 성향을 가진다;

(조건 B) S는 명제p를 믿는 사람으로서 할만한 행동들을 하는 성향을 가진다.

하지만 조건 A 에는 해당하지 않지만 조건 B 에 해당하는 믿음에 대해 S 는 투명성 절차를 통해서는 자기지식을 얻을 수 없다. 이 경우 주체는 p를 믿을 증거를 인지하지 못하므로 자신이 p 를 믿는다는 것을 관찰적 행동적 증거에 기반해 알 수 있을 뿐이기 때문이다. 그리고 이 사례들은 p 를 믿는 것처럼 행동하지만 실제로는 p 를 믿지 않는 예외적인 사례가 아니라, 불완전한 인간으로서 겪게 되는 일상적인 예시이다. (2011: 10-13)

나는 기존의 투명성 절차가 믿음에 대한 형성과 이미 있는 믿음에 대한 발견을 혼동하는 경향이 있다는 거틀러의 비판에 동의한다. 비록 암묵적 믿음을 믿음의 정당한 외연으로 인정한다고 하더라도, 믿음의 과대생산을 막기 위해서는 최소한 p 를 믿을 성향과 성향적 믿음 p 를 구분하기 위한 기준이 필요하다. S 가 창 근처에 있어서 지금 비가 오는지를 확인할 수 있는

상황에 놓여있지만 실제로는 확인하지 않은 상황이라면, 만일 누군가 비가 오는지 물어본다면 쉽게 창 밖을 바라볼 수 있는 상황에 있으므로 S 에게는 비가 온다는 것을 실제로 믿을 성향이 있다. 그러나 이 경우 S 에게 비가 온다는 믿음을 귀속시키는 것은 믿음을 과대생산하는 결과를 낳게 된다. 위의 상황에서 S 는 모르게 밖을 볼 수 없도록 커튼을 쳤다가, 잠시후 또다시 밖을 볼 수 있도록 창문을 열었다고 하자. 만일 믿을 성향도 성향적 믿음에 포함된다면, S 에게는 어떠한 내재적 변화없이 지금 비가 온다'는 믿음이 사라졌다가 생긴다는 받아들이기 어려운 귀결이 따라 나온다. 그보다 상식적인 생각은 창문이 열리고 닫힘과는 상관없이 S 에게는 지금 비가 온다는 성향적 믿음이 애시당초 없다는 것이다. 나는 믿음의 과대생산이 포함하는 이와 같은 이상한 귀결을 피하기 위해서 해당 시점에서 행위자 자신이 가지고 있는 증거만을 가지고 믿음을 형성하는 것으로 기준을 잡는 거틀러의 제안이 타당할 수 있는 가장 너그러운 믿음의 범위라고 생각한다.

하지만 해당 순간에 주체가 가지고 있는 증거들만으로 믿음 귀속의 여부를 판단해야 한다는 주장이 기존의 투명성 이론가들의 주장과 양립하기 어렵다는 거틀러의 비판(2011: 10)은 옳지 않다. 투명성 이론가들이 말하고자 하는 바는, 명제 '나는 p 를 믿는다'의 참 거짓을 판단하기 위해서 나의 내부에 있는 p 에 대한 믿음 상태를 주목하는 것이 아니라 사태 p 에 대한 증거에 주목한다는 것이기 때문이다. 투명성 이론에서 '내적(inward)인 고려'는 **심적 상태 자체**를 고려함을, '외적(outward)인 고려'는 **비-심적 사태**를 고려함[29]을 의미한다. 따라서 설령 주체가 믿음에 대한 자기지식을 얻기 위해서 이미 가지고 있는 내적인 자료가 사용된다고 하더라도, 만일 이것이 심적 상태 자체가 아니라 심적 상태의 대상을 지향하는 과정이라면 투명성 이론은 여전히 자기지식 획득 절차에 대한 적절한 이론이 될 수 있다.

---

[29] 물론 일차 믿음의 대상이 심적 상태인 경우, 심적 상태가 외적인 고려의 대상이 될 수 있다.

성향적 믿음 일반에 투명성 절차를 적용할 수 없는 거틀러의 사례들도 자기지식 획득에서의 투명성 절차의 역할을 논파하기에 불충분하다. 사실 거틀러가 제시하는 예시들은 모란이 *Authority and Estrangement* (2001)에서 제시하는 전형적인 자기소외의 예시들이다. 물론 어떤 믿음을 가진 것처럼 행동하는 것 또한 *믿음*의 기능성을 예화하고 있는 것으로서 믿음의 사례라고 생각될 수도 있으며, 이처럼 믿음의 현상성을 대면하진 못하지만 기능성을 예화함으로써 발생하는 자기소외의 상황은 거틀러의 주장처럼 흔하게 발생하는 사례일 수도 있다. 그러나 우리가 일상적으로 사용하는 믿음 개념이 애매한 개념인 것과 별개로, 인식 주체의 관점에서 무어의 역설을 역설로서 만드는 방식으로 믿음이 나에게 인지되는 방식과 그러한 증거와 상관없이 행동 성향적 증거로부터 이론적 추론을 통해 믿음이 나에게 인지되는 방식은 철학적으로 분명한 기준을 가지고 구분할 수 있다.[30]

요약하자면, 거틀러는 해당 순간에 투명성 절차를 통해 믿음을 형성하는 것과 이미 있는 믿음을 발견하는 것을 구분해야 한다는 것을 적절하게 지적했다. 만일 믿음의 형성을 이미 있는 믿음의 발견과 구분하지 않는다면, 우리는 성향적 믿음과 믿을 성향을 잘 구분하지 못하게 되며 믿음의 과대생산 문제를 맞닥뜨리게 된다. 그러나 믿음의 형성과 발견을 구분하더라도, 거틀러의 비판은 인식 주체의 일인칭 시점에서 직접적인 방식으로 믿음을 인지하는 투명성 방식과 행동 성향적 증거를 토대로 이론적 추론을 통해 인지하는 방식 사이의 유의미한 구분 자체를 허물 수는 없다.

---

[30] 비슷한 문제 의식에서 겐들러(Tamar Gendler)는 인식적 증거를 바탕으로 형성되는 믿음과 우리의 신체에 보다 원초적인 반응으로서 내재되어 있는 심적 상태 '알리프(alief)'를 구분하자고 제안한다("Alief and Belief", 2008). 하지만 거틀러는 겐들러의 구분이 믿음을 비개연적으로 합리적인 개념으로 만들 것이라는 이유로 이 제안을 거부한다(2011: 23). 그러나 인식적 증거를 바탕으로 형성된 믿음과 기능적 역할을 가리키는 믿음의 믿음 개념 내에서의 구분은 거틀러도 암묵적으로 하고 있으며(2011 :10) 이 글은 믿음 개념을 개념적으로 둘로 나누어야 한다는 것까지 개입하지는 않으므로 거틀러의 입장에서도 믿음 내의 구분은 부당하지 않다.

따라서 이어지는 3장에서는 심적 상태의 투명성을 일인칭 특권적 자기지식을 설명하는 방식으로 고려하는 학자들의 입장을 보다 자세히 살펴볼 것이다.

제 3 장 투명성과 자기지식

3.1. 심적 상태의 투명성 이해하기

일인칭 주체의 시점에서 자기지식을 얻는 방법의 핵심으로 심적 상태의 투명성에 주목한 세 철학자–번, 모란, 보일–을 살펴보기에 앞서, 1 절에서는 에반스의 "투명성"을 이해하기 위해 도움이 될 몇 가지 논점을 살펴볼 것이다. 이는 첫째로 에반스의 투명성과 데카르트적 투명성 사이의 혼동을 줄이기 위함이다. 비록 일반적으로 에반스의 투명성이 마음의 일인칭 특권의 핵심으로 지목되는 데카르트적 투명성과 언어 표현 '투명성'을 공유하고 있지만 실상은 완전히 구분되는 마음의 특성을 가리키기 때문이다. 투명성 전반에 대한 논점을 살펴보는 두 번째 이유는 에반스의 언급에 주목하는 학자들 사이에 구체적으로 투명성이 어떤 특성인지 그리고 이와 자기지식 사이의 관계를 어떻게 해명해야하는지 대한 일치된 견해가 없으므로 논의의 공통 지형을 분명히 할 필요가 있기 때문이다. 비록 앞으로 논할 투명성 이론가들이 에반스(1982: 225)를 마음의 투명성에 가장 앞서 주목한 사람 중 하나로 언급하지만, 에반스 스스로는 이러한 마음의 일인칭 상대적 특성을 명시적으로 '투명성'으로 부른 바가 없을 뿐 아니라 무엇보다 그 스스로 마음의 일인칭 상대적 특성의 본성에 대해서 별달리 설명한 바가 없다. 따라서 에반스에 공통적으로 주목하더라도 그들은 모두 제각기 다른 의미로 투명성을 사용하며 그 설명 방식도 모두 다르다.

심적 상태의 자기-고시성, 심적 판단에 대한 오류불가능성과 교정불가능성을 포괄하는 데카르트적 투명성에서 '투명성'은 나의 인식의 대상인 심적 상태와 나 사이에 어떠한 **불투명한** 장애물이 없어서, 그 대상이 나에게 있는 그대로 보인다는 특성을 가리킨다. 반면 에반스의 '투명성'은 심적 상태 자체에 적용되는 독특한 성질로서, 나의 심적 상태는 **그 자체로** 그것을 통과해 바깥의 대상을 지향하게 해주는 통로로서 바깥의 대상을

인식하게 만들어주는 일종의 투명한 유리의 역할을 한다는 심적 상태의 특성을 나타낸다. 이 비유적 표현은 이미 '투명성'이 데카르트적 의미에서 사용되어온 오랜 역사를 갖기 때문에 그렇게 적절해 보이지 않을 수는 있다. 하지만 그것을 통해야만 밖을 볼 수 있지만 너무나도 투명해서 내 앞의 대상들과 구분될 수 없는 유리가 있다고 할 때 나의 초점은 그 유리가 아니라 유리 너머의 대상에 닿을 수밖에 없음을 생각해보면, 우리의 믿음도 그 자체가 세계의 사태를 들여오는 투명한 통로이기 때문에 믿음 자체에 자기 인식의 초점을 맞출 수 없다는 것을 나타내고자 하는 의도에는 충분히 부합한다.

한 가지 제기될 수 있는 의문은 투명성을 가지는 믿음의 범위에 대한 것이다. 지각적 경험이 투명하다는 것은 에반스(1982) 이외에도 많은 학자들에게 주장된 바 있다(Harman 1990, Dretske 1995, Tye 2000). 그들은 주체가 *지각적* 경험을 인식할 때 지각적 경험의 내재적인 특성에 주목하는 것이 아니라, 지각적 경험의 대상이 되는 바깥 대상의 특성에 주의를 기울인다는 것에 주목한다. 빨간 장미의 빨강을 지각할 때, 그 지각 경험을 알기 위해서 주체가 주목하는 것은 빨간 장미에게 있는 빨강 속성이지, 그외에 추가적으로 인지되는 스스로의 경험의 내재적 속성은 없다는 것이다. 그들은 P 에 대한 지각적 상태를 알기 위해서 P 이외에 경험의 내재적 속성을 들여다 볼 필요가 없다는 점에 동의한다는 점에서 지각적 믿음의 투명성에 개입한다. 반면 그들은 지각적 믿음 이외의 믿음에 대해서는 명시적으로 투명성에 개입하지는 않는 것으로 보인다. 이는 그들의 관심사가 현상성을 인과적으로 분석할 수 있는 표상 개념으로 설명하기 위한 표상주의(representationalism)의 이론적 배경 하에서 현상성을 마음의 신비한 속성이 아니라 외부 대상이 가지는 객관적 속성으로 설명하는 데 있었기 때문일 것이다.

에반스도 지각적 상태의 투명성을 통한 자기귀속이 세계 삼차 대전에 대한 믿음의 자기귀속과는 다른 방식으로 이해될 수 있는 또 다른 모형을 따르는 것으로 제시한다(1982: 226-230). 그는 지각 상태를 구성하는 것으로 비-

개념적인 내적 정보적 상태(internal informational state)와 정보적 상태에 근거하여 개념을 적용한 인지적 상태인 지각적 판단을 구분한다. 중요한 것은 내적 정보적 상태에 근거하여 얻은 지각적 판단은 그 자체로 내적 정보적 상태에 **대한** 것은 아니라는 것이다. 이는 주체가 지각적 판단을 식별하기 위해 내적 정보적 상태를 살피는 것이 아니라 다시금 바깥 대상에 주목한다는 것에서 알 수 있다. 이처럼 내적 정보적 상태는 주체에게 그 자체로는 인식의 대상이 아니며 다만 주체가 지각적 판단을 할 수 있게 하는, 지각적 상태를 구성하는 상태이다. 물론 내적 정보적 상태에 대해서 어떠한 외부적(extraneous) 지식을 배제하는 방식으로, '나에게는 ~하게 보인다'는 접두어를 사용하여 자신의 정보적 상태에 체계적으로 의존하는 내용을 갖는 지식을 얻을 수 있다. 하지만 이 경우에도 내적 정보적 상태는 나의 인지 *대상*이 아니다. 바깥의 대상에 대해 인지할 때 정보적 상태가 그 대상과 맺는 관계가 정보적 상태 그 자체에 대해서는 전혀 주어지지 않기 때문이다. 즉, 지각적 상태를 구성하는 내적 정보적 상태는 그 자체로는 지각의 대상이 아니라, 바깥의 대상에 대한 지각적 판단을 가능하게 하는 구성요소이다. 그렇기 때문에 예컨대 나무를 본다는 지각 상태를 주체가 스스로 귀속할 때 그는 그의 내적인 정보적 상태를 대상으로 내성하는 것이 아니라, 다만 그 내적인 정보적 상태가 지각할 수 있게 만들어주는 구체적 대상인 나무에 주의를 기울인다.

반면 세계 삼차 대전의 사례에서 자기지식의 대상인 일차 믿음의 대상은 세계 삼차 대전이라는 미래의 가능한 사태이며, 미래의 가능한 사태는 빨간 장미처럼 나에게 직접적으로 지각될 수 있는 대상이 아니다. 빨간 장미가 있다는 지각적 믿음을 갖고 있는지 알기 위해 주체는 다시금 바깥의 대상인 빨간 장미를 *바라본다*고 할 때, 이때 바라봄 관계는 대상과 감각 주체 사이에 존재하는 인과적인 관계로 손쉽게 분석될 수 있을 듯하다. 반면 세계 삼차 대전이 일어난다는 믿음을 갖고 있는지 알기 위해서 세계 삼차 대전이라는 미래의 사태를 *바라보는 것*이 세계 삼차 대전과 문자 그대로 바라봄의 인과적 관계를 맺는 것일 수는 없으며, 다만 세계 삼차 대전의

사태를 *고려한다*거나 혹은 세계 삼차 대전에 대해 *생각한다*는 것이므로 일종의 대상에 대한 지향적 관계를 맺고 있는 것이다. 따라서 지각적 믿음 이외의 믿음이 갖는 투명성에 대해서 다음과 같은 의문이 제기될 수 있다. 지각적 믿음, 예컨대 여기 빨간 장미가 있다는 믿음에 대해서 빨간 장미가 있다고 믿는지 묻는다면 주체는 빨간 장미에 대한 믿음을 내성하는 것이 아니라, 바깥의 빨간 장미를 (문자 그대로) 바라본다. 반면 세계 삼차 대전이 일어날 것을 믿는지 묻는다면 주체는 세계 삼차 대전이 일어날 것이라는 믿음을 내성하는 것이 아니라, 세계 삼차 대전이 일어날지를 고려한다. 하지만 이때 나와 미래의 사태 사이에 성립하는 고려함-관계는 무엇인가? 미래의 세계 삼차 대전은 구체적 대상이 아니라 추상적 대상으로 이해되어야 할 것 같고, 이 고려함-관계는 자연스럽게 추상적 대상에 대한 지향함의 관계로 이해될 수 있을 텐데, 그렇다면 바깥의 구체적 대상과의 관계가 아니라 내 심적 세계 내부의 추상적 대상에 대한 지향성에 대한 의식을 필요로 하며, 따라서 내성의 능력을 사용하는 것은 아닌가?

지향적 대상의 본성이 무엇인지, 지향적 대상과 지향의 주체가 맺는 관계가 무엇인지는 그 자체로 심원한 탐구를 요하는 주제다. 그러나 지각적 믿음과 그렇지 않은 믿음 사이의 경우 인식 주체가 대상과 맺는 관계의 유형이 다르다고 해서 이것이 지각적 믿음을 제외한 믿음들의 투명성에 의문을 제기할 근거가 되지는 않는다. 지각적 경험에 근거한 믿음과 그렇지 않은 믿음에 대한 앎은 어디까지나 나의 내적 상태가 아니라 바깥 대상에 대한 지향을 통해서 얻어지게 된다는 점에서 주체의 시점에서 P 를 믿는지 알기 위해 P 를 살펴보는 투명성 현상을 설명하는 데 있어서 본질적인 차이를 갖지 않기 때문이다. 심적 상태의 투명성을 통해 자기지식을 획득하기 위해서 주체에게는 세계의 사태에 대해 지향할 것만이 요구될 뿐, 자신이 무엇을 지향하며 그 지향의 대상이 무엇인지를 또다시 일종의 내성을 통해서 발견하는 것까지 요구되지는 않는다. 세계 삼차 대전이 일어날 것을 믿는지 질문을 받은 주체는 자연스럽게 세계 삼차 대전이 일어날지를 생각해보고 그렇다고 생각한다. 그리고 그 결과 세계 삼차 대전이 일어날

것이며 자신이 세계 삼차 대전이 일어날 것이라고 믿는다고 믿는다. 이 과정에서 주체가 자신이 생각하는 것이 세계 삼차 대전의 발생인지 그리고 자신이 세계 삼차 대전에 대해서 생각하고 있는지에 대해서는 내성을 통해서 다시금 발견할 필요는 없다. 다만 주체는 세계 삼차 대전의 발생이 심적 태도의 자기기귀속을 위해 고려할 사태라는 것을 알고, 세계의 사태에 대해 생각하는 일반적인 인지 능력의 사용을 통해서 자신에게 믿음을 귀속한다. 비록 세계의 사태와 맺는 지향적 관계가 어떻게 이해되어야 하는지는 지각 상태에서의 인과적 관계의 경우처럼 분명하지는 않지만, 그렇다고 해서 이 과정에서 나의 심적 상태를 대상으로 하여 직접 살펴보는 내성이 필요하지 않음은 분명하다.

