철학석사 학위논문

기능주의적 내적 감각 이론에 관한 소고

2023년 2월

서울대학교 대학원

철학과 서양철학전공

김 현 채

기능주의적 내적 감각 이론에 관한 소고

지도교수 한 성 일

이 논문을 문학석사 학위논문으로 제출함
2022년 10월

서울대학교 대학원
철학과 서양철학전공
김 현 채

김현채의 석사 학위논문을 인준함
2022년 11월

위 원 장 김 기 현 (인)

부위원장 한 성 일 (인)

위 원 이 우 람 (인)

국문초록

주요어 : 자기 지식, 내성, 내적 감각, 심리적 기능주의
학 번 : 2020-27285

본 논문은 자기 지식에 관한 이론 중 하나인 내적 감각 이론이 심적 상태에 관한 기능주의를 받아들일 때 다양한 심적 상태들의 내성에 관하여 가장 적절한 설명을 제시할 수 있을뿐더러 환원적 물리주의를 전제하지 않음으로써 심신 문제와 관련해 보다 다양한 형이상학적 입장들을 수용할 수 있게 된다고 주장하고자 한다. 우선 내적 감각 이론은 우리가 스스로의 내부를 들여다보는 내성이라는 과정을 통해 자신의 심적 상태에 대한 지식을 획득하며, 이때 내성은 외부 대상에 대한 지각과 유사한 방식으로 이루어진다고 말한다. 현대의 분석철학적 맥락에서 내적 감각 이론을 발전시켰던 학자들로는 암스트롱, 라이칸, 니콜스와 스티치, 그리고 골드만이 있는데, 본 논문에서는 이들이 제시한 내적 감각 이론의 내용을 이들이 각자 심적 상태에 대한 기능주의에 관하여 어떠한 입장을 갖고 있는지와 연관지어서 살펴보려고 한다. 내적 감각 이론은 심적 상태에 대한 내성을 설명하고자 하는 이론인 만큼, 기능주의를 받아들인다는 것은 내성의 대상이 되는 심적 상태가 기능적으로 정의된다고 보는 것을 의미한다. 그렇다면 기능주의적 내적 감각 이론에서는 내적 감각이 심적 상태를 감지하는 것이 심적 상태의 기능적 속성을 통해서 이루어진다고 말할 수 있다.

기존의 내적 감각 이론가들 가운데서 암스트롱, 라이칸, 니콜스와 스티치는 모두 기능주의를 받아들이고 있다. 구체적으로 암스트롱은 분석적 실현자 기능주의를, 라이칸은 심리적 역할 기능주의를 받아들이며, 니콜스와 스티치는 심리적 기능주의를 받아들이지만 역할 기능주의와 실현자 기능주의 사이에서 특정한 입장을 취하지는 않고 있다. 한편, 가장 최근에 내적 감각 이론을 이어받아 발전시켰던 골드만은 기능주의를 거부하며, 내적 감각의 작용이 심적 상태의 기능적 속성을 대상으로 한다는 기존 견해를 강하게 비판한다. 그는 대안으로서 내적 감각이 심적 상태의 신경적 속성을 대상으로 삼는다는

신경적 내적 감각 이론을 제시한다. 본 논문에서는 골드만의 견해가 분명 설명적인 장점도 있으나, 내성에 대한 설명을 순수하게 신경적인 속성을 통해서만 하고자 한다는 점에서 심적 상태들을 분석하고자 했던 기존 철학적 논의와 단절되며, 또한 환원적 물리주의를 전제해야만 한다는 점에서 이전 이론들에 비해 형이상학적인 개입을 강하게 해야 한다는 