마지막으로 P 에 대한 믿음을 자기귀속하기 위해서 주체가 스스로 P 가 참인지를 물어보는 절차를 투명성 절차라고 할 때, 투명성 절차의 메커니즘은 크게 두 단계로 이루어져 있다는 것을 분명히 할 필요가 있다. 요컨대 투명성 절차는 크게 세계의 사태가 성립한다(P)는 판단에 도달하는 첫 번째 단계와, 이로부터 '나는 P 를 믿는다'는 자기지식에 도달하는 두 번째 단계로 이루어져 있다. 이 구분이 중요한 이유는 앞으로 살펴 볼 모란, 번, 보일이 중심적인 설명의 대상으로 두는 단계가 각각 다르기 때문이다. 자기지식을 얻는 과정에서 모란(2001)에게 핵심적 역할을 하는 숙고(deliberation)와 결의(resolution)는 P 가 참일 이유들을 살펴봄으로써 P 라는 결론에 도달하는 과정, 즉 첫 번째 과정을 설명한다. 반면 번(2018)의 투명성 추론은 두 번째 단계를 일종의 신빙성있는 추론으로 설명하는 것에 거의 모든 초점이 맞추어져 있으며 다만 그러한 추론의 전제가 되는 판단에 어떻게 도달하는지, 즉 첫 번째 단계에 대해서는 아무런 설명을 제공하고 있지 않다. 보일(2019)의 "반성"도 P 가 참이라는 결론에 도달하는 과정이 아니라, 'P 이다'로부터 '나는 P 를 믿는다'는 자기지식의 결론에 도달하는 과정으로 제시되고 있다. 앞으로의 논의에서 나는 세계의 사태에 대한 판단에 도달하는 첫 번째 단계를 해명하는 문제를 '**판단의 메커니즘의 문제**'로, 'P'로부터 '나는 P 를 믿는다'로 이행하는 두 번째 단계를 해명하는

문제를 번(2018: 4, 77 등)과 보일(2019: 1014) 등을 따라 '**투명성의 퍼즐**'로 부르겠다. 특히 두 번째 단계가 '투명성의 퍼즐'인 까닭은, 어떻게 세계의 사태를 나타내는 'P'와 나의 심적 상태를 나타내는 '나는 P 를 믿는다' 사이에 증거 관계가 없음에도 불구하고 일인칭 시점에서는 이러한 이행으로 얻어진 믿음 귀속 문장이 인식적으로 정당한지에 대한 합의된 설명이 없기 때문이다. 이어지는 2 절부터 투명성 이론가들이 각각의 퍼즐에 대해 어떠한 대답을 내어놓고 있는지를 살펴볼 것이다.

3.2. 번의 투명성 추론주의 이론

3.2.1. 투명성과 행위자 중립적 추론

번(2018)은 인식 주체가 에반스가 말하는 심적 상태의 투명성을 통해 특유한 접근과 인식적 특권성을 설명한다. 그의 이론은 투명성 절차가 자기지식 획득을 위한 일인칭 고유의 방식이면서 동시에 인식적 정당성을 부여한다고 본다는 점에서 3 절에서 볼 모란의 입장과 유사하다. 하지만 모란과 달리 번은 자기지식을 획득하는 과정에서 **추론**이 필요함을 명시적으로 밝힌다. 또한 자기지식을 설명하기 위해서 합리적 행위자성을 가정하지 않았으며 자신의 추론주의가 합리적 행위자성의 발현을 필요로 하지 않는 다른 모든 심적 상태들에 대한 자기지식을 일률적으로 설명하는 이론이라고 주장한다.

번에 따르면 주체는 심적 상태를 포함하지 않는다는 의미에서 **행위자 중립적**(agent-neutral)으로 세계의 사태를 표현하는 명제로부터 자신의 믿음 사태를 표현하는 명제로의 추론을 통해 자기지식을 얻는다. 범례가 되는 것은 역시 믿음이다. 그에 따르면 주체는 일차 믿음의 대상이 되는 'P'가 유일한 전제이고 '나는 P 를 믿는다'를 결론인 다음의 믿음 도식을 통해 믿음에 대한 자기지식을 얻는다.

P
———
그러므로, 나는 P 를 믿는다

물론 번은 'P'와 '나는 P 를 믿는다' 사이에는 어떠한 내용적인 증거 관계도 성립하지 않는다는 것을 인정한다. 둘 사이에는 어떠한 논리적 관계도 없으며, 형이상학적으로도 내가 전지(omniscient)에 가까운 존재가 아닌 이상 각각의 사태 P 와 나의 P 를 믿음 사이에는 충분히 강한 귀납적 관계가 존재하지 않기 때문이다. 그는 논란의 여지가 있는 도식 형식을 사용하는 대신에 자기지식을 얻게 만들어주는 일종의 추론 규칙 BEL 을 제안한다.

BEL 만일 P 라면, 당신이 P 를 믿는다는 것을 믿어라.

번은 BEL 이 다음과 같은 의미에서 "강하게 자기 확정적(strongly self-verifying)"이라고 주장한다: 만일 믿음 도식을 앞에 둔 주체가 전제를 받아들이고, 따라서 믿음 도식을 따라 추론을 진행한다면 주체는 반드시 믿음에 대하여 참인 자기지식에 도달하며 이는 심지어 전제의 참 거짓과 상관없이 그러하다. 따라서 만일 이 도식을 따라 자기지식을 얻는다면 자신의 믿음 도식이 매우 안전한 도식이며, 따라서 이렇게 얻어진 지식인 쉽게 거짓이 될 수 없다는 점에서 안정성 조건[31]을 만족한다고 주장한다.[32] 요컨대 번은 두 번째 단계를 해명하는 투명성 퍼즐을 안전한 지식을 산출하는 추론 규칙을 통해 해결하려고 시도한다. 하지만 번은 여전히 첫 번째 퍼즐에 대해서 실질적인 대답을 제공하지는 않는 듯하다.

---

[31] 지식의 안전성 조건은 소사(1999) 그리고 윌리엄슨(2000: ch. 5)이 제시한 바 있다.

[32] 이외에도 번은 주체에게 그렇게 보이는 바를 통해서 다양한 감각, 기분, 내적 발화 등에 대해서도 투명성 규칙을 제공한다. (2018: Chapter 6-8)

비판과 교훈

나는 번의 투명성 추론 규칙이 우리에게 어떤 의미에서 인식적으로 정당한 것으로 받아들여질 수 있는 규칙인지에 대한 번의 정당화가 성공적이지 않다고 생각한다. 우선 우리가 무엇을 정당한 추론 도식 혹은 규칙이라고 생각하는지를 검토해보자. 일반적으로 모든 전제가 사실이면 결론의 사실이 논리적으로 따라나오는 연역적 규칙을 일컬어 타당한 추론 도식이라고 부른다. 하지만 더 넓게 생각해서 전제와 결론 간에 논리적 관계가 성립하지 않더라도 전제의 내용과 결론의 내용 사이에 귀납적 관계가 성립하여 전제가 참이라면 결론의 참이 일반적으로 따라나오는 그러한 추론 도식이라면 이 또한 정당한 추론 도식이라고 볼 수 있을 것이다. 그리고 그렇기 때문에 정당한 추론이라면 전제와 결론 사이에 성립하는 일종의 규범적 이유 관계를 우리에게 알려주는 역할을 해야 한다.

하지만 믿음 도식에서 번은 스스로 인정하듯이 전제와 결론 사이의 적절한 이유 관계를 신경쓰지 않는다. 다만 전제와 결론 사이에 있는 모종의 안전한 관계 덕분에 자신의 도식이 좋은 추론 도식이라고 주장할 뿐이다. (2018: 109-12, 208) 하지만 번의 도식이 일견 안전한 추론 도식처럼 보일 수 있는 까닭은 일반적으로 **도식 사용자가 도식의 전제에 개입한다**는 우연적인 사실 덕분이다. 어떤 도식이 정말로 추론 도식인지는 전제에 개입하는 것과 전제를 단순히 참으로 가정하는 것을 구분해보면 명확하게 드러난다. 도식이라면 전제의 참을 믿지 않음에도 전제가 참이라는 *가정하에서* 정당하게 따라나오는 참인 결론이 무엇인지를 알기 위해 사용될 수 있어야 한다.

(전제) 텍사스의 주도는 휴스턴이다.

(결론) 텍사스의 주도는 휴스턴이나 댈러스이다.

실제로 텍사스의 주도는 어스틴이다. 하지만 우리는 텍사스의 주도가 휴스턴이라는 것을 참이라고 가정할 수 있다. 그리고 이러한 *가정 하에서* 결론은 참이다.

그러나 믿음 도식은 전제를 참이라고 가정하더라도 결론은 참이 아니다. 텍사스의 주도가 어스틴이라는 것을 아는 사람에게, 다음의 믿음 도식을 주었다고 하자.

(전제) 텍사스의 주도는 휴스턴이다.

(결론) 나는 텍사스의 주도는 휴스턴이라고 믿는다

그는 텍사스의 주도가 어스틴이라고 믿지만, 도식의 사용자로서 텍사스의 주도는 휴스턴이라고 **가정**할 수 있다. 그러나 이런 가정 하에서도 결론의 참은 따라나오지 않는다. 그는 텍사스의 주도가 휴스턴이 아니라 어스틴이라고 믿기 때문이다.

일반적으로 도식 사용자는 자신이 참이라고 믿는 명제를 전제에 대입해서 내가 믿어야 하는 결론을 얻기 위해서 도식을 사용한다. 그리고 이는 전제의 객관적 참과 그것에 개입함이 도식 사용자 시점에서 전혀 구분될 필요가 없는 상황이다. 하지만 일반적으로 추론 도식이라면 내가 개입하지 않는 전제더라도 **참으로 가정할 때** 어떤 결론을 믿어야 되는지를 알아보는 방식으로 사용될 수 있어야 한다고 여겨진다. 그리고 이 점에서 번의 추론 규칙은 전제를 단순히 가정하고 사용할 수 있어야 한다는 추론의 일반적 성격을 전혀 갖추지 못하고 있다. 따라서 번의 추론 규칙은 그것을 추론이라고 부를 근거가 매우 미약하다.

그래도 어찌됐건 간에 명제 'P'로부터 명제 '나는 P 를 믿는다'로의 이행이라는 점에서 이상하긴 하지만 추론으로 부를 수 있으며, P 의 참에 개입하는 사람이라면 이 추론 규칙은 매우 안전한 지식을 산출하므로 이렇게 산출된 자기귀속 믿음은 (외재적으로) 정당한 자기지식이라는 번의 주장까지 받아들여 보자. 번의 입장에서는 추론 관계의 필요조건으로 전제와 결론

사이에 내재적인 이유 관계가 있어야 한다는 입장은 내재주의자의 입맛에 따른 논점선취적 주장이라고 반박할 수 있다. 이에 따라 번은 투명성의 퍼즐에 대해서는 나름의 대답을 제시하고 있다고 하자. 하지만 더욱 중요한 것은 번의 추론의 전건이 그가 주장하는 것처럼 행위자 중립적인 사태를 나타내는지에 대해서도 매우 의문스럽다는 것이다. 번은 아래와 같은 BEL 과 유사한 투명성 규칙들의 변형들을 통해서 다양한 지향적 태도들에 대한 자기지식을 지식의 안정성 조건을 만족하는 방식으로 얻을 수 있다고 주장한다.

NOTBEL 만일 -P 라면, 당신이 P 를 믿지 않는다는 것을 믿어라.

NOVIEW 만일 당신이 P 인지에 대한 형편없는 인식적 위치에 놓여있다면, 당신이 P 를 믿지 않는다는 것을 믿어라.

CONFIDENCE 만일 당신이 P 를 믿고, 당신의 믿음이 높은 인식적 확신도를 가진다면, 당신이 P 를 높은 확신과 함께 믿는다고 믿어라.

NOTBEL 과 NOVIEW 는 이를 따르면 자신이 P 를 믿지 않는다는 자기지식에 안전하게 도달할 수 있는 추론 규칙이다. NOTBEL 는 일견 BEL 처럼 스스로에 대한 자기지식을 선행적으로 요구하지 않는 행위자 중립적인 명제인 것처럼 보인다. 하지만 이 중 NOVIEW 는 명확하게 행위자 자신에 대한 선행적인 지식-나는 P 인지에 대한 형편없는 인식적 위치에 놓여있다-을 포함하고 있다. 따라서 NOVIEW 의 규칙을 따르기 위해서는 모종의 자기지식이 이미 필요하다.

번은 NOVIEW 의 전건이 행위자 중립적으로 이해될 수 있는 한 가지 방식을 제안하는 것처럼 보인다. 예컨대 "나는 뉴델리에 비가 온다고 믿지 않는다"는 믿음을 얻기 위해서 NOVIEW 규칙을 사용한다고 해보자. 스스로 뉴델리에 비가 오는지에 대한 형편없는 인식적 위치에 놓여있는지 판단하기 위해서 주체는 뉴델리에 대한 날씨 뉴스를 접한 바가 없다거나 혹은

뉴델리의 오늘자 신문을 읽은 바가 없다는 등의 행동적 증거만을 고려할 수 있다. 그리고 이 경우 주체는 자신의 인식적 위치를 고려하기 위해서 자신에 대한 비-심리적 사실을 고려했으므로 NoView 의 전건은 행위자 중립적이다.

하지만 주체가 직접적인 방식으로 자신의 믿음에 대해서 아는 것이 자기지식의 인식적 특권성의 핵심임을 고려한다면, 자신의 행동적 증거만으로 P 를 믿지 않는다는 것을 판단하더라도 이렇게 얻어진 자기지식은 특권적 자기지식이 될 수 없다. 그리고 이는 단지 양적인 측면에서 자신이 어떤 인식적 상황에 놓여있는지에 대한 정보를 더 갖고 있기 때문에 가질 수 있는 자기지식이라는 점에서 2 장에서의 방법적 구분에서 보았던 것처럼 특유한 접근이지도 않다.

또한 스스로 무엇을 믿지 않는다는 것을 알고자 하는 일반적 상황을 가정하더라도 번의 모형은 별다른 설명력을 갖지 않는다. 뉴델리에서 비가 올 것이라는 기상 예보를 들었고 따라서 자신이 뉴델리에서 비가 올 것이라는 것을 알 만한 인식적 위치에 놓여 있다고 판단하더라도, 실제로 오늘 뉴델리에서 비가 올 것인지에 대해서는 확신이 없을 수 있으며 바로 이 확신의 결여에 근거해 뉴델리에서 비가 올 것이라고 믿지 않는다고 믿는 것 또한 특정 믿음이 없음을 믿는 자연스러운 과정인 것처럼 보인다. 즉, 나의 인식적 위치에 대한 판단은 일반적으로 나의 주관적 확신성에 근거하여 내려지는 것 같다. 만일 그렇다면 NoView 는 Confidence 와 유사한 규칙, 즉 자신이 P 에 대해 낮은 확신을 가진다는 믿음의 산출 규칙으로 이해될 수 있으며, 설령 그것을 행위자 중립적으로 이해할 수 있는 방향이 열려있다고 하더라도 특권적 자기지식을 산출하는 규칙이기 위해서는 행위자 중립적이지 않은 규칙이어야 한다.[33]

중요한 것은 NotBel 와 Bel 그리고 NoView 는 모두 믿음에 대한 자기지식을 얻기 위한 규칙이라는 사실이다. 그렇다면 어떤 차이로 인해

---

[33] 번 또한 스스로 NoView 가 중립적 규칙이 아님을 명시적으로 밝히고 있다. (Byrne, 2018: 121)

어느 것은 행위자 중립적이지 않은 전건을, 다른 둘은 행위자 중립적인 전건을 가지게 되는 것인가? 세 규칙은 믿음을 지식의 대상으로 한다는 점에서 차이가 없다. 그렇다면 BEL 과 NOTBEL 의 전건은 행위자 중립적이지만 NOVIEW 는 그렇지 않은 이유가 무엇인가? 나는 세 규칙이 전건의 행위자 중립성에서 차이를 가지게 만들 어떠한 다른 요소도 갖지 않는다고 생각한다. 이는 세 규칙이 유형적으로 분명히 구분되는 세 지향적 태도를 알기 위한 규칙이 아니라 단지 정도에서만 차이를 가지는 지향적 태도들에 도달하는 규칙이기 때문이다. 이처럼 만일 NOVIEW 의 전건을 행위자 중립적인 것으로 만들 수 없고, 세 규칙 사이에 전건의 행위자 중립성에 있어서 차이를 가져올 어떠한 요소도 찾기 어렵다면, NOTBEL 와 BEL 의 전건이 확실히 믿지도, 확실히 믿지 않지도 않는 상태에 대한 자기지식을 산출하는 NOVIEW 와 같은 규칙의 전건처럼 행위자 중립적이지 않다는 결론이 우리에게 남는 가장 합리적인 선택지이다.

그렇다면 번의 이론이 성공적이기 위해서 추론 규칙의 전건은 반드시 행위자 중립적이어야 하는가? 그는 왜 자신의 규칙의 전건이 행위자 중립적임을 보이고자 하는가? 추론 규칙의 전건이 행위자 중립적이지 않아야 한다는 것은 내적 감각 이론과 자신의 이론을 구분하고자 하는 번 스스로의 목적과는 연관되어 있지만, 특권적 자기지식의 모형에서 전건이 반드시 행위자 중립적이어야 하는 이유는 없다. 하지만 만일 번의 추론 규칙, 특히 가장 핵심이 되는 믿음 규칙이 실제로는 행위자 중립적인 규칙이 아니라면, 번의 이론은 일반적인 내성 이론, 특히 내적 감각 이론과 구분되지 않을 위험에 처하는 것처럼 보인다. 실제로 번(2018: Chpater 2)은 자신의 입장이 내적 감각 이론과 공유하는 부분이 많다고 진술하며, 책의 한 장을 내적 감각 이론에 대한 기존의 비판에 대한 재반박에 할애하고 있다. 그럼에도 불구하고 번은 중립적 전건으로부터 자기지식으로 이행하는 규칙에는 내적 감각이 불필요하며, 따라서 자신의 투명성 추론주의가 내적 감각 이론보다

경제적이라는 점에서 이론적 강점을 찾는다.[34] 따라서 만일 번의 규칙이 실제로 행위자 중립적이지 않다면 번은 내성 이론으로 회귀하거나 혹은 자신의 이론을 투명성 이론으로 불 수 있는 새로운 설명을 추가적으로 제공해야 할 것이다.

하지만 번의 논의로부터 우리는 다음의 교훈을 얻을 수 있다. 이전까지 자기지식 논의에서 주체 자신에게 있는 믿음 P 를 의식하는 것과 자신이 P 라고 믿는다는 것을 아는 것은 크게 구분되지 않았다. 예를 들어서 내성 이론 중 대면 이론의 경우에도, (만일 현상적 지향성 이론이 맞다면) 엄밀히 말해서 나에게 대면되는 것은 내용 P 이며 그것이 '나의 믿음'이라는 사실까지 대면의 대상이 되는 것은 아니다. 즉, 믿음 P 를 의식하는 것과 나는 P 라고 믿는다는 지식을 얻는 것 중에서 진정한 의미에서의 자기지식은 후자이며, 믿음 P 에 대해 의식함으로부터 나는 P 라고 믿는다는 지식을 얻는 상태에 이르는 데에는 또 다른 인식적 도약이 더 필요하다는 것이다.[35] 번이 신빙성있는 추론을 통해서 채우고자 했던 인식적 틈은 바로 이 지점에 있다. 즉, 자기지식 이론이라면 자기지식을 얻기 위해 주체의 시점에서 'P'로부터 '나는 P 를 믿는다'로의 이행을 어떻게 정당화할 것인지에 대한 답, 즉 투명성의 퍼즐에 대한 답을 제공해야 한다는 것을 번은 잘 지적하고 문제화하였다. 하지만 번이 자신의 "추론 규칙"이 정말로 일반적인 의미에서 정당한 추론으로 여겨질 수 있는지에 대해서 충분히 설득력있게 제시하고

---

[34] Byrne (2018: 26)

[35] 내적 감각 이론으로 묶을 수 있는 니콜스와 스티치(2002)가 제안하는 자기지식 생성 메커니즘은 이러한 인식적 도약을 필요로 하지 않는 것으로 보인다. 그들이 제시하는 자기지식 생성 모형에 따르면, 우리는 믿음을 일종의 "믿음 상자"에 보관하고 있다가, 믿음 상자에 들어있는 P 를 꺼내어 이를 'I believ that _____'이라는 통사론적 틀에 넣고 다시 믿음 상자에 집어 넣는 방식으로 자기지식을 생성하는 특정한 인과적 메커니즘에 따라서 자기지식을 얻는다. 하지만 나는 1 장에서 제시한 내적 감각 이론에 대한 비판에 더해 "믿음 상자"와 같은 장치를 비유가 아닌 의식에 존재하는 형이상학적 요소로서 사용하는 것을 받아들일 이유가 충분한 이유가 없다고 생각한다.

있는지, 따라서 투명성의 퍼즐에 대한 답을 잘 제공하고 있는지 분명하지 않다.[36] 결정적으로 그의 추론 규칙이 행위자 중립적이지 않다면, 그는 결국 투명성 퍼즐 단계 이전에 사태에 대한 판단에 도달하는 메커니즘에 대하여 설명하는 바가 아무 것도 없다는 점에서 일인칭 특권적 자기지식을 설명하기에 충분한 모형을 제시하는 데 실패한다.

3.3. 모란의 합리적 행위자 이론

3.3.1. 합리적 행위자성과 숙고적 자기지식

지향적 태도에 대한 자기지식의 논의에 에반스의 투명성을 본격적으로 끌어들인 것은 모란이다. 모란은 *Authority and Estrangement* (2001)에서 에반스의 투명성과 합리적 행위자성 사이의 새로운 통찰을 제시함으로써 일인칭 특권적인 자기지식의 본성에 대한 이론을 제공하고자 하였다.

모란의 자기지식 이론을 이해하기 위해서는 먼저 자기소외적 자기지식과 자기소외적이지 않은 숙고적 자기지식의 구분을 알아야 한다. 모란에 따르면, 합리적 행위자는 자기지식을 얻을 때 이론적 관점에서 얻을 수도 있지만 이론적이지 않은 관점에서 얻을 수도 있다. 어떤 심적 상태를 가졌을 때 보일 법한 증거들을 근거로 심적 상태를 이론적 추론을 통해 정당하게 귀속시킨다면 해당 주체는 이론적 관점에서 자기지식을 얻은 것이다. 반면 합리적 행위자는 투명성 절차를 통해서 **이론적 증거없이도** 직접적으로 자신에게 귀속시킬 심적 상태를 알 수 있다. 투명성 조건을 만족하는 방식으로 얻어진 자기지식은 심적 상태를 가짐 자체에 대한 증거없이 비관찰적이고 직접적인 방식으로 얻어진다는 점에서 마음에 대한 이론적인 증거를 바탕으로 얻어지는 자기지식과 구분된다. 한편 이론적 관점에서

---

[36] 번의 추론 규칙이 정말로 "추론"으로 생각될 수 있는지에 대하여는 보일과 세티야가 별도로 논의한 바 있다. 보일은 번을 비판하였고, 세티야는 번의 입장에서 그의 추론 규칙을 옹호한다. Boyle (2011)과 Setiya (2012) 참고.

얻어졌지만 투명성 조건을 만족하지 않는 자기지식은 자기소외적 자기지식이며, 투명성 조건을 만족하는 방식으로 얻어진 지식은 자기소외적이지 않은 자기지식이자 행위자가 자인하는(avow), 즉 스스로 정말 자신의 것으로 진심으로 인정할 수 있는 숙고적 자기지식이다. 여기서 자기소외적이지 않은 자기지식이 숙고적 자기지식인 까닭은, 이것이 주체가 합리적 행위자로서 행위자성을 발현하여 유관한 세계의 사태에 대해 숙고하고, 그 결과를 근거로 행위자성을 발휘해 형성한 심적 상태에 대한 자기지식이기 때문이다. (Moran 2001, Chapter 3-4.)

모란에게 자기지식의 특권성은 숙고적 자기지식을 구성하는 합리적 행위자성에 근거한다. 그렇다면 자기소외적 자기지식과 숙고적 자기지식의 구분은 자기지식의 특유한 접근과 특권성을 어떻게 설명하는가? 모란은 숙고적 자기지식이 자기지식의 인식적 특권에 대해 어떤 대답을 줄 수 있는지를 명시적으로 이야기하지 않는다. 하지만 모란은 이후 몇몇 학자들이 제기한 문제[37]를 다루는 과정에서, 합리적 행위자라면 만족하는 **초월적 가정(transcendental assumption)**을 도입하면서 연관 물음에 답하려는 것처럼 보인다.

> "만일 나의 [비가 온다는] 이 믿음이 [비가 온다는 것을 옹호하는] 이유들에 대한 반성을 통해서 결정된 것이라고 가정할 수 있다면, 나는 비가 온다는 것을 옹호할 이유에 대한 반성이 비가 옴에 대해 내가 믿는지에 대한 질문에 대답을 제공한다고 가정할 권리를 가질 것이다." (2003: 405-6)

모란은 인식 주체가 어떤 사태 P 가 성립함을 옹호하는 이유들에 대한 반성을 통해 P 라는 내용을 가지는 나의 믿음이 결정된다고 가정하면 일차 믿음과 이차 믿음(자기지식) 사이의 인식적 틈이 메워진다고 주장한다. 그리고 이 가정은 합리적 행위자라면 그렇게 가정할 권리를 갖는 그러한

---

[37] 예를 들어, Shoemaker (2003)

가정이다.[38] 그리고 P 가 참일 좋은 이유들로부터 P 의 참에 개입할 것을 결의한 주체는 합리적 행위자에게 허용되는 일종의 초월적 가정을 통해 자신이 P를 믿는다는 것을 정당하게 알 수 있다.

요컨대 모란은 합리적 행위자가 숙고적 자기지식을 얻기 위해 다음과 같은 과정을 거친다고 주장한다.

(1) 스스로가 P 를 믿는지를 알기 위해서, 합리적 행위자는 P 와 관련된 증거들을 숙고한다.

(2) P 와 관련된 증거들을 숙고한 결과, 이로부터 합리적으로 P 에 개입할 것을 결의한다.

(3) P에 개입할 것을 결의한 합리적 행위자는 초월적 가정에 따라서 자신이 P를 믿는다는 것을 정당하게 알게 된다.

모란에 따르면 합리적 행위자는 스스로의 믿음을 믿음 상태와 관련된 증거를 통해서 이론적으로 발견하는 것이 아니라 그 믿음이 지향하는 사태가 참인지를 고려함으로써 알게 된다. 그리고 그 믿음이 지향하는 사태가 참인지를 고려하는 것은 그것을 믿을 다른 이유들을 고려하는 과정, 즉 참일 좋은 이유가 있는지를 고려하는 과정이다.

---

[38] 모란의 설명은 충분히 명료하지는 않다. 이는 합리적 행위자라면 그렇게 가정해야 한다는 것인가? 아니면 그렇게 가정하고 있다는 것인가? 그렇게 가정해야 한다는 것이라면, 나는 모란의 주장에 동의한다. 하지만 그렇게 가정해야 한다고 해서 실제로 행위자가 그렇게 가정한다는 것이 따라나오는 것은 아니므로, 이 경우 모란의 설명은 자기지식의 실제 획득 메커니즘을 설명하는 것에 이르지 못한다. 모란이 투명성을 통해 자기지식을 얻는 것을 비-추론적인 방식이라고 명확히 밝히고 있다는 점을 고려할 때(2001: 92, 134), 이러한 초월적 가정을 이해하는 한 가지 방식은 초월적 가정이 행위자가 암묵적로 받아들이는 가정이라는 것이다. (모란의 초월적 가정을 암묵적인 가정으로 이해하는 방식은 서울대 강진호 교수님이 제안해주셨다.)

비판

이후 모란의 합리적 행위자 이론에 가장 많이 제기된 질문은, 모란의 주장과는 달리 나의 믿음을 알기 위해서는 모란이 말하는 합리적 행위자성과 숙고가 굳이 필요하지 않아 보인다는 것이다. 그것이 참일 규범적 이유를 고민해봄으로써 믿음의 대상이 되는 사태의 참에 개입하기를 결의하는 것은 믿음의 형성, 즉 열린 질문을 닫는 것에 해당하며, 이미 갖고 있는 믿음에 대해 다시 확인하는 경우에는 우리는 일반적으로 해당 사태를 믿을 규범적 이유를 살펴보지 않는다는 것이다. 또한 많은 경우 주체는 어떤 믿음을 어쩔 수 없이 믿을 수밖에 없는 것처럼 보이는데, 모란의 숙고 개념은 이러한 믿음의 일반적인 불수의적 성격에 어울리지 않아 보인다.[39]

후자의 비판에 대하여 모란은 자신이 "숙고"와 "결의"를 통해서 말하고자 했던 바가 믿음에 대한 주의주의(voluntarism)적인 주장에 개입하려는 것은 아니었다고 해명한다. 모란에 따르면, 자신이 말하는 "합리적 행위자성"이란 이유를 살펴보고 그로부터 그 이유가 가리키는 결론에 도달한다는 의미에서 약한 의미의 행위자성, 혹은 **이유-반응성**이다. 즉, 숙고와 결의 그리고 합리적 행위자성으로서 모란이 이야기하고 싶었던 바는 합리적 행위자의 지향적 태도가 세계의 대상에 대한 다양한 규범적 이유에 대한 행위자의 이유-반응성을 통해서 결정되며, 이러한 이유-반응성으로서의 합리성이 자기지식의 특유한 접근과 일인칭 특권성을 설명하는 것에도 핵심적인 역할을 한다는 것이다.[40]

그러나 나는 다음의 이유에서 모란의 해명이 만족스럽지 않다고 생각한다. 합리적 행위자성이 '숙고'와 '결의'가 일반적으로 의미하는 적극적인 행위자성이 아니고 모란의 주장처럼 이유-반응성으로 이해되더라도, 믿음을 제외한 나머지 지향적 태도들에 대해서 합리적 행위자성은 자기지식의 특유한 접근과 일인칭 특권을 설명하는 데 실질적인 역할을 하지 못한다.

---

[39] Shoemaker (2003), Heal (2004), Gardner (2004)

[40] Moran (2003)

2 장에서 언급했던 무어의 역설을 다시 생각해보자. 무어의 역설은 다음의 형식을 취하는 두 진술의 연언이 세계의 사태에 대한 보고와 일차 믿음에 대한 보고 사이에 존재하는 유사-모순성을 보여준다는 것이다.

[무어의 역설적 연언] 'P, 그리고(그러나) 나는 P 를 믿지 않는다.'
'-P, 그리고(그러나) 나는 P 를 믿는다.'

1 장에서 나는 무어의 진술이 역설로 읽히는 것이 보여주는 바가 다음과 같다고 주장했다: 자신의 믿음에 대해서 얻은 자기지식을 보고할 때, 우리는 믿음의 자기귀속 진술이 삼인칭 시점에서 이론적 증거를 근거로 보고되는 것이 아니라 **일인칭 시점에서** 사태와 동기화되는 방식으로 얻어진다는 것을 자연스럽게 생각한다. 이는 일인칭 시점에서 무어의 연언을 이해했을 때 세계의 사태에 대한 명제 'P'의 참에는 개입하면서 동시에 자신의 믿음을 보고하면서는 P 를 믿지 않는다고 진술하는 것은 마치 모순처럼 이상하게 느껴진다는 사실에서 알 수 있다.

모란의 제안에 따라서 사태의 성립에 관한 규범적 이유들을 살펴본 결과 이유들이 사태가 성립함을 가리킨다고 해보자. 그리고 규범적 이유들을 살펴본 결과 참일 좋은 이유가 있다는 판단과 그러한 명제를 믿는다는 판단 사이에는 틈이 거의 없어 보인다. 따라서 믿음의 경우, 규범적 이유들을 살펴본 결과 참에 개입하고 이로부터 자신의 믿음을 알게 되는 것은 그럴 듯하다.

하지만 다른 지향적 태도들 일반에 대해서 그러한 태도를 가질 좋은 이유와 실제로 그러한 태도를 가짐 사이에 존재하는 인식적 틈을 무시할 수 있는가? 우리가 〈믿음〉을 제외한 다른 지향적 태도들의 개념들을 어떻게 사용하는지 알아볼 수 있는 한 가지 방법은 각각의 지향적 태도에 대해서도 일종의 무어적 진술을 만들어보고 역설로서 읽히는지를 확인하는 것이다. 믿음 이외의 명제적 태도에 대하여, 특정 사태가 성립할 좋은 이유가 있음에도 그러한 좋은 이유에 따라 개념적으로 연관되는 특정 태도를 갖지는

않는다는 진술을 확인해보았을 때, 다른 지향적 태도들에 대해서는 다음의 진술들이 역설처럼 느껴지지 않는 듯하다.

[연언의 변주]

'P 가 바람직할 좋은 이유가 있다. 그리고 나는 P 를 바라지(욕구하지) 않는다.'

'P 가 부러울 만한 좋은 이유가 있다. 그리고 나는 P 를 부러워하지 않는다.'

'P 에 분노할 만한 좋은 이유가 있다. 그리고 나는 P 에 분노를 느끼지 않는다.'

이 진술들은 믿음에 대한 무어식 진술과는 달리 일인칭 시점에서 진술되는 것으로 생각해도 충분히 자연스럽게 역설로 이해하지 않을 수 있다. 이는 이상의 지향적 태도들은 그 개념이 애시당초 그러한 태도를 취해야 할 규범성을 따름을 개념적으로 함축하고 있지 않기 때문이다. 이러한 태도들에 대해서는, 모든 지향적 태도를 그러한 태도를 취할 규범적인 이유에 따라서만 형성하는 극도로 합리적인 행위자가 아닌 이상 모란이 말하는 이유-반응성이 자기지식의 특유하고 특권적인 접근을 설명해줄 수 없게 된다. 물론 만일 모란이 가정하는 이유-반응성이 이처럼 모든 지향적 태도를 그러한 태도를 취할 좋은 이유에 따라서 형성하는 극도로 강한 이유-반응성이라면 모란의 이론은 자기지식에 대해 적절한 설명을 제공해줄 수 있을 지도 모르겠다. 하지만 그러한 이론은 거의 모든 실제 사람들에게는 적용될 수 없는 공허한 자기지식 이론이 될 것이다.[41]

---

[41] 이는 번의 욕구 규칙과 의도 규칙에 대해서도 마찬가지로 적용될 수 있는 비판이다. 번은 아래의 두 규칙을 각각 욕구와 의도에 대한 자기지식을 얻는 추론 규칙으로 제시한다.

DES 만일 ø 하는 것이 바람직한 선택지라면, 당신이 ø 하기를 원한다는 것을 믿어라.

물론 이 비판은 성공적이더라도 적어도 **믿음**에 대한 자기지식을 얻는 것에 있어서 합리적 행위자성이 핵심적인 역할을 한다는 것까지 기각하지는 않는다. 믿음의 경우에는 사태에 대한 규범적 이유-반응성이 실제로 갖게 되는 심적 태도가 무엇일지와 충분히 긴밀하게 연결되어 있다. 이러한 사실은 믿음의 대상이 되는 사태가 참일 규범적 이유 진술과 일인칭 자기 귀속 진술의 연언에 무어의 역설이 적용된다는 사실에서 알 수 있다.

하지만 나는 모란의 자기지식 이론이 믿음에 대한 자기지식의 특권성을 설명하는 데 있어서도 부족하다고 생각한다. *Authority and Estrangement* 에서 모란은 P 의 참에 도달하는 메커니즘이 무엇인지를 묻는 판단의 메커니즘의 문제에 대한 해답으로 규범적 이유에 대한 숙고와 결의를 제시하는 것처럼 보인다. 즉, 주체는 P 가 참인지 확인하기 위해서 P 가 참일 좋은 이유들을 숙고하고, 이를 토대로 P 에 개입하기로 결의한다. 하지만 이어지는 후속 논문(2003)에서 모란은 일반적으로 P 가 참인지 확인하기 위해 결의는 물론 굳이 숙고의 과정도 거치지 않는 것 같다는 슈메이커(2003: 395-396)의 비판에 자신이 말하는 합리적 행위자성이 숙고 과정을 거치는 것까지 요구하지는 않으며 다만 이유-반응성으로서의 약한 행위자성이라고 대답한다.

하지만 모란의 2003 년의 입장은 투명성 절차를 구성하는 두 가지 단계 중 첫 번째 단계인 판단의 메커니즘의 문제에 대한 해답을 더 이상 제공해주지 못하는 것처럼 보인다. 그는 2001 년 글에서 숙고와 결의를 통해 주체가 미래의 세계 삼차 대전의 발발에 대한 개입에 도달하고 이로부터 세계 삼차

INT 만일 당신이 ø 할 것이라면, 당신이 ø 하기를 의도한다고 믿어라.

두 규칙에 대해서도 마찬가지로 개념적으로 믿어야 할 좋은 이유가 있음과 실제로 믿음 사이의 틈이 작은 〈믿음〉과 달리 〈욕구〉는 욕구할 좋은 이유가 있음과 욕구함 사이의 틈이 넓다는 비판이 적용될 수 있다. 번은 바람직함에도 불구하고 바라지 않는 것이 충분히 가능하다는 것을 인정하면서도 나름의 이유를 들어 욕구 규칙과 의도 규칙이 자기지식을 산출하기에는 충분히 자기-확정적이며 따라서 안전한 규칙임을 보이기를 시도한다. Byrne (2018: Chapter. 7) 참고.

대전이 벌어질 것이라는 믿음을 정당하게 보고할 수 있다고 설명한다. 하지만 2003 년 글에서는 숙고와 결의의 과정이 실질적으로 필요하지 않음을 인정하면서 명제의 참에 개입하기 위해 주체가 겪게 되는 실질적인 메커니즘에 대한 설명을 더이상 공란으로 남겨두게 되었다.

모란의 입장에서 가능한 하나의 대답은 주체가 합리적 행위자이기 때문에 설령 어떤 명제를 믿을 규범적 이유를 실제로는 떠올리지 않더라도, 즉 규범적 이유에 대해 숙고하지 않더라도 **자동적으로** 자신이 아는 규범적 이유에 반응하는 방식으로 P 의 참에 개입한다는 것이다. 즉, 이유-반응성으로서의 합리성은 행위자로 하여금 그 이유를 의식적으로 떠올릴 것까지는 요구하지 않는다는 것이다. 실제로 모란은 가장 최근의 논문(2012)에서 이와 유사한 방식으로 이유-반응성을 생각하는 것 같다. 2012 년의 글에서 모란은 세계의 사실에 대한 질문을 물어봤을 때 적절하게 대답하면서도 자신이 어떤 이유에서 이러한 대답을 하고 있는지는 대답하지 못하는 주체의 사례를 합리적 주체의 사례로 인정한다. 이 주체는 설령 그렇게 믿을 좋은 이유가 무엇인지 자신의 의식으로 불러내지 못하더라도 어찌됐건 적절한 대답을 제시하는 것처럼 보이는데, 이는 사태에 대한 판단에 그의 합리성이 항시 작동하고 있기 때문이다. 또한 주체는 질문에 대한 적절한 답이 왜 이 대답인지 명시적으로 말하지는 못함에도 자신감있게 대답하는데, 이는 비록 그것이 사실일 좋은 이유가 나에게 보이지 않음에도 불구하고 합리적인 주체는 자신에게 주어진 물음이 무엇을 묻는 질문인지 잘 이해하고 있고, 주어진 사태가 성립할 가능성이 주체 자신에게 매우 그럴 듯해보이기 때문이다.

나는 합리적 행위자라면 세계에 존재하는 다양한 규범적 이유들을 인지하고 이유들에 반응하여 믿음을 끊임없이 형성하고 수정하며 합리성의 규범에 따라 조정해나간다는 모란의 핵심 주장에 전적으로 동의한다. 또한 설령 이유-반응적인 행위자가 자신이 그렇게 믿는 이유를 의식으로 불러내지 못한다고 해서 그가 더이상 합리적 행위자가 될 수 없다고 생각하지도 않는다. 다만 내가 말하고자 하는 바는 자기지식이 어떤 메커니즘을 통해

얻어지고 이 메커니즘이 자기지식에 어떻게 인식적 특권성을 부여하는지를 논하는 지금의 맥락에서, 판단의 메커니즘을 구성하는 구체적인 방법으로 숙고와 결의를 제시했던 이전의 입장과는 달리 합리성을 약한 이유-반응성으로 생각하는 최근의 입장은 실제로 주체가 겪는 심리적인 메커니즘이 무엇인지, 그러한 메커니즘과 자기지식의 인식적 특권성 사이의 관계에 대해서는 구체적인 설명을 제공하지 않는다는 것이다.

정리하자면, 2001 년의 모란과 2012 년의 모란은 자기지식의 일인칭 특권성을 설명하는 데 있어 제각기 다른 어려움에 처하고 있다. 숙고와 결의를 자기지식을 얻는 투명성 절차의 핵심적인 과정으로 보았던 초기 모란의 입장은 합리적 행위자가 믿음의 투명성을 사용하여 자기지식을 얻을 수 있는 과정을 제시하는 듯 보였다. 하지만 문제는, 우리가 실제로 자기지식을 얻을 때 대부분의 경우 정말로 이유에 대해 살펴보고 명제의 참에 개입하기를 결의하지 않는다는 것이다. 이후 여러 학자들의 지적을 받아들인 모란은 합리적 행위자가 사태에 개입하기 위해서 굳이 규범적 이유에 대해서 숙고할 필요가 없을 뿐만 아니라, 더 나아가 규범적 이유를 더이상 떠올릴 것까지도 요구되지 않는다고 입장을 바꾼다. 하지만 새로운 입장에서 모란은 자기지식 획득에서의 숙고와 결의 과정을 포기하면서 초월적 가정 이전에 사태에 대한 판단에 이르게 되는 메커니즘을 더이상 설명하지 못하게 되었다.

교훈

모란이 발견한 일인칭 자기지식의 특징은 다음과 같다:

(1) 일인칭 시점에서 자기지식은 투명성 조건을 만족시키는 방식으로 얻어진다.

(2) 일인칭 시점에서 믿음에 대한 자기지식은 마음에 대한 이론적 추론을 통해서 얻어지는 것이 아니라, 자기지식의 대상이 되는 믿음이 지향하는 사태가 성립하는지 대한 고려를 통해서 얻어진다.

(3) 투명성 절차를 통해서, 즉 사태에 대한 고려를 통해서 얻어진 자기지식은 투명성 절차를 통해서 얻어질 수 없는 자기지식과 달리 자기소외적이지 않은 자기지식이다.

모란이 주는 중요한 교훈은 자신의 심적 상태에 대한 일인칭 특권과 타인에 대한 지식 사이의 불일치가 자기소외성과 연관이 있다는 것을 보여주었다는 데 있다. 즉, 일인칭 특권적으로 얻어진 자기지식이라면, 행위자는 이 믿음을 단순히 이론적 관점에서 알게 된 것이 아니라 진심으로 자인하는 것으로서 알아야 한다. 물론 나는 자기소외적이지 않음의 근원이 그러한 태도를 가질 규범적 이유를 인지하거나 혹은 규범적 이유에 반응하는 것에 있다는 모란(2001)의 주장에는 동의하지 않지만, 그럼에도 그는 합리적 주체에게 사태에 대한 판단으로부터 자기지식에 도달하는 투명성 절차를 만족하는 방식으로 자기소외적이지 않은 자기지식이 얻어진다는 것을 설득력있게 보여주고 있다.[42]

이에 더하여 나는 초월적 가정이 합리적 주체가 암묵적으로 가정하는 바로 이해되었을 때 일인칭 시점에서의 얻어지는 자기지식의 인식적 보증을 설명해줄 수 있는 좋은 대답이라고 생각한다. 그의 초월적 가정은 투명성의 퍼즐에 대한 상식적인 대답이며, 만일 초월적 가정을 받아들인다면 번이 제안한 불분명한 추론 규칙을 받아들이지 않으면서도 성공적으로 투명성 퍼즐을 해결할 수 있다.

---

[42] Moran (2001: Chapter 3-4)

[초월적 가정]

"믿음의 주제인 P 를 옹호할 이유에 대한 나의 반성이 P 에 대한 나의 실제 믿음에 대한 질문과 어떤 관계를 가지고 있다고 생각할 그러한 권리가 나에게 있는가? ... 이에 대한 나의 생각은 다음과 같다: 만일 나의 [비가 온다는] 이 믿음이 [비가 온다는 것을 옹호하는] 이유들에 대한 반성을 통해서 결정된 것이라고 **가정**할 수 있다면, 나는 비가 온다는 것을 옹호할 이유에 대한 반성이 비가 옴에 대해 내가 믿는지에 대한 질문에 대답을 제공한다고 가정할 권리를 가질 것이다. [비가 온다는 이 믿음이 무엇인지가 비가 온다는 것을 옹호하는 이유들에 대한 반성을 통해서 결정된 것이라는] 이 **가정**은 [내가 비가 오는지를 믿는가와 비가 오는가의] 두 질문 사이의 연결고리를 제공해준다. 이제, 물어보자. 나는 어떤 문제에 대해서 결론을 내리는 나의 방식을 생각하는 과정에 있을 때마다 이러한 가정을 내리고 있지 않는가?" (2003: 405-6)[43]

나는 'P'로부터 '나는 P 를 믿는다'로 이행하는 것이 신빙적인 과정으로서 인식적 권리를 얻는다는 번의 추론 규칙은 거부했지만, 대신 합리적인 행위자가 일인칭 시점에서 P 의 참을 자인하는 것으로부터 P 에 대한 믿음을 귀속시키는 것은 정당하다는 모란의 초월적 가정의 경우는 매우

---

[43] "What right have I to think that my reflection on the reasons in favor of P (which is one subject-matter) has anything to do with the question of what my actual belief about P is? ... then my thought at this point is: I would have a right to assume that my reflection on the reasons in favor of rain provided an answer to the question of what my belief about the rain is, if I could assume that *what* my belief here is was something determined by the conclusion of my reflection on those reasons. An assumption of this sort would provide the right sort of link between the two questions. And now, let's ask, *don't* I make just this assumption, whenever I'm in the process of thinking my way to a conclusion about some matter?" (2003: 405-6)

그럴듯하다고 생각한다. 이는 모란 스스로 말하는 것처럼 대상이 되는 사태가 참인지에 대한 고려가 믿음 귀속 질문에 대한 적절한 답을 준다는 가정이 합리적 행위자라면 암묵적으로 가정하고 있을, 행위자의 시점에서 정당한 가정이기 때문이다. 행위자는 자신이 세계의 사태가 어떠한지 판단하고, 이떨 때는 이러한 판단을 근거로 다른 사태에 대하여 이전과는 달리 판단을 내리기도 한다. 그리고 이러한 판단은 그 자체로 행위자가 바로 그 세계의 사태에 대해 어떤 믿음을 가지고 있는지와 연결되어 있다. 따라서 초월적 가정은 'P 라고 믿는가?'라는 질문을 받고 P 가 참인지를 고려하는 주체에게 왜 믿음 귀속 질문에 대해 답하기 위해 믿음의 대상이 되는 사태를 고려하는지 물었을 때 자연스럽게 대답할 법한 이유이다. 그리고 만일 초월적 가정이 정당하다면, 모란이 주장하는 것처럼 P 로부터 내가 P 를 믿는다는 결론으로 나아가는 단계에서 발생하는 투명성 퍼즐은 자연스럽게 더이상 퍼즐이 아니게 된다.

3.4 보일의 반성주의 이론

3.4.1. 비-정립적 의식과 반성적 자기지식

보일(2019)은 사르트르의 의식 이론으로부터 의식에 대한 정립적 의식(positional consciousness)과 비-정립적 의식(non-positional consciousness)의 구분을 받아들인다면 투명성의 퍼즐을 풀 수 있다고 주장한다. 보일에 따르면, 사르트르는 의식이 언제나 무엇에 대한 의식임을 받아들이고 어떤 의식이 이처럼 대상을 가질 때 그것을 '정립적 의식'이라고 부른다. 이때 '정립'은 사르트르의 용어로, 그것의 대상과 의식이 가지는 관계를 가리킨다. 사르트르에 따르면 대상을 가지는 모든 정립적 의식이 그 자체로 동시에 스스로에 대한 "비-정립적" 의식이라는 것은 의식에 대한 근본 명제 중 하나다. 다음은 비-정립적 의식의 존재를 논증하는 사르트르에 대한 보일의 인용이다:

"무언가를 아는 의식(knowing consciousness)이 그것의 대상에 대한 앎이 되기 위한 필요 충분 조건은, 이 의식이 스스로를 이러한 아는 의식으로 의식해야 한다는 것이다. [다음]은 무언가에 대한 의식이기 위한 필요조건이다: 만일 나의 의식이 책상에 대해 의식함에 대한 의식이 아니라면, 그것은 책상에 대한 의식이지만 스스로에 대한 의식은 아닐텐데, 이는 비의식적 의식(unconscious consciousness)이다. 그리고 이는 터무니없다. 그리고 [다음]은 무언가의 의식이기 위한 충분조건이다: 이 책상에 대한 의식임을 의식하는 것은 실제로 내가 [이 책상]을 의식하는 것에 충분하다."[44]

보일은 대상은 갖지만 그 자체로 의식되지 않는 의식의 존재를 대부분의 현대의 철학자들이 받아들일 것이기 때문에, '비의식적 의식'이 자기모순이라는 사르트르의 생각을 받아들이지 않을 것임을 인정한다. 하지만 그는 '반성을 가능하게 하는 것은 비-반성적 의식이다'라는 사르트르의 또 다른 의식에 대한 근본 명제를 받아들인다면 사르트르의 비-정립적 의식 논증을 보충할 수 있다고 생각한다. 즉, 반성을 통해 의식을 그 자체로도 우리의 대상으로 정립할 수 있는 것은 그 의식 상태에 내재적으로 속해있는 비-정립적 의식이 존재하기 때문에 비로소 가능하다는 것이다.

보일은 일반적인 정립적 의식과 그에 언제나 동반하는 비-정립적 의식의 구분을 심적 상태의 투명성과 연결시킨다. 자기지식이 투명성 절차, 즉

---

[44] "[T]he necessary and sufficient condition for a knowing consciousness to be knowledge of its object is that it be consciousness of itself as being this knowing. This is a necessary condition: if my consciousness were not consciousness of being conscious of the table, then it would be consciousness of the table without consciousness of being such, i.e., a consciousness ignorant of itself, an unconscious consciousness – which is absurd. It is a sufficient condition: that I am conscious of being conscious of this table suffices for me in fact to be conscious of it."
(Sartre 1956, *Being and Nothingness* lii/18, Boyle 2019; 1029 에서 재인용)

P 자체에 대한 고려를 통해서 획득되는 까닭은, P 를 지향하는 의식이 있다면 언제나 이미 인식 주체에게 전제되어야만 하는 지향의 방식에 대한 비정립적 의식이 있으며, 이 비정립적 의식으로부터 자기지식이 얻어지기 때문이다. 주체는 P 에 대한 지향적 태도를 가질 때 그것이 바로 그러한 유형의 지향적 태도임을 결정해주는 특정한 방식으로 P 를 의식한다. 예를 들어서 똑같은 사태에 대해서도 주체는 그것을 확신하는 방식으로 믿거나 혹은 그보다는 조금 더 약한 방식으로 믿을 수 있고 혹은 반신반의할 수도 있다. 그리고 의식의 대상에 대해 접근하는 이러한 특정한 방식이 없다면 우리는 그 대상에 대한 의식 자체를 가질 수 없다는 점에서 대상에 대한 비정립적 의식은 정립적 의식이 있기 위한 구성적인 조건이다. 그리고 이처럼 주체가 P 에 대한 정립적 의식을 구성하는 비-정립적 의식을 알기 위해서 추가적인 정보는 필요없으며 단지 기존의 의식에 대한 반성만 하면 된다.

비판과 교훈

보일에 따르면 의식을 지향하는 방식에 대해 주체는 비록 명시적으로는 아니지만 의식하고 있으며, **반성**을 통해서 이러한 의식을 지식으로 만들 수 있다. 문제는 보일이 이러한 반성 작용이 무엇인지에 대해 별다른 설명을 제공하지 않고 있다는 것이다. 반성이 의식의 어떤 작용이기에 사태에 대한 믿음으로부터 기존에 없던 새로운 자기 귀속 믿음을 얻게 해준다는 것인가?

일견 그는 반성이 어떤 과정인지, 혹은 심적 행위인지 형식적으로 명확하게 제시하지는 않지만 군데군데 자신이 생각하는 반성에 대한 그림을 제시하고 있는 것처럼 보인다. 보일에게 반성이란, 대상에 대한 자신의 지향에 이미 포함되어 있는 것 이외의 어떠한 추가적인 정보도 필요로 하지 않는, 지향적 태도의 방식에 적절한 개념을 적용할 수 있는 능력이다(2019: 1032).

보일의 반성 작용이 무엇인지 파악하기 위해, 번의 의도 규칙에 대한 보일의 비판을 살펴보는 것이 도움이 된다. 보일은 의도에 대한 번의

추론주의적 설명을 다음과 같이 비판한다: 번에 따르면 주체는 신빙성있고 안전한 방식으로 〈나는 ø 할 것이다〉로부터 〈나는 ø 하기를 의도한다〉를 추론할 수 있다. 하지만 보일이 보기에 이러한 추론이 신빙성있고 안전한 방식이기 위해서는 '할 것이다'가 단순히 내가 ø 할 것이라는 것을 이론적인 증거를 토대로 예측하는 의미로 사용되는 것이 아니라, 미래 행위에 대한 행위자의 결심을 표현하는 방식으로 사용되는 것으로 이해되어야만 한다. 만일 '할 것이다'가 단지 이론적 증거에 기반한 미래에 대한 예측의 의미로서 사용된 것이라면, 번의 의도-도식은 지식을 산출하지 못할 것이다.[45] 따라서 번의 의도-도식이 의도에 대한 자기지식에 적절한 도식이기 위해서는 적어도 행위자는 자신의 지향성에 대한 특정한 이해에 근거해야 하며, 따라서 의도-도식은 번의 주장처럼 행위자 중립적인 전제로부터 자기지식으로의 추론이 아니다.

번에 대한 보일의 비판에서 알 수 있는 것은, 그가 말하는 반성이 세계의 비-심리적 사태 자체를 근거로 지식을 산출하지는 않으면서도 동시에 대상에 대한 지향적 의식에 포함되는 정보 이외의 어떠한 추가적인 정보를 필요로 하지 않는다는 것이다. 즉, 세계의 사태 자체가 자기지식의 근거가 되지는 않지만, 자기지식의 근거는 세계의 사태에 대한 지향에 암묵적으로 전제되어 있는 바 이상을 필요로 하지 않는다. 또한 보일은 반성을 '비정립적 의식에 대해 심리적 개념을 적절하게 적용하는 것'으로 설명하기도 한다. 정리하자면, 보일이 말하는 반성은 새로운 자료에 대한 고려가 아니라 대상에 대한 지향적 의식이 이미 가지고 있는 정보에 대한 적절한 개념의 적용이다.

그렇다면 주체로부터 개념화되지는 않았어도, 즉 반성되지는 않았어도 지향적 의식이 이미 갖고 있고 또한 주체에 의해서 개념의 적용을 받는 이

---

[45] 이해를 돕기 위해 앤스컴이 든 예시를 들어보자(Anscombe 1957: 1-2). 앤스컴의 예시를 빌려볼 수 있겠다. 우리는 '나는 시험에서 낙제할 것이다'를 우리는 두 가지 방식으로 읽을 수 있는데, 하나는 내가 공부를 너무 하지 않았다는 사실에 근거하는 미래 사실에 대한 예측으로 읽을 수 있고, 다른 하나는 모종의 이유로 시험에서 낙제할 의도를 가지고 있다는 결심의 표현으로 읽힐 수 있다.

정보는 무엇인가? 이는 우리의 의식에 포착되지 않는, 무의식적인 인지 상태일 수는 없다. 자기지식이 되기 위해서는 주체가 정보에 대해 의식적으로 개념을 적용해야 하기 때문이다. 그렇다면 이것은 나의 지향적 상태를 나타내는 언어적 표현으로부터 일종의 유추를 통해서 얻어지는 것인가? 보일은 우리가 예컨대 고양이가 그르렁댐을 지각한다는 것을 어떻게 알 수 있는지를 설명하면서, 지시사 '이것(this)'를 통해서 나타내어 진다는 점에 주목한다. 그리고 반성을 위해서는 이처럼 대상을 표상할 수 있는 특별한 지시적 방식(special demonstrative mode of presentation)과 주체 스스로의 지각적 상태 사이의 관계에 대해서 이해하기만 하면 된다고 말한다. 하지만 여전히 보일은 주체가 대상을 지향하면서 가지는 "비-심적 대상을 표상하는 특별한 방식"을 주체가 어떻게 의식할 수 있는 것인지에 대해서 구체적인 설명을 제시하지 않는다.

혹시 보일이 주장하는 바가, 예컨대 주체 스스로의 지향의 언어적 방식에 대한 인지, 즉 자신이 지시사를 사용하는 방식으로 세계의 사태를 지향한다는 사실에 대한 인지를 근거 삼아 스스로가 지각 상태에 있는 것이라고 판단한다는 것일까? 하지만 주체는 지각적 대상을 꼭 지시사를 통해서 지향하지 않더라도 여전히 자신이 지각하고 있음을 알 수 있다. 지시사를 쓴다는 것은 지각의 방식을 보여주는 좋은 근거지만, 지각의 방식으로 지향한다고 해서 세계의 사태에 대한 지향이 반드시 지시사로 표현될 필요는 없다.

그렇다면 지각 경험의 사례에서 지시사를 사용해서 세계의 사태를 포착하게 만들어주는 우리의 의식 상태는 무엇인가? 자연스러운 대답은 바로 우리가 고양이를 주체의 일인칭 시점에서 **시각 경험**으로서 의식하고 있으며, 시각 경험으로 대상을 의식한다는 것이 곧 〈지각〉을 적용하기 위한 적절한 조건이기 때문이라는 것이다.

그렇다면 믿음의 경우는 어떠한가? 보일은 세계의 사태에 지향할 때 주체가 사태에 대해 지향하는 다양한 방식이 이미 전제가 되어 있음을 다음과 같이 보여준다: 우리는 세계 삼차 대전이 정말로 벌어질 것이라고

생각하는 방식으로 지향할 수도 있지만, 그러한 미래의 가능한 사태에 대해서 단순히 가정하는 방식으로 지향할 수도 있다. 또는 우리는 세계 삼차 대전이 벌어질지 묻는 질문에 대해, 미래의 가능한 사태인 세계 삼차 대전을 지향하면서 '세계 삼차 대전이 벌어질 것이다(There **will be** a third world war)'라고 표현할 수도 있고 '세계 삼차 대전이 벌어질 지도 모른다(There **might be** a third world war)'라고 표현할 수도 있는데, 여기에는 세계 삼차 대전 자체에 대한 고려에 이미 내가 미래 사태를 지향하는 구체적인 방식-확신하는 방식으로 또는 반신반의의 방식으로-이 암묵적으로 전제되어 있다.

나는 사태가 정말 성립할지에 대한 고려에 이미 사태를 지향하는 방식이 암묵적으로 전제되어 있다는 보일의 주장이 설득력있다고 생각한다. 하지만 보일의 생각처럼 지향성의 형이상학적 구조를 굳이 비-정립적 의식과 같은 분명치 않은 개념을 통해서 이해해야 하는지는 불분명하다. 보일도 인정하듯이 비의식적 의식이 모순이라고 생각했던 사르트르와는 달리, 분석철학에서는 이미 지향성과 현상성의 구분(Kim 1998), 현상적 의식과 접근적 의식의 구분(Block 1995), 주체의 의식 상태과 심적 상태의 의식성의 구분 등 보다 명료한 개념적 구분이 정립되어 왔으며 그 결과 비-정립적 의식이 없으면 발생하는 비의식적 의식, 즉 주체에게 의식되지 않는, 하지만 대상을 갖는 의식에는 어떠한 개념적 모순도 없다고 받아들인다. 이처럼 의식에 대한 논의에서 지금의 맥락에서 보다 명료하게 사용될 수 있는 개념들이 있기 때문에 사르트르의 의식 개념을 아이디어를 제공하는 것 이상으로 진지하게 받아들여야 할 필요는 없는 듯하다.

결정적으로, 대상을 지향하는 방식이 있다는 주장과, 그러한 방식이 무엇인지 주체가 알 수 있다는 것은 서로 구분되는 사실이다. 하지만 보일은 믿음에 대해서도 역시 내가 어떻게 나의 지향적 방식을 의식하고 개념을 적절히 적용할 수 있는 것인지에 대해서 실질적인 설명을 주지 않는다. 세계의 대상을 지향하는 방식에 적용되는 반성 작용은 보일이 생각하는 것만큼 직관적으로 분명한 심적 작용이 아니며, 의식의 구조에 대한 보일의 형이상학적인 제안을 받아들이더라도 우리가 어떤 방식으로 대상을

지향하는지에 대한 자기의식이 어떤 심리적 메커니즘을 통해서 얻어지는지에 대한 인식론은 해명되지 않은 채 남아 있다.

3.5. 세계의 대상에 대한 고려와 투명성

나는 2 장을 시작하기에 앞서 번, 모란, 보일을 투명성 이론가로 분류했다. 이는 일차적으로 그들이 공통적으로 에반스(1982)가 말하는 심리적 상태의 투명성에 주목해서 자기지식을 설명하고자 하였기 때문이다(Moran 2001: 61, Byrne 2018: 57-8, Boyle 2019: 1012). 하지만 그들은 심적 상태의 **투명성**과 그 퍼즐이 무엇인지에 대해 서로 이견을 갖고 있기 때문에 서로 다른 자기지식 이론을 펼치고 있다. 4 장으로 넘어가기 전에 이들의 투명성이 어떤 측면에서 다르고 어떤 측면에서 공통점을 갖는지 정리할 것이다.

세 학자가 공통적으로 주목하는 것은 믿음을 자기 귀속할 때 주체가 그러한 믿음의 대상인 세계의 사태에 대해서 생각한다는 에반스의 발견이다. 일인칭 믿음 귀속에 대한 에반스의 관찰에 철학자들이 특별히 관심을 기울이는 까닭은 데카르트 이래로 스스로의 심적 상태에 대해서는 내면을 살펴보는 특별한 능력인 내성을 통해서 대상으로서의 심적 상태를 살펴본 결과 자기지식을 얻는다는 것이 일반적으로 공유되고 있던 직관인 반면, 그가 포착한 자연스러운 현상은 내성의 능력 없이 주체가 스스로의 심적 상태에 대한 자기귀속을 한다고 말해주는 듯 보이기 때문이다. 하지만 실상 에반스는 자기귀속 질문에 대답하는 데에서 나타나는 자연스러운 모습에 대한 서술 이상으로 무언가를 말해주지는 않는다. 세계의 사태에 대해 대답하는 것처럼 바깥의 현상에 주목한다는 것이 어떤 심리적 행위인지, 우리가 어떤 방법으로, 그리고 무엇을 근거로 자기지식을 얻는지에 대해서는 구체적으로 이야기해주는 바가 없으며, 또한 세계의 대상에 대해 고려하는 것이 어떻게 자신의 심적 상태를 묻는 질문에 적절한 대답이 될 수 있는지를

말해주지 않는다. 이 두 공백을 메우는 것이 심적 귀속의 소위 투명성에 주목하는 학자들에게 남겨진 설명 과제다.

번과 모란은 에반스로부터 "지향적 태도에 대한 자기지식을 얻기 위해서 세계의 비-심적 사태에 대한 인지 능력만이 필요하다"는 통찰을 가져간다. 그들은 투명성을 설명하기 위해 믿음을 지향적 태도의 전형으로 삼는다.

번은 'P'를 전제로 하여 결론 '나는 P 를 믿는다'로 이행하는 안전한 추론으로 이해하면 투명성 퍼즐을 해소할 수 있다고 주장한다. 주체가 세계의 대상을 고려하는 여타 다른 방식처럼 P 에 대해 고려하고, 그 결과 스스로에게 P 에 대한 믿음을 귀속시키는 것은 그것이 안전한 추론이기 때문에 외재적으로 정당화된다. 하지만 번은 판단의 메커니즘 문제에 대해서 별다른 설명을 제공하지 않는다. 오히려 그는 사태에 대한 판단에 도달하는 과정에 대한 별다른 설명을 필요로 하지 않는 것이 그의 이론이 가지는 경제성의 덕목이라고까지 말하고 있다.

반면 판단의 메커니즘 문제에 대해서는 별달리 고려하지 않는 번과는 달리, 모란(2001)은 이 과정을 "P 가 참일 좋은 이유를 숙고하는" 과정으로 진술하였다. 그리고 번과 달리 모란은 그렇게 개입하게 되는 'P'로부터 자기귀속 명제인 '나는 P 를 믿는다'로의 이행을 **초월적 가정**으로 정당화한다. 이때 초월적 가정은 단지 이 이행을 신빙성있는 것으로 만들어주는 가정일 뿐만 아니라, 실제로 인식 주체라면 암묵적으로 가정하고 있을 법한, 따라서 이 이행이 어떻게 내적으로 정당화되는지에 대한 물음에 대한 답으로 제공할 수 있는 정당한 가정이다. 하지만 모란에게도 번과 마찬가지로 과연 믿음 이외의 지향적 태도에 대해서 비-심적인 사태의 성립 여부만을 고려하여 자기지식을 얻을 수 있는지에 대한 질문이 제기된다. 번과 달리 모란은 합리적 행위자성을 가정하고, 적어도 합리적 행위자라면 그가 예컨대 P 의 바람직함이라는 비-심적이고 규범적인 세계의 사태에 대한 고려와 개입으로부터 스스로 P 에 대한 욕구를 정당하게 귀속할 수 있다고 주장한다. 합리적 행위자라면 적어도 이유-반응적인 지향적 태도에 대해서는 규범적 사태에 대한 고려를 통해서 정당하게 자기지식을 얻을 수 있다는 것이다.

하지만 일반적으로 욕구 개념과 규범적 사태에 대한 인지가 이처럼 밀접한 것으로 생각되지는 않으며, 다른 지향적 태도 개념들을 모두 규범적 사태에의 개입으로 수렴시키는 것은 지향적 태도 전반에 투명성 절차를 적용하는 댓가로 모든 지향적 태도를 그러한 태도를 가질 규범적 이유에 부합하는 방식으로 갖지는 않는다는 일상적인 경험을 무시하는 결과를 낳는다.

반면 보일은 투명성 절차를 통해 자기지식을 얻을 때 주체는 단순히 비-심적 사태만이 아니라, 비-심적인 세계의 사태에 대한 지향의 방식, 즉 행위자 자신의 심적인 측면까지도 참조한다고 본다는 점에서 두 학자와 차이를 보인다. 보일에 따르면 세계의 사태를 지향할 때 주체는 이미 특정한 방식으로 지향하며, 스스로가 그러한 특정한 방식으로 지향한다는 것은 사태에 대한 지향과 동시에 이미 암묵적으로 의식되고 있다. 그리고 자기지식을 얻는 것은 반성을 통해 이러한 지향의 방식에 대한 의식을 명시적으로 의식하는 것이기 때문에 세계의 사태에 대한 지향적 의식에 포함되는 것 이외의 추가적인 정보를 필요로 하지 않는다. 하지만 그는 지향적 의식에 포함되는 정보가 무엇이기에 주체에게 명시적 믿음은 아니면서도 의식되고 있는지에 대해서는 말하지 않는다. 자기지식의 근거가 세계의 사태 자체라기보다는 세계의 사태를 지향하는 데 동반되는 주체의 의식의 특정 측면이라고 주장하지만, 그래서 구체적으로 그러한 의식의 특정 측면이 무엇이고 어떤 과정으로 주체가 이 특정 측면으로부터 자기지식을 얻어낼 수 있는지에 대한 인식론을 결여하고 있는 것이다.

요컨대, 모란과 번의 이론은 주체가 세계의 비-심리적 사태에 대해 고려하고, 그렇게 고려된 비-심리적 사태에**만** 근거해도 자기지식을 얻을 수 있다고 주장한다. 하지만 믿음을 제외한 대부분의 지향적 태도에 대한 앎은 그 지향적 태도 개념의 특성상 비-심리적 사태에 대한 개입으로는 얻어질 수 없는 것 같다. 심지어 참에 개입하는 태도에 대한 자기지식에 대해서 조차도, 자신이 완전히 확신하거나, 꽤나 강하게 믿거나, 반신반의한다는 것을 주체가 구분할 수 있다는 사실은 비-심적 사태에 근거하는 것만으로는 설명되기

어려워 보인다. 반면 보일에 따르면 고려된 비-심리적 사태와 함께 비-심리적 사태에 대한 고려에 동반하는 지향의 방식에 대한 의식까지도 자기지식의 근거가 된다. 보일의 이론은 투명성 이론이면서도 설명되는 자기지식의 영역을 다양한 지향적 태도에 대한 자기지식으로 확장시킬 수 있는 한 가지 개연적인 방향을 제공하고 있지만, 구체적으로 지향의 방식에 대한 의식이 무엇인지, 이로부터 어떻게 정당한 자기지식을 이끌어낼 수 있는지에 대한 설명은 결여하고 있다.

제 4 장 자기지식에 대한 현상적 투명성 이론

4.1. 자기소외적 믿음과 일인칭 시점에서의 경험

자기지식의 일인칭 특권성을 설명하는 대안적 견해를 제시하기에 앞서, 3 장의 1 절에서는 모란과는 다른 방식으로 자기소외적인 심적 상태를 구분할 수 있는 한 가지 새로운 기준을 제시하겠다. 이 기준은 앞으로 제시할 특유한 방법으로 자기지식을 얻을 수 있는 심적 상태를 구분하는 기준이 될 것이다.

모란은 규범적 이유가 가리키는 이상적 태도와 실제 태도 사이의 괴리를 '자기소외성'으로 표현하였다. 모란에 따르면 자기지식과 타인에 대한 지식 사이의 불균형의 근원은 내가 나의 심적 태도에 자기소외되지 않는다는 점에 있다. 그리고 자기소외는 심적 상태의 투명성과 다음의 측면에서 긴밀하게 연결되어 있다: 행위자가 심적 상태의 투명성을 사용해서 자신의 믿음을 알게 된다면, 그는 자신의 믿음을 단순히 이론적 관점으로부터 아는 것이 아니라, 일인칭 시점에서 규범적 이유를 따라 자인하는 것으로서 알게 되며, 따라서 그 믿음으로부터 자기소외되지 않는다.

모란이 맞다면 행위자는 그렇게 믿을 규범적 이유에 대한 인식을 통해 사태의 참에 도달할 수 있고, 합리적 행위자가 가정할 권리를 가지는 초월적 가정을 통해서 특권적 자기지식을 얻는다. 하지만 나는 다음의 이유에서 모란의 입장에 동의하지 않았다. 첫째, 지향적 태도를 가질 규범적 이유에 따라서만 지향적 태도를 형성하는 극도로 합리적인 행위자가 아닌 이상, 믿음을 제외한 다른 지향적 태도들에 대해서 그러한 태도를 가질 규범적 이유를 아는 것은 자신의 실제 지향적 태도를 아는 것에 별달리 기여하지 않는다. 둘째, 설령 합리적 행위자를 가정하더라도, 행위자는 지향적 태도를 가질 규범적 이유를 실제로 살펴보지 않더라도 자신이 어떤 태도를 가지는지

알 수 있다. 즉, 행위자는 자신이 어떤 지향적 태도를 갖고 있는지 알기 위해서 해당 지향적 태도를 가질 규범적 이유를 실제로 살펴볼 필요가 없다.

짚고 넘어가야할 부분은, 모란의 자기지식 이론을 거부하는 입장에서도 합리적 주체라면 규범적 이유에 따라 심적 상태를 가진다(혹은 가져야만 한다)는 모란의 논지에 반대할 이유가 전혀 없다는 것이다. 논점은, 모란의 이론은 실제로 어떤 믿음을 갖고 있어야 합리적인가, 그리고 합리적인 사람이라면 어떻게 믿음을 형성해야 하는가를 설명해주는 이론일 수는 있지만, 실제로 자신이 어떤 심적 상태를 갖고 있는지를 알기 위해 주체는 규범적 이유를 실제로 살펴볼 필요가 없다는 점에서 일상적인 자기지식 이론으로 부적절하거나 혹은 사태에 대한 판단에 이르는 일련의 메커니즘에 대해서 어떤 실질적인 설명을 제공하지 않기 때문에 자기지식의 이론으로 부족하다는 것이다.

규범적 이유를 아는 것이 심적 상태의 투명 절차를 밟는 것의 핵심이라는 모란의 주장을 거부한다면, 지향적 태도의 투명성은 자기지식을 얻는 과정에 어떤 관계가 있는가? 나는 투명성 절차와 자기지식 사이의 관계를 이해하는 또 다른 방식에 대한 한 가지 힌트를 다양한 현상적, 감각적, 지각적 상태에 대한 자기맹의 예시들을 생각해 봄으로써 얻을 수 있다고 생각한다.

현상적, 감각적 상태에 있어서 일인칭적 자기지식이 결여된 상황의 예시들, 즉 자기맹의 상황을 생각해보자. 주체가 특정 감각적 상태에 대하여 자기맹의 상태에 놓여있을 때, 그는 왜 자기맹의 상태에 있다고 여겨지는가? 의식과 현상성의 논의에서 자주 사용되는 맹시 현상(blindsight)의 예시를 들어보자. 맹시는 일차 시각 피질(primary visual cortex)의 일부가 파괴된 환자에게서 발견되는 현상으로, 맹시 현상을 겪는 환자는 시야의 일정 부분이 의식되지 않는다고 보고하면서도 의식되지 않는 시야에서 벌어지는 일을 지각하고 있다고 볼 수밖에 없는 반응을 보인다. 예컨대 환자가 시야의 결손을 호소하는 부분에 장애물을 놓더라도 환자는 높은 확률로 그 장애물을 마치 지각을 가진 사람처럼 피해간다거나, 혹은 수평선과 수직선을 놓고 둘 중에 하나를 선택하기를 강요했을 때 지각 상태에 있지 않다는 것으로는

설명하기 어려울 높은 확률로 정답을 맞추는 모습을 보인다는 것이다. 의식에 관심을 갖는 철학자들은 맹시 현상에 대해서 예컨대 현상적 의식은 없지만 인지적 의식은 있는 상태라거나, 고차 의식을 갖지 않는 현상적/감각적 상태에 있다는 방식으로 설명하고자 하였다.[46] 물론 맹시현상의 메커니즘은 아직까지 명확하게 밝혀진 바 없으나, 앞선 1 장에서의 논의를 기반으로 맹시현상을 현상성에 대한 대면은 결여하고 있지만 기능성은 예화하고 있는 지각 상태의 일례로 들 수 있을 것이다. 즉, 맹시의 상황에서 행위자는 일반적으로 그러한 지각 상태를 가질 때 갖추고 있는 기능적/성향적 속성을 갖고 있지만, 일반적으로 그러한 지각 상태를 가질 때 갖추고 있는 현상성은 대면하고 있지 않다..

어떤 심적 상태의 기능성과 현상성이 공외연적이라는 것은 형이상학적으로 필연적인 사실은 아니기 때문에, 다른 복합적인 심적 개념에 대해서도 그것의 기능성은 예화되어 있지만 현상성을 대면하고 있지는 못한 상황을 상상해볼 수 있다. 예컨대 고통을 겪을 만한 상황에서 고통을 겪는 듯한 반응을 보이면서도 고통 특유의 질적인 경험을 의식하고 있지는 못하는 경우, 가려움을 느끼면 보일 행동을 하면서도 가려움 특유의 질적인 경험을 의식하고 있지는 못하는 경우는 마치 맹시 현상처럼 기능성을 예화하는 심적 상태를 가지면서도 그러한 심적 상태의 특유의 현상성을 대면하지는 않는 상태이다.[47]

그리고 바로 이처럼 어떤 복합적 속성 M 의 기능적 속성은 예화하면서도 현상성은 예화하지 않는 심적 속성 M 의 사례 M_1 을 가지는 경우가, 주체가

---

[46] 맹시에 대한 전자의 설명은 Block (1995)을, 후자의 설명은 Carruthers (2000)를 참고하라.

[47] 곧 이어서도 언급하겠지만, 〈고통〉이 그 현상성이 본질인 개념이라면 고통에 대한 자기맹은 개념적으로 불가능할 것이다. 하지만 〈고통〉을 기능적 개념으로 생각하는 학자들도 있으며(e.g. Shoemaker 1994) 만일 〈고통〉이 기능적 개념이라면 고통에 대한 자기맹의 불가능성은 〈고통〉 개념 자체에서 따라나오지는 않는다. 여기서는 〈고통〉 개념에 대한 특정한 형이상학적 전제를 갖지는 않겠다.

M_1에 있으면서 동시에 자신의 심적 상태 M_1으로부터 자기소외되는 경우이다. 즉, 이 경우들은 주체가 삼인칭적 시점에서는 M 을 귀속할 만한 행동적인 증거들을 보여주면서도, 그것을 일인칭 시점에서 주관적인 현상적 경험으로 **의식**하지는 못하는 경우이다.

나의 제안은, 이 논의를 지향적 태도로 확장시켜서 지향적 태도가 기능성은 예화하면서도 현상성은 예화하지 않거나 혹은 현상성에 주의를 기울이지 못해 심적 개념을 적용하지 못하는 상황을 자신의 지향적 태도에 대해 일인칭 시점에서의 자기지식을 얻지 못하는 자기소외의 상황으로 생각하자는 것이다. 무언가를 욕구하는 것처럼 행동하지만 스스로는 욕구함이 가지는 특유의 느낌을 대면하지 못하거나, 어떤 대상에 대해서 분노 혹은 두려움을 느끼는 사람처럼 행동을 하지만 스스로는 그러한 분노 혹은 두려움 특유의 느낌을 의식적으로 대면하지 못하는 사람은, 삼인칭 시점에서는 그 대상을 욕구하거나 그 대상에 대해 분노 혹은 두려움을 느낀다고 여기는 것으로 보이면서도 일인칭 주체의 시점에서 그러한 욕구, 분노, 두려움을 특권적 방식으로 자기 귀속하는 것에는 실패한다.

이 사례들은 모란의 자기소외적 사례와 어떻게 다른가? 그는 특정 태도를 가질만한 좋은 이유가 없다고 생각하면서도 실제로는 그러한 지향적 태도를 가지는 주체를 자기소외의 예시로 든다. 그가 드는 예시는 일인칭 시점에서 상대방에게 화를 낼 규범적 이유를 찾을 수 없음에도 그에게 화를 내는 상황이다.[48] 나는 이처럼 규범적 이유가 가지길 요구하는 태도와 실제 주체가 갖는 태도 사이의 불일치도 일종의 자기소외의 상황일 수 있다고 생각한다. 만일 모든 이유를 생각했을 때 매일 아침 운동을 가는 것을 욕구할 만하다고 생각하는 사람이 그럼에도 불구하고 의지박약 때문에 운동을 욕구하지 않는다면, 그는 아마도 이상적 상황과 자신의 현실적 욕구 사이에서 괴리를 느낄 것이며, 자신이 아는 이유가 가리키는 대로 심적 태도를 갖지 못하는 스스로가 그 자체로 완결적이지 않다고 느낄 것이다. 하지만 모란의

---

[48] Moran (2001: 84-86)

자기소외는 일인칭 시점에서 자기지식을 얻는 것과는 별 상관이 없다. 모란의 화를 내는 사람 예시에서도, 비록 그는 자신이 화를 낼 이유를 잘 모른다는 점에서 모란의 의미에서는 자기소외적이지만, 그럼에도 화의 현상성을 경험하고 있기 때문에 자신이 화가 난다는 자기지식을 얻는 데에는 아무런 문제가 없다. 매일 아침 운동을 가는 것을 욕구할 만하다고 생각하면서도 운동을 욕구하지 않는 사람 또한 그럼에도 불구하고 자신이 운동을 욕구하지 않는다는 것을 직접적으로 알 수 있다. 그가 아침 운동에 대해 욕구의 느낌을 경험하지 않기 때문이다. 물론 자신이 화를 내는 동기적 이유조차 스스로 모르는 경우에 그 주체의 태도로부터 어떠한 이유-반응성을 찾기 어렵다는 점에서 해당 주체는 그다지 합리적이라고 할 수 없을 것이다. 그럼에도 불구하고 주체가 대상에게 화를 느끼고 있는 이상, 주체는 일인칭 시점에서 자신이 화가 났다는 것을 알기에 충분한 위치에 놓여 있다. 그가 상대방에 대해서 화남 특유의 경험을 하고, 이 느낌이 개념 〈화〉를 적용할 수 있는 그런 느낌임을 알고 있는 한 그는 자신이 화가 났다는 것을 분명히 알 수 있기 때문이다. 만일 어떤 사람이 자기지식을 정당하게 얻을 수 없을 만큼 자기소외되었다면, 그가 처한 상황은 모란의 자기소외적 상황처럼 화를 낼 규범적 이유를 알지 못하면서도 대상에 대한 화남을 예화하는 상태가 아니라, 화의 현상성을 경험하지 못하면서도 마치 화가 난 사람처럼 말하고 행동하는 상황일 것이다. 즉, 화의 현상성을 대면하거나 혹은 이에 주목하지는 못하지만 기능성은 갖고 있는 그러한 심적 상태를 가지는 경우다. 이러한 상황에 놓여있지 않은 이상, 비록 주체는 자신이 그러한 지향적 태도를 가질 좋은 이유를 알지는 못하더라도 자신이 어떤 지향적 태도를 갖고 있는지는 정당하게 알 수 있다.

이는 믿음에 대해서도 적용될 수 있다. 어떤 주체는 삼인칭 시점에서는 마치 P 가 참이라고 생각하는 사람이 보일 법한 반응들을 보이면서도 일인칭적 시점에서는 P 를 참이라고 의식할 수 없는 상황에 놓여 있다. 앞서 거틀러(2011)가 든 예시처럼 '나쁜 일이 있으면 소금을 뿌려야 한다'는 것이 결코 참이 아니며 미신이라고 생각하면서도 부지불식간에 나쁜 일이 생길

때마다 소금을 찾아 뿌리는 사람이 있다면, 그는 나쁜 일이 있으면 소금을 뿌려야 한다는 규범적 사태에 대한 믿음의 기능적 상태는 가지면서도 그에 동반하는 현상성에 주목하지는 못하는 상황에 처해 있으며 그렇기 때문에 일인칭 시점에서 투명성 절차를 통해 믿음을 알 수 없는 자기소외의 상태에 빠져 있다고 할 수 있다.

4.2. 믿음 속성 특유의 현상성

앞 절에서 나는 일인칭의 시점에서 자기지식을 얻을 수 없는 상황을 주체가 특정 심적 속성 M 의 사례 M_1 을 가지고 있을 때 M_1 이 M 의 기능성은 예화하고 있지만 현상성에는 주목하지 못하는 상황이 바로 자신의 심적 상태 M_1 으로부터 자기소외된 상황이라고 분석하였다. 이 절에서는 이 분석이 믿음에도 적용되기 위해서는 믿음 태도 특유의 현상성이 있어야 함을 주장하고 믿음 태도 특유의 현상성이 있다고 생각할 이유들을 제시할 것이다.

우선 어떤 심적 상태에 대해 현상성을 통한 자기지식을 얻을 수 없고 얻을 수 있는지를 살펴보자. 현상성이 그 심적 속성을 구성하는 전부인 경우 현상성을 예화하지 못하는 상황은 곧 그 심적 속성의 사례를 갖지 않는 그런 상황이기 때문에 자기소외가 발생할 수 없다. 예를 들어 차머스가 말하는 **순수 현상적 개념**은 경험의 순수한 내재적 성질에 의해서 그 개념의 구성요소가 결정되는 심적 개념이다.[49] 만일 순수 현상적 개념이 정당한 심적 개념이라면, 순수 현상적 속성은 순수히 현상성만으로 구성되어 있으므로, 순수 현상적 속성의 현상성이 예화되지 않는다면 주체는 필연적으로 순수 현상적 속성의 사례를 가질 수 없을 것이다. 그렇다면 순수 현상적 상태에 대한 자기소외는 불가능할 것이다. 마찬가지로 현상성이 심적 속성의 본질적 속성인 경우에도 유관한 의미에서의 자기소외는 발생할 수 없다. 만일 고통이 그것의 현상성을 본질로 가지는 속성이라면 고통 상태를 가지는

[49] Chalmers (2003: 220-272)

주체는 반드시 고통의 현상성을 경험하고 있어야만 하므로, 고통의 현상성을 경험하고 있지 않는 주체는 필연적으로 고통 상태에 놓여 있을 수 없을 것이다. 즉, 앞서 말한 고통의 자기맹 상태는 필연적으로 불가능할 것이다. 고통의 자기맹 상태가 가능하려면 반드시 고통 특유의 현상성이 고통의 본질이 아니어야 하며, 〈고통〉은 기능적 측면과 현상적 측면을 모두 본질로 가지는 개념이어야 할 것이다.[50]

또한 만일 그 심적 속성이 그 개념의 내포 안에 현상성을 포함하지 않는다면 유관한 의미에서의 자기소외는 불가능할 것이다. 예컨대 어떤 작업에 숙련됨이 그 예시가 될 수 있다. 어떤 작업에 숙련된 상태를 작업이 목적한 바를 성취하는 능력을 가지는 것으로 이해한다면 〈숙련됨〉의 본질은 기능성이다. 그리고 무언가에 숙련되는 것에는 고유한 현상성이 없는 것처럼 보인다. 그렇다면 주체는 자신의 숙련됨로부터 위와 같은 의미에서 자기소외적일 수 없다. 숙련됨에 동반되는 현상성이 없기 때문에 주체는 현상성에 근거해서 자신의 숙련됨을 알 방법이 없기 때문이다.

따라서 어떤 심적 상태가 나에게 자기소외적이기 위해서는 다음의 두 가지 조건이 만족되어야 한다. 첫째, 그 심적 개념은 현상성을 본질로 가지는 개념이 아니어야 한다. 자기소외의 상태는 심적 상태가 예화되면서도 동시에 현상성에 주목하지는 못하는 상태이기 때문에, 현상성을 주목하지 못하는 상황에서도 여전히 그 심적 상태가 예화되는 것이 가능해야 한다. 둘째, 해당 심적 개념이 가리키는 심적 상태의 특유의 현상성이 있어야 한다. 일인칭 시점에서 현상성에 대한 대면만으로 자신이 그러한 상태에 있다는 것을 알 수 있어야 하기 때문이다. 요컨대, 그 심적 개념은 복합적 심적 개념이어야만 한다.

---

[50] 고통의 자기맹을 사고실험의 재료로 쓴 슈메이커(1994)는 〈고통〉을 기능주의적 관점에서 규정한다. 하지만 〈고통〉이 현상적 개념인지, 혹은 복합적 개념인지를 확실하게 결정하는 것은 이 글의 관심에서 벗어난다.

나의 목표는 지금까지의 논의를 확장시켜, 현상성의 결여를 통하여 믿음에 대한 자기소외적이지 않은 상황을 분석하는 것이다. 따라서 위에서 제시한 자기소외성의 분석이 믿음에 대해서도 적용되기 위해서는 마찬가지로 위의 두 조건이 만족되어야 한다. 따라서 첫째, 믿음에 대한 자기소외가 가능해야 하므로 적어도 믿음은 현상성을 본질로 갖는 속성이 아니며, 둘째, 현상성에 대한 대면을 통해서 그것이 심적 속성 믿음의 예화임을 알아야 하므로 〈믿음〉과 긴밀하게 연결되는 현상성, 즉 믿음-특유의 현상성이 있어야 한다. 그중 첫째 가정은 믿음이 현상성을 본질로 가진다고 생각하는 학자들이 오늘날에는 거의 없다는 점에서 문제적이지 않다. 하지만 둘째 가정, 즉 믿음-특유의 현상성에 대해서는 철학자들 사이에서 동의된 바가 없을 뿐더러 논의도 많이 되지 않은 것처럼 보인다.[51]

현상성 특유의 주체의 경험에 의존하는 특징과 그로 인해 공적 언어로 표현하기 어렵다는 특징은, 어떤 현상성이 정말로 존재하는가에 대한 엄밀한 논의를 어렵게 만든다. 이는 나는 단지 스스로의 경험만을 경험할 수 있을 뿐이며 상대의 경험은 결코 경험할 수 없기 때문에 나와 상대의 경험이 정말로 같은 현상성을 예화하는지 확인할 길이 없기 때문이다. 하지만 그럼에도 불구하고 우리는 공통적으로 경험하는 고통, 가려움, 즐거움, 슬픔과 같은 현상성들에 대해 이야기한다. 상대가 아픔을 느낀다고 할 때, 당연하게도 그가 내가 느끼는 아픔과 같은 유형의 현상성을 경험한다고 생각한다. 우리는 자연스럽게 어떤 현상적 사례들을 묶는 현상성의 유형이 있음을 받아들이는 것 같다. 이러한 직관을 받아들인다면 태도 특유의 현상성 유형을 무리없이 가정할 수 있으며 이 논의에서는 그러한 현상성 유형이 있음을 가정하겠다.

---

[51] 주의할 것은, '믿음-특유의 현상성'이 어떤 내용 토큰을 포함하는 믿음-개별자 각각에 특유한 현상성이 있다는 현상적 지향성의 주장을 의미하는 것은 아니라는 것이다. 여기서 믿음-특유의 현상성은 믿음 상태와 형이상학적으로 동반되는 그러한 현상성이다.

그렇다면 믿음-특유의 현상성은 있는가? 믿음-특유의 현상성이 있는 지 아직까지 철학적 맥락에서 풍부하게 다루어지지는 않았다. 이는 믿음 뿐만 아니라 지향적 태도 일반에 대해서 그러하다. 여기에는 몇 가지 이유가 있을 수 있는데, 우선 지금까지 그 현상성을 근거로 자기지식을 얻을 수 있는 대상이 주로 지각적 상태나 지각적 상태에 한정되어서 이루어졌기 때문이다. 믿음과 같은 지향적 태도를 현상성에 근거하여 설명하려는 시도는 호건과 티엔슨(2002)을 기점으로 이루어지는 현상적 지향성 논의에서 이루어지기 시작했으며 아직 그 역사가 길지 않다. 현상적 지향성 이론의 기점으로 여겨지는 그들의 2002 년 논문에서 그들은 명제 태도의 **내용** 뿐만 아니라 명제 태도의 **태도 유형**에도 그것의 특유한 현상성이 있다고 주장한다.[52] 호건과 티엔슨에 따르면 모든 지향성은 현상성으로 환원될 수 있으며, 이때 지향적 태도의 현상성은 일반적으로 말하는 감각적 현상성들과 구별되면서도 지향적 태도를 결정할 수 있을만큼 풍부하다는 것이다.[53] 예컨대 어떠한 감각적 현상성을 동반하지 않으면서 토끼가 꼬리를 가지고 있다고 생각하는 상황을 떠올려보자. 이때 토끼가 꼬리를 가지고 있다는 명제(혹은 토끼의 꼬리를 가지고 있는 사태)는 이 생각의 내용이다. 그들이 제시하는 논변은 우리는 분명 토끼의 분리되지 않은 부분들의 모음이 꼬리-부분을 가지고 있다고 있다는 생각을 떠올리는 것과 단지 토끼가 꼬리를 가지고 있다는 생각을 떠올리는 것을 구분하며, 만일 이것이 맞다면 두 생각의 내용을 구분하게 해주면서도 감각적 현상성으로 분류될 수 없는 내용 특유의 현상성이 있다는 것이다. 더 나아가 그들은 지금 의식되고 있는 '토끼가 꼬리를 가지고 있다'는 생각의 내용에 대하여 주체는 자신이 그것을 단지 생각하고 있는지, 그것을 지금 의심하고 있는지, 혹은 그것을 믿고 있는지와 같은 태도 유형들을 구분할 수 있다고 주장하는데, 이는 그러한 태도

---

[52] Horgan and Tienson (2002 : 522-523)

[53] 이외에도 현상적 지향성 이론의 핵심적인 논제를 알기 위해서 Pitt (2004)의 논문이 도움이 된다.

유형들이 만일 주체가 주의를 기울인다면 놓칠 수 없는 그러한 현상적 경험의 측면을 가지고 있기 때문이다.[54] [55]

이 글에서 나는 어떤 지향적 태도 토큰이 그와 같은 유형이면서 동시에 다른 내용을 가지는 지향적 태도 토큰과 구분되는 고유한 현상성을 가지고 있다는 현상성 지향성론자들의 주장에까지 개입하지는 않는다. 이는 이것이 그 자체로 별도의 많은 정당화를 요구하는 작업이기도 하지만, 설령 현상적 지향성 이론이 성공적이라고 해도 나의 주장에서 여전히 심적 내용이란 여전히 내성을 통해서 결정되는 것이 아니라 투명성 절차를 통해서 결정되는 것이기 때문이다. 이 글의 입장에 따르면, 주체는 스스로의 지향적 태도를 알기 위해서 내 머릿속에 떠다니는 생각의 내용을 구성하는 현상성을 내성을 통해서 의식할 필요가 없다. 다만 투명성 절차를 통해서 고정된, 내가 지향하는 사태를 알고 있으면 되는 것이다.

자기지식을 설명하기 위해 태도 유형 특유의 현상성이 있다는 주장을 구체적으로 전개한 학자는 현상적 지향성 논의 바깥에서도 많지 않다. 지향적 태도에 대한 자기지식을 설명하기 위해서는 자신이 갖는 태도-예컨대 믿음인지, 의문인지, 가정인지-를 구분하는 것이 선행되어야 함에도 태도

---

[54] 지향성을 현상성에 근거짓는 것이 현상적 지향성 이론가들의 궁극적 목표이기 때문에, 그들의 주된 관심사는 생각의 내용이 다른 감각적 상태와 유형적으로 구분되며 더 나아가 내용들 사이에서도 구분되는 그러한 고유한 현상성을 가지고 있다는 것을 보이는 데 있다. 그렇기 때문에 이후 그들은 내용 특유의 그리고 고유의 현상성이 있다는 것을 보이는 것에 집중하며 태도 유형 특유의 현상성이 있다는 주장을 특별히 구체적으로 개진하지는 않는 듯하다.

[55] 이것이 그들이 태도 유형 특유의 현상성이 있다는 것을 암묵적으로 그럴 듯한 것으로 받아들이고 있기 때문에 그러한 것인지, 혹은 내용 특유의 그리고 고유의 현상성을 보이는 것이 더 시급한 이론적 과제이기 때문에 미뤄두고 있는 것인지는 분명치 않다. 다만 현상적 지향성 이론이 궁극적으로 지향적 태도를 현상성으로 환원적으로 설명하려는 목적을 가지고 있다면 생각의 내용 뿐만 아니라 태도 유형에 특유한 현상성이 있다는 것을 받아들여야 하는 것은 분명해 보인다.

유형을 구분하는 문제가 구체적으로 다루어지지 않은 까닭은, 지향적 태도에 대한 자기지식을 설명하는 이론들이 제각기 다른 목표를 두고 제시되었기 때문이다. 예컨대 합리성의 가정에 기댄 초월적 논변에 근거한 이론들은 자기지식 획득의 구체적인 메커니즘에 대해 개입하고 있지 않다.[56] 또는 내적 감각 이론은 자기지식을 얻는 구체적인 메커니즘을 설명하는 것에 그 목적을 두고 있지만, 그들의 관심사는 믿음에 집중하여 믿음의 내용을 인과항이 될 수 있는 표상으로 설명하고 그러한 표상을 어떻게 내적 감각을 통해 자기귀속 믿음이 표상하게 만드는지, 즉 내용에 대한 내성이 어떻게 이루어지는지에 초점이 맞추어졌기 때문에 태도 유형을 구분하는 것은 큰 관심사가 아니었다.[57]

하지만 호건과 티엔슨과는 다른 동기로부터, 골드만(1993)은 일상적으로 주체가 스스로에게 다양한 지향적 태도를 내성적으로 귀속하는 현상을 설명하기 위해서는 태도 특유의 현상적 경험을 가정해야 한다고 주장하였다. 그는 행위자가 일상적으로 심적 속성을 정당하게 자기귀속하는 방식을 다음과 같이 설명한다. 골드만에 따르면, 유능한 언어 사용자라면 심적 용어들을 특정한 범주 표상(category representation)들과 연결하고 그러한 범주 표상이 현재 주체에 활성화되어 있는 사례 표상(instance representation)과 짝짓는 과정을 통해서 심적 속성을 자기귀속한다. 이때 범주 표상이란 유능한 언어 사용자가 각각의 심적 용어에 알맞게 갖고 있는 의미론적 표상이며, 사례 표상은 개인의 상태에 대한 활성화되어 있는 정보이다. 주체는 이러한 활성화된 사례 표상을 특정 범주 표상과 짝지으며, 이로부터 범주 표상과 연결된 심적 단어를 일인칭 술어화함으로써 심적 속성을 자기귀속하게 된다.

골드만은 심적 용어와 연결되어 있는 범주 표상과 이와 짝지어지는 사례 표상은 기능적 역할에 대한 표상일 수 없고 신경 상태에 대한 표상일 수도

---

[56] Burge (1996), Shoemaker (1994)

[57] 예를 들어, Nichols and Stich (2003)

없으며, 남는 것은 오직 개인 경험의 질적인 측면, 곧 현상성이어야 한다고 주장한다. 우선 심적 상태를 자기귀속하는 사람은 자신의 심적 상태를 야기한 원인 혹은 뒤따르는 결과를 모르면서도 자신이 어떤 심적 상태에 있는지 알 수 없으므로, 심적 용어는 기능적 역할을 표상하는 것일 수는 없다. 이에 대해 만일 행위자가 바로 이 상태의 원인이나 결과를 알 필요는 없고 다만 심적 상태와 연관된 가능한 인과 관계를 알아야 된다고 주장한다면, 이러한 요구는 심적 상태와 연결된 다른 심적 상태들, 그리고 그 심적 상태들과 연결된 또 다른 심적 상태들에 대한 가정적 인과 관계에 대한 표상을 가져야 한다는 불가능한 요구로 이어진다. 따라서 행위자가 심적 상태를 자기귀속하기 위해 심적 상태의 기능적, 관계적 속성을 알아야 된다는 주장은 일상적으로 이루어지는 심적 상태의 정당한 자기귀속을 설명해주지 못한다. 그렇다면 자기귀속을 위한 범주 표상은 심적 상태와 연관된 비-관계적 내재적인 속성에 대한 것이어야 할텐데, 이 중 행위자 레벨에서 주체는 신경적 속성에 대해서 접근하지 못하기 때문에 유일하게 남는 것은 현상성 뿐이다. 따라서 주체는 자신의 현상성을 심적 용어의 표상으로 삼아 자기귀속을 한다.

골드만의 논변을 단순하게 요약하면 다음과 같다. 우리는 일반적으로 심리적 속성을 정당하게 자기귀속한다. 이 과정은 현재의 사례 표상을 특정 심적 개념의 적절한 범주 표상과 짝짓고 해당 심적 개념을 스스로에게 귀속시키는 과정으로 이루어진다. 이때 이 범주 표상은 주체가 의식적으로 접근할 수 없는 기능적 역할 혹은 신경적 속성에 대한 것일 수 없으며 유일하게 남는 적절한 후보는 현상성이다. 그리고 그의 논변은 일반적으로 이루어지는 심적 귀속이 정당하다는 전제로 이루어진다. 우리가 일상적으로

얻는 자기지식이 정당함을 설명하기 위해서는 심적 개념 특유의 현상성을 이용한다고 봐야 한다는 것이다.[58]

궁극적으로 자기지식을 설명하는 이론적 방향성 측면에서 호건과 티엔슨, 그리고 골드만은 서로 다른 입장에 서있다. 하지만 그들은 공통적으로 지향적 태도에 대해서 주체가 의식하기 위해서는 태도 특유의 현상성이 있어야 된다고 주장하는데, 이는 우리가 일상적으로 스스로가 무엇을 믿는지, 의심하는지, 욕구하는지에 대한 의식적으로 안다는 사실을 설명하기 위해서는 내성된 태도가 어떤 태도 유형인지 주체가 의식할 수 있도록 태도 특유의 현상성이 있어야 하기 때문이다.

혹자는 믿음이 형성되고 수정되는 인지 작용이 인과적인 과정이며, 따라서 내가 어떤 태도를 취하는지는 그 태도를 갖게 된 원인을 의식적으로 알게 되기 때문에 구분할 수 있는 것이라고 주장할 수도 있겠다. 하지만 우리는 자신이 무엇을 믿는다거나 혹은 긴가민가 의심한다는 것을 그것의 원인에 대한 의식에 근거해서 알게되는 것 같지는 않다. 믿음이 형성되었다거나 혹은 폐기되어 더 이상 믿지 않는다는 것이 인과적인 인지 작용의 결과라고 하더라도, "주체의 관점에서" 분별하기 위해서는 그것이 주체에게 의식되어야만 한다. 예컨대 세계 삼차 대전이 발생할 것인지를 생각해본 결과 어떤 주체 S 가 '세계 삼차 대전이 발생할 것이다'를 완전히 확신하는 정도로 믿게 되었고 이를 스스로 의식하게 되었다고 하자. 그런데 어느날 S 는 신문에서 미국과 중국의 관계가 개선되고 있다는 정보를 얻었다. 이 새로운 정보는 내 안의 인지 과정에 무의식적이나 인과적인 방식으로 영향을 주어 내가 의식하지 못하는 동안 '세계 삼차 대전이 발생할 것이다'에 대한

---

[58] 골드만은 태도 유형을 식별할 수 있게 해주는 현상성이 있다는 주장에 명시적으로 개입한다. (1993, 23-5) 이후 그는 태도 유형을 식별할 수 있게 하는 요소가 필요하다는 주장을 견지하면서도 내적 감각 이론을 통해 자기지식의 인과적 메커니즘을 설명하면서, 태도 유형을 식별하게 하는 요소로 가장 가망있는 것으로 인과항이 될 수 있는 신경적 속성을 제시한다. 하지만 그는 여전히 가능한 입장 중 하나로 현상성을 통한 태도 유형의 식별을 제시한다. Goldman (2006: Chapter 9)

나의 주관적 확신 정도를 꽤나 낮추었다. 몇 달 후 나는 인지적 과정에서의 원인-미국과 중국의 관계가 개선되고 있다는 정보를 얻음-을 인식하지는 못하면서도 전과는 달리 '세계 삼차 대전이 발생할 것이다'를 믿지 않게 되었다는 사실을문득 의식하게 되었다. 그리고 자신이 '세계 삼차 대전이 발생할 것이다'를 믿었던 몇달 전과는 달리 더이상 세계 삼차 대전이 발생할 것이다'를 믿지 않는다는 것을 알게 되었다. 이 사례에서 지향적 태도의 내용은 동일하며, 달라진 것은 그 내용의 참에 개입하는지의 여부이다. 그리고 S 는 자신의 인지적 체계 내에서 원래 갖고 있던 믿음을 폐기하게 만든 원인에 대해서는 인지하지 못하면서도 같은 명제에 대해 믿었다가 더 이상 믿지 않게 되었다는 것을 의식하고 있다. 이때 두 지향적 태도의 차이를 주체가 의식할 수 있는 근거는 믿음의 내용의 차이도, 하위-인간적(sub-personal) 인지 과정도 아니다. 따라서 이 차이를 의식하게 해줄 수 있는 근거는 믿음-특유의 현상성이어야 한다.

믿음이 현상성을 가진다는 것을 받아들일 만한 또 하나의 이유는 현상적/감각적 상태와 마찬가지로 우리의 믿음이 정도적인 측면이 있다는 것이다.[59] 이는 믿음 뿐만 아니라 예컨대 욕구, 의도에 대해서도 마찬가지인데, 어떤 것이 의식 속에서 정도적인 측면을 가진다는 것은 현상성과 감각질의 가장 큰 특징 중 하나이다. 우리는 즐거움을 약하게도, 강하게도 느끼고 이들을 구분한다. 예컨대 병원에서는 보편적인 자가진단 방식으로 고통의 정도를 스펙트럼을 통해서 스스로 보고하게 한다. 많은 경우 주체는 그렇게 믿을 이유를 의식적으로 떠올릴 수 있는 것과 별개로 자신의 어떤 믿음이 다른 믿음보다 더 확신을 갖는 믿음이라고 판단할 수 있는 것처럼 보인다. 마치 영화 속에서 모든 가시적 증거가 그가 범인임을 지목함에도 직감적으로 그가 진범이 아니라고 강하게 믿게 되는 형사처럼

---

[59] 골드만 또한 우리가 어떤 것을 지향적 태도를 가질 때 그것을 강하게 갖거나, 덜 강하게 가질 수 있다는 사실과 이러한 정도를 표면적으로 집어내는 심적 용어들-확신, 자신감 있음, 의심스러움 등-이 있다는 점에 주목한다.(1993: 24)

많은 경우 주체는 의식적으로 그렇게 믿을 이유를 살피지 않고 설령 그 이유가 다른 방향을 가리키는 경우에도 자신이 무엇에 개입하고 있는지, 그리고 그것이 얼마나 강한 개입인지를 의식할 수 있다. 그리고 만일 믿음에도 특유한 현상성이 있음을 가정한다면, 이 특유한 현상성은 행위자 스스로가 자신이 무언가를 믿고 있고 이 믿음이 강한지 혹은 약한지를 즉각적으로 의식할 수 있는 현상을 단순하게 설명해줄 수 있다.

4.3. 믿음에 대한 현상적 투명성 이론과 질문들

4.3.1. 현상적 투명성 이론

이제 지금까지의 논의를 바탕으로 현상적/감각적 상태에 대한 자기귀속에 대한 정식화를 다음과 같이 제시한다:

[현상성의 대면(주목)에 근거한 현상적/감각적 상태에 대한 자기지식]

S 는 현상성의 대면에 근거해 현상적/감각적 상태 M_1 에 대한 자기지식을 얻는다:

iff (i) S 는 심적 속성 M 의 사례인 심적 상태 M_1 를 가진다.

(ii) M_1 은 심적 속성 M 특유의 현상성 P_M 을 예화한다.

(iii) S 는 M_1 이 예화하는 현상성 P_M 을 대면(주목)하고 여기에 심적 개념 〈M〉을 적용하여 스스로에게 귀속한다.

고통을 예로 들자면 내가 현상성의 대면에 근거해 고통 상태를 스스로에게 귀속한다는 것은 첫째, 나는 심적 속성 고통의 사례인 심적 상태를 가지며, 둘째, 내가 가지고 있는 이 심적 상태는 고통 특유의 현상성, 즉 아픔의 느낌을 예화하고 있고, 셋째, 나는 이 심적 상태에 예화된 아픔의 느낌을 주목하고 여기에 심적 개념 〈고통〉을 적용하여 스스로에게 귀속하는 과정이다. 이때 심적 속성 M 이 반드시 세부적인 다른 속성들로 다시

쪼개어질 수 없는 단순한 속성이라고 가정될 필요는 없다. 고통도 예를 들어 시린 듯한 고통, 얼얼한 고통, 찌르는 듯한 고통처럼 그 속성 아래에서 묶일 수 있는 하위 속성들을 포함한다고 해도 문제될 것이 없다.

현상성의 대면에 근거해 현상적/감각적 상태에 대한 자기귀속을 하기 위해서 주체는 이미 학습을 통해서 P_M 이 개념 〈M〉을 적용하기에 적절한 바로 그러한 느낌임을 알고 있어야 한다. 유아가 고통의 현상성을 느끼면서도 자신이 고통 상태에 있다는 자기지식을 가질 수 없는 까닭은 개념 〈고통〉을 알지 못하고 따라서 자신의 고통스러운 느낌에 〈고통〉을 적용할 수 있는 능력을 갖고 있지 않기 때문이다. 하지만 성장하면서 유아는 개념을 배우고 어떤 현상성에 〈고통〉을 적용할 수 있는지에 대해서 학습하는 과정을 통해서 성공적으로 자기지식을 얻는 능력을 갖추게 된다.

다음은 지향적 태도인 믿음에 대해서 현상성에 근거해 얻어지는 직접적인 자기지식에 대한 정식화이다:

[투명성 절차를 통한 믿음에 대한 자기지식 획득]

S는 **투명성 절차**를 통해 P에 대한 믿음에 대한 자기지식을 얻는다

iff (i) S는 심적 속성 믿음의 사례인 믿음 B_1을 가지고 있다.

(ii) S가 투명성 절차를 따라 명제 P가 **참인지** 고려할 때,

B_1은 **믿음-특유의 현상성** B_M(표현하자면, P가 참이라는 느낌)을 예화한다.

(iii) S는 B_M(P가 참이라는 느낌)을 대면하여 심적 개념 〈믿음〉을 적용하고, 투명성 질문의 대상인 P 를 내용으로 갖는 믿음을 스스로에게 귀속한다.

앞서 주체가 자신이 P를 믿는지를 알기 위해서 사태 P에 대해서 고려하는 과정을 투명성 절차로 제시하였다. "세계 삼차 대전이 발생할 것이라고 믿나요?"라는 질문을 받으면 주체는 자연스럽게 "세계 삼차 대전이 발생할까요?"로 이행하고, P 가 참인지 생각해본 후 맞다면 앞의 질문에 대해 그렇다고 판단하는 것이 투명성 절차를 통해 믿음에 대한 자기지식을 얻는

과정이다. 하지만 투명성 절차를 거치기 위해서 심적 상태에 대하여 명시적 질문, "세계 삼차 대전이 발생할 것이라고 믿나요?"가 반드시 선행되어야 한다고 가정할 필요는 없다. 심적 상태에 대한 질문, 예컨대 '지금 바깥에 비가 오고 있다고 믿나요?'가 선행된다면 이로부터 자연스럽게 투명성 절차에 들어가 '지금 바깥에 비가 오는가?'를 고려하겠지만, 지금 바깥에 비가 오는지를 고려하기 위해서 반드시 믿음에 대한 명시적인 질문이 선행적으로 제기될 필요는 없기 때문이다.

P 를 고려할 때 언제나 자신이 P 를 믿는다는 결론에 도달하는 것도 아니다. 일반적인 상황에서 주체는 세계의 사태에 대한 판단을 내리면서 그로부터 자기지식을 명시적으로 얻지는 않는다. 의학 기사를 보면서 '코로나19를 극복하려면 집단 면역이 필요하겠다'라고 판단한다고 해서 곧장 '나는 코로나 19 를 극복하려면 집단 면역이 필요하다고 믿는다'라는 자기지식을 얻지는 않는다. 이는 'P'라고 판단할 때마다 '나는 P를 믿는다'는 자기지식을 얻을 필요가 없기 때문에, 주체가 매번 스스로가 갖는 참이라는 느낌에 믿음의 개념을 적용하지는 않기 때문이다. 하지만 만일 자기지식을 얻는 것이 요구되는 상황-예를 들어 버지가 말하는 비판적 추론의 상황처럼 나의 믿음들의 정합성을 검토하려는 상황-에서는 어떠한 질문이 명시적으로 선행되지 않더라도 투명성 절차를 진행할 수 있다.

또한 믿음에 대한 투명성 절차에서 "P 는 참인가?"라는 질문을 **고려한다**는 것은 모란이 말하는 **숙고**처럼 규범적 이유를 찾아보는 과정일 필요는 없다. 예를 들어서, 누군가 '중국의 수도가 베이징이라고 믿나요?"라고 나에게 묻는다면[60] 중국의 수도가 베이징인지를 뒷받침하는 좋은 이유를 살펴보지 않더라도 당연히 중국의 수도가 베이징이라는 명제가 참이라는 느낌을

---

[60] 물론 너무도 당연한 질문이라고 생각되기 때문에 이 경우 질문 받은 사람은 질문자가 이 질문을 한 숨겨진 의도-'내 상식을 테스트하나?'-를 고민하거나, 질문자가 질문할 합리적 이유를 떠올려보려고 하거나-'설마 모르는 사이에 상하이가 수도가 되었나?'-할 수도 있을 것이다. 하지만 이 논의의 사례는 정말로 단순히 상식을 묻는 상황으로, 따라서 질문의 대상이 되는 사태가 정말로 성립하는지를 굳이 생각해볼 필요가 없는 상황으로 가정한다.

대면하고, 그렇다고 이야기할 것이다. 이는 어떤 명제가 당연히 참이라고 확신한다면 굳이 그 명제의 참에 개입하기 위해서 숙고할 필요는 없기 때문이다. 물론 해당 명제가 확실히 참이라는 느낌이 들지 않는다면 주체는 이를 뒷받침해주는 이유들을 살펴보는 (모란이 말하는) 숙고의 과정에 들어갈 것이다. 하지만 숙고를 통하여 이유를 살펴볼 때에도 여전히 떠오르는 이유들이 해당 명제가 참이라고 느껴지게 만들어주지 않는 한, 다른 이유들을 더 찾기 위해 숙고를 계속하거나 혹은 결론적으로 이유들이 세계 삼차 대전이 일어날지를 확신할만큼 충분히 강하지 않다고 판단하고 개입하지 않을 것이다. 만일 어떤 이유가 그 명제가 참으로 느끼게 만들어 줄 때에야 주체는 세계 삼차 대전이 일어난다는 것에 비로소 개입하게 된다.

현상적 투명성 이론이 그 현상성을 자기지식의 근거로 삼음에도 내성 이론이 아니라 투명성 이론의 일종이라고 생각될 수 있는 이유는 다음과 같다: 자기지식을 얻기 위해서 주체가 근거로 삼는 현상성, 즉 개념 〈믿음〉을 적용하는 그 느낌은 바로 우리가 참인 명제를 고려할 때 경험하는, 따라서 사태의 판단에 반영되는 **참이라는 느낌**이다. 주체가 어떤 사태가 성립하는지 판단하기 위해서 근거로 삼는 참이라는 느낌이 바로 주체가 자기지식을 얻기 위해 투명성 절차가 시행되는 맥락에 〈믿음〉을 적용하는 믿음-특유의 현상성이며, 바로 그렇기 때문에 이 주체가 얻는 자기지식은 사태에 대한 판단과 동기화되는 방식으로 얻어질 수밖에 없다.

마지막으로 이때 사태의 판단에 반영되는 현상성이라는 것이 정확히 무엇인지 이해하기 위해 다음의 진술들 생각해보자:

(a-1) 그의 행동은 **화날만큼** 무례하다.

(a-2) 나는 그의 무례한 행동에 대해 **화가 난다.**

(b-1) 내일 비가 올 것이 **확실하다.**

(b-2) 나는 내일 비가 올 것이라고 **확신한다.**

(c-1) 세계 삼차 대전이 일어난다는 것은 **참이다.**

(c-2) 나는 세계 삼차 대전이 일어날 것이라고 **믿는다.**

a-1 은 상대방의 행동에 대한 판단이다. 그리고 a-2 는 상대방의 행동에 대한 나의 태도를 보고한 것이다. 자기지식을 얻을 필요가 없는 맥락에서,그의 행동에 대해 화가 난 주체는 그의 행동이 화날만큼 무례하다고 생각한다. 이때 주체는 꼭 자신이 그의 무례한 행동에 대해 생각할 때 느끼는 화의 현상성에 〈화〉를 적용할 필요는 없으며 다만 그렇게 느낀 화는 세계에 대한 나의 판단에 이미 반영되어 있을 뿐이다. 하지만 그의 무례한 행동에 대한 화를 스스로에게 귀속할 때 〈화〉가 적용되는 대상은 자기지식을 얻지 않는 맥락에서 그의 행동 자체에 대한 판단의 근거가 되는 나의 현상성과 같다. 마찬가지로 자기지식을 얻을 필요가 없는 맥락에서, 나는 확신에 찬 태도로 내일 비가 올 것이 확실하다고 생각한다. 하지만 확신에 찬 태도로 내일 비가 올 것이 확실하다고 생각하기 위해서 자신이 느끼는 확신에 〈확신〉을 적용할 필요는 없다. 다만 자기지식을 얻기 위해서는 나는 나의 확신하는 느낌을 대면하고 그에 근거해 심적 개념 〈확신〉(혹은 상위 개념인 〈믿음〉)을 적용한다. 마지막으로, 세계 삼차 대전이 일어난다는 명제가 참이라고 판단하기 위해서 주체가 반드시 자신이 느끼는 참이라는 느낌에 어떤 심적 개념을 적용할 필요는 없다. 하지만 자기지식을 얻을 때, 나는 나의 이 참이라는 느낌에 근거하여 〈믿음〉을 적용하고 자기지식을 얻는다.

4.4. 특유한 접근과 일인칭 특권

마지막 절에서는 현상적 투명성 이론이 어떻게 특유한 접근과 일인칭 특권을 보이는지 설명할 것이다.

우선 투명성 절차와 심적 속성 특유의 현상성에 주목하는 입장이 지향적 태도에 대한 일인칭 주체의 **특유한 접근**을 설명한다는 것은 분명하다. 자신의 지향적 태도를 알기 위해서 투명성 절차를 따를 때, 사태에 대한

고려와 동반하는 현상성을 대면하는 것은 오직 일인칭 주체에게만 허용되기 때문이다.

투명성 절차를 통하여 세계의 사태에 대한 판단에 동반하는 현상성에 심적 개념을 적용하여 스스로에게 심적 태도를 귀속시킴으로써 얻어지는 자기지식이 어떻게 인식적인 보증을 가질 수 있는가? 나는 이 보증이 모란의 초월적 가정과 유사한, 일종의 주체의 합리성에 대한 가정으로부터 나온다고 제안한다. 즉, 내가 만일 최소한으로 합리적인 사람이라면, 나는 나의 심적 상태가 예화하는 현상성을 대면하고 주목하며 이에 적절한 심적 개념을 적용하여 자신이 어떤 심적 상태를 갖고 있는지 직접적으로 알 수 있다. 따라서 최소한으로 합리적인 주체라면 자신의 현상성이 실제로 자신이 갖고 있는 심적 상태를 적절하게 반영하는 현상성이며, 현상성에 근거한 판단이 스스로에게 심적 상태를 귀속하는 적절한 근거라고 가정할 수 있는 인식적 권리를 가진다.

여기서 최소한으로의 합리성은 모란의 이유-반응성으로서의 합리성보다는 훨씬 넓은 외연을 가진다. 모란에게 자기지식을 정당하게 얻기 위해 필요한 합리성은 믿을 만한 좋은 이유가 있는 것만을 실제로 믿고, 바랄 만한 좋은 이유가 있는 것만을 실제로 바라는, 즉, 규범적 이유에 대한 반응성으로 이해된다. 하지만 현상성과 관련된 합리성은 주체가 자신이 갖고 있는 심적 상태들을 적절히 반영하는 방식으로 스스로에게 주어지는 현상성을 대면하고 이에 주목할 수 있는 능력을 가리킨다. 이것이 '최소한으로의 합리성'인 까닭은, 설령 규범적 이유에 따라서 자신의 심적 상태를 형성하는 데에는 실패하는 주체이더라도 자신의 (규범적 이유에 대한 합리성의 측면에서) 비합리적인 믿음을 현상적으로 의식하는 것은 충분히 가능하기 때문이다.

한편 현상성에 대한 합리성 가정에 대하여 다음과 같은 이유에서 그 타당성에 의문을 제기할 수 있어 보인다. 만일 심적 상태에 적절히 조응하는 방식으로 주어지는 현상성을 대면하는 능력으로서의 합리성이 있다면, 이는 현상적인 느낌에 적용되는 일종의 규범성이 있다는 주장을 함축하는 것처럼 보인다. 합리적이지 않은 행위자가 경험하는 현상성은 많은 경우 자신의

심적 상태에 대하여 부적절한 것이 되기 때문이다. 하지만 주관적인 느낌이 어떤 의미에서 규범적일 수 있는가? 이에 대한 나의 답은 현상성에 일종의 규범성을 적용하는 것은 일상적인 직관에서도 전혀 이상할 것이 없다는 것이다. 다음과 같은 사례에서 현상성에 대해서도 그것이 적절한/부적절한 현상성인지를 판단할 수 있다는 점에서 일종의 규범이 있음을 생각해볼수 있다. 예컨대 내가 벌레를 보면 벌레를 무서워 하는 사람처럼 말하고 행동하는 방식으로 행동한다고 하자. 하지만 그러한 내가 현상적으로는 벌레를 볼 때 불쾌감을 대면하지 않고 일종의 쾌감을 대면하고 있다면, 우리는 이 쾌감이 적절하지 않다고 말한다. 이때 내가 대면해야 할 적절한 현상성은 벌레를 무서워함 혹은 혐오함이기 때문이다. 다만 여기서 말하는 합리성의 규범성은 모란이 말하는, 세계의 객관적인 사태들에 대해 심적 상태가 올바르게 반응해야 한다는 이유-반응성으로서의 합리성의 규범성이 아니라, 인식 주체의 심리 상태가 기능성과 현상성 간의 내적 일관성을 갖고 있어야 한다는 의미에서 내적 일관성의 규범성이다.

일견 이처럼 심적 상태와 적절하게 조응하여 현상성이 주어질 것만을 요구하는 내적 일관성은 대부분의 사람들이 만족하는, 일반적으로 생각하는 이유-반응성으로서의 합리성보다 훨씬 더 조건을 만족하기 쉬운 합리성이며 따라서 어쩌면 합리성으로 일컫기 어려운 공허한 개념인 것처럼 보인다. 하지만 무엇을 믿고 또 욕구해야 하는지를 결정하기 위해 규범적 이유를 별로 숙고하지 않는 행위자라도 자신이 무엇을 믿고 욕구하는지를 아는 능력을 갖고 있으며, 그렇기 때문에 기존의 모란의 숙고와 결의가 자기지식을 정당하게 얻기 위한 필요한 조건으로서 너무 강하다는 것은 2 장에서 모란에게 제기된 가장 핵심적인 문제였다. 이에 비추어 볼 때, 주체의 시점에서 자기지식을 얻기 위해 필요한 내적 일관성의 합리성의 조건을 만족하기 쉽다는 것은 대부분의 사람들이 일상적으로 스스로의 심적 상태에 대한 믿음을 성공적으로 귀속한다는 경험적 사실을 비추어 볼 때 자기지식 현상을 잘 설명해준다는 점에서 오히려 장점이 된다.

심적 체계 내에서 기능성과 현상성의 내적 일관성이 왜 일반적으로 성립하며, 이러한 내적 일관성이 깨어지는 이상 상황은 어떤 경우 발생하는지를 설명하는 것은 이 논문의 범위를 벗어난다. 다만 나의 잠정적인 추측은 기능성과 현상성의 내적 일관성은 이유-반응성으로서의 합리성이 성립하기 위해서 선제적으로 갖춰져야 하는 조건이며, 따라서 객관적 이유를 따라서 사태의 변화에 합리적으로 대응하기 위해서는 이러한 내적 일관성이 요구된다는 것이다. 객관적인 규범적 이유를 따라서 자신의 심적 상태를 조정하는 능력이 합리적인 주체라면 반드시 가지고 있어야 할 능력이라는 사실에는 이미 모란과 슈메이커, 그리고 버지를 포함하여 많은 철학자들이 주목해왔다. 하지만 그들은 실제로 어떤 능력 덕분에 자신의 심적 상태를 알고 합리성의 규범을 따를 수 있는지에 대해서는 구체적인 설명을 제공하지는 않는다. 나는 현상성을 통해 자신의 기능적 상태에 대하여 즉각적이고 직접적으로 아는 것이 이유-반응성으로서의 일반적인 능력을 가능하게 해주는 가능 조건이라고 생각한다. 기능성과 현상성의 내적 일관성의 능력 덕분에 어떤 심적 상태에 현재 놓여있는지 즉각적으로 알 수 있으며, 그 덕분에 합리성의 규범을 통해 자신이 지금 갖고 있는 다양한 심적 상태를 조정할 수 있는 것이다.

제5장 결 론

지금까지의 논의를 통해 나는 자기지식에 있어서 인식 주체에게 허용되는 것처럼 보이는 일인칭 특권에 대한 한 가지 설명을 제시하였다. 이 글의 주된 목표는 심적 상태 중에서도 지향적 태도, 그리고 그 중에서도 믿음에 대한 일인칭 특권을 설명하는 것이었다. 그리고 그 중에서도 믿음의 투명성을 사용하여 자기지식을 얻을 수 있다는 투명성 이론가들의 이론에 주목하였다. 투명성 이론이 맞다면 주체는 자신의 마음 속에 있는 표상을 살펴볼 필요가 없이 단지 세계의 사태를 직접 고려하는 능력만으로도 스스로의 믿음에 대해서 정당하게 알게 된다. 그렇다면 지금까지 내성 이론가들에게 제기된 어려움-지향적 태도의 내용에 대한 내성이 어떻게 가능한가-에 대한 추가적인 설명 없이 지향적 태도에 대한 직접적인 자기지식을 설명할 수 있다는 점에서 이론적으로도 장점을 가진다.

하지만 나는 보다 근본적인 차원에서 투명성 이론이 자기지식에 대한 올바른 그림을 제공하고 있다고 생각한다. 주체의 심적 상태는 주체와 구분되는 대상이 아니며, 다만 주체는 스스로의 심적 상태 **안에서** 세상을 바라볼 수 밖에 없기 때문에 자신의 심적 상태를 알기 위해서는 자신이 세계를 어떠한 방식으로 바라보고 있는지를 살펴보아야 한다. 이는 세계의 객관적 사태에 대한 판단과 그러한 사태를 내용으로 갖는 지향적 태도에 대한 판단이 동기화되지 않는 상황이 마치 모순적으로 보일 만큼 비합리적인 상황이라는 무어의 역설이 함축하는 바로부터 알 수 있다.

기존의 투명성 이론가들은 사태에 대한 판단과 자기지식의 동기화가 일인칭 시점에서의 특권적 자기지식의 핵심이라는 것에는 정확히 주목했지만, 구체적으로 어떤 메커니즘을 통하여 주체가 세계의 사태에 대한 판단에 도달하는지를 설명하지는 못하였다. 현상적 투명성 이론은 이처럼 자기지식의 맥락에서 기존의 투명성 이론가들이 별 관심을 갖지 않았던 판단의 메커니즘에 대한 실질적인 설명을 제공하는 것을 목표로 하고 있다.

나는 기존의 투명성 이론이 판단의 메커니즘에 대한 대답을 결여하고 있다는 본고의 문제 의식과 그에 대한 현상적 투명성 이론의 대답이 자기지식에 대한 투명성 이론이 나아갈 지평에 조금이나마 기여할 수 있기를 기대한다.

참고 문헌

Armstrong, D. (1993), A Materialist Theory of the Mind, London; New York : Routledge.

Boyle, M. (2009), "Two Kinds of Self-Knowledge", Philosophy and phenomenological research, 2009-01, Vol.78 (1), 133-164.

_____ (2011), "Transparent Self-Knowledge", Supplementary volume/Aristotelian Society, 2011-06-01, Vol.85 (1), 223-241.

_____ (2019), "Transparency and Reflection", Canadian journal of philosophy, 2019-10-03, Vol.49 (7), 1012-1039.

Boghossian, P. (1989), "Content and Self-Knowledge", Philosophical topics, 1989-04-01, Vol.17 (1), 5-26.

BonJour, L. (1980), "Externalist Theories of Empirical Knowledge", *Midwest studies in philosophy,* 1980-09, Vol.5 (1), 53-74.

Burge, T. (1988), "Individualism and Self-Knowledge", The Journal of philosophy, 1988-11-01, Vol.85 (11), 649-663.

_____ (1993), "Content Preservation", The Philosophical review, 1993-10-01, Vol.102 (4), 457-488.

_____ (1996), "Our Entitlement to Self-Knowledge", Proceedings of the Aristotelian Society, 1996-01-01, Vol.96, 91-116.

Byrne, A. (2011), "Transparency, Belief, Intention", Supplementary volume/Aristotelian Society, 2011-06-01, Vol.85 (1), 201-221.

_____ (2018), *Transparency and Self-Knowledge*, Oxford, UK: Oxford University Press, 2018.

Carrutehrs, P. (2011), *The Opacity of Mind: An Integrative Theory of Self-Knowledge,* Oxford, New York: Oxford University Press, 2011.

Carruthers, P. & Gennaro, R. (2001), "Higher-order theories of consciousness", *Stanford Encyclopedia of Philosophy (First published Tue Apr 3, 2001; substantive revision Wed Sep 2, 2020) (https://plato.stanford.edu/entries/consciousness-higher/)*

Cassam, Q. (2014), *Self-Knowledge for Humans*, Oxford, UK; New York, NY: Oxford University Press, 2014.

Chalmers, D. (1996), *The Conscious Mind: In a Search of a Fundamental Theory,* New York: Oxford University Press, 1996.

_____ (2003), "The Content and Epistemology of Phenomenal Belief", *Consciousness: New Philosophical Perspective*, Oxford: Oxford University Press, 2003, 220-272.

_____ (2004), "The representational character of experience", *The Future for Philosophy*, Oxford University Press, 153-181.

Dretske, F. (1994), "Introspection", Proceedings of the Aristotelian Society, 1993-01-01, Vol.94, 263-278.

Evans, G. (1982), *The Varieties of Reference*, Oxford: Oxford Univesrity Press.

Fricker, E. (2016), "Unreliable Testimony", In Brian McLaughlin & Hilary Kornblith (eds.), *Goldman and his Critics*. Malden, MA: Wiley-Blackwell. 88-120.

Fumerton, R. (2005). "Speckled hens and objects of acquaintance", *Philosophical Perspectives*, *19*, 121-138.

Gendler, T. (2008), "Alief and Belief", *The Journal of philosophy*, *105*(10), 634-663.

Gertler, B. (2001), "Introspecting Phenomenal States", *Philosophy and Phenomenological Research*, *63*(2), 305-328.

_____ (2003), "Self-Knowledge", *Stanford Encyclopedia of Philosophy* (*First published Fri Feb 7, 2003*, *substantive revision Mon May 18, 2015*) (https://plato.stanford.edu/entries/self-knowledge/)

_____ (2011), "Self-Knowledge and the Transparency of Belief", In Anthony Hatzimoysis (ed.), *Self-Knowledge*. Oxford: Oxford University Press.

Goldman, A. (1993a), "The Psychology of Folk Psychology", *Behavioral and Brain sciences*, *16*(1), 15-28.

_____ (1993b), "Consciousness, Folk Psychology, and Cognitive Science", *Consciousness and Cognition*, *2*(4), 364-382.

_____ (2006), "Self-Attribution", *Simulating Minds,* Oxford; New York: Oxford University Press, 2006. Chapter. 9

Heal, J. (2004), "Moran's "Authority and Estrangement", *Philosophy and Phenomenological Research*, Vol. 69, No. 2, 427-432.

Horgan, T. & Tienson, T. (2002), "The Intentionality of Phenomenology and the Phenomenology of Intentionality", *Philosophy of mind: Classical and contemporary readings*, 520-533.

Kriegel, U. (2007), "Phenomenal Epistemology: What is Consciousness That We May Know It So Well?", *Philosophical Issues*, *17*, 123-144.

_____ (2013), "The Phenomenal Intentionality Research Program", Oxford: Oxford University Press.

Kripke, S. (1980), *Naming and Neccessity*, Malden, MA; Oxford, UK: Blackwell.

Mckinsey, M. (1991), "Anti-Individualism and Privileged Access," *Analysis*, 51: 9-16.

_____ (1994), "Acepting the Consequences of Anti-Individualism," *Analysis*, 54(2): 124-128.

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

_____ (2003), "Reponses to O'brien and Shoemaker", *European Journal of Philosophy 11:3*, 402–419.

_____ (2012), "Self-Knowledge, 'Transparency', and the forms of activity", *Introspection and consciousness*, 211-236.

Nichols & Stich (2003), *Mindreading*, Oxford: Oxford University Press.

Nisbett, E. & Wilson, T. (1977), "Telling more than we can know: Verbal reports on mental processes", *Psychological Review*, 84: 231–259.

O'Brien, L. (2003), "Moran on Agency and Self-Knowledge", *European Journal of Philosophy,* 11:3, 375-390.

Pitt, D. (2004), "The Phenomenology of Cognition or What is It Like to Think that P", *Philosophy and Phenomenological Research*, *69*(1), 1-36.

_____(2011), "Introspection, Phenomenality and the Availability of Intentional Content", *Cognitive phenomenology*, 141-173.

_____(2018), "Consciousness and Intentionality", In *The Routledge Handbook Of Consciousness*, NY: Routledge, 260-270.

_____(2019), "Acquaintance and Phenomenal Concepts", *Cambridge classic arguments series: The knowledge argument*, 87-101.

Peacocke, C. (1998), "Conscious Attitudes, Attention, and Self-Knowledge", *Knowing Our Own Minds* (edited by Wright, C. et al.), Oxford: Oxford University Press.

Putnam, H. (1975), "The Meaning of 'meaning'", *Philosophical papers*, *2*.

Schwitzgebel, E. (2003), "A Phenomenal, Dispositional Account of Belief", *Noûs*, *36*(2), 249-275.

_____ (2008), "The Unreliability of Naive Introspection", *Philosophical Review*, *117*(2), 245-273.

_____ (2019), "Belief", *Stanford Encyclopedia of Phiolsophy (First published Mon Aug 14, 2006; substantive revision Mon Jun 3, 2019) (https://plato.stanford.edu/archives/fall2019/entries/belief/)*

Setiya, K. (2012), "Transparency and Inference", Proceedings of the Aristotelian Society, 2012-01-01, Vol.112, p.263-268

Shoemaker, S. (1994), "Self-Knowledge and Inner Sense", *Philosophy and phenomenological research*, *54*(2), 249-269.

_____ (2003), "Moran on Self-Knowledge", *European Journal of Philosophy*, *11*(3), 391-401.

Abstract

Phenomenal Transparency Account on Self-Knowledge

Shin, Hui-Jae
Philosophy (Western Philosophy)
The Graduate School
Seoul National University

This paper aims to propose 'Phenomenal Transparency Account', by examining and diverging from previous theories that could be classified as transparency accounts. Transparency, in this context, indicates the status of mind through which the subject recognizes the object and the world. The subject encounters the world only through his/her 'transparent' mental states, as we see things through sheets of transparent glass when we are wearing eyeglasses. I focus on 'belief' as the examplar of the object of self-knowledge that is acquired by first-person transparency procedure. The ability to recognize one's own belief has been considered to be closely tied to the subject's rationality, and therefore it is important to explicate the nature of self-knowledge of belief when it comes to the topic of rational agency.

I will first argue that intropection accounts, which explains that self-knowledge of intentional states can be fully achieved by one's introspection, fail to explain the subject's self-knowledge that is acquired in one's first-person perspective. The discovery models of

self-knowledge do not satisfy the requirement of Moore's paradox, the precondition that the self-knowledge in first-person perspective must be acquired by ways that are synchronized with the world-directed judgement.

On the other hand, transparency accounts satisfy the requirement suggested by Moore's paradox. Recent transparency accounts examined in this paper--by Byrne, Moran, and Boyle--argue that the subject self-ascribes his/her intentional states only through world-directed considerations. These accounts, however, fall short of explaining the mechanism by which the subject progresses toward the judgment for the states of affiars, therefore cannot fully account self-knowledge in first-person perspective.

In the last chapter, I propose 'Phenomenal Transparency Account' that explicate the judgment mechanism by which the subject arrives at his/her world-directed judgment. In this account, the acquaintance with attitude-type-specific phenomenality, through which the subject arrives at world-directed judgment, becomes the basis of self-attribution. This acquaintance is the peculiar access in the sense that it is permitted only to the subject him/herself, and the subject's self-knowledge by the peculiar access is epistemically direct as it is not acquired by theoretical evidence regarding the belief states. Also, the subject has the epistemic entitlement to ascribe his/her intentional states by the transparency procedure on the assumption that the functionality and the phenomenality of his/her mental states are internally consistent. By this assumption, the self-ascribed belief, which is acquired through the transparency procedure and the acquaintance with phenomenality that comes along with world-

directed considerations, is successfully warranted as proper self-knowledge.

Keywords : Self-Knowledge, Transparency Accounts, Privileged Access, Intentional States, Attitude-type-specific Phenomenality, Moore's Paradox, Philosophy of Mind

Student Number : 2019-25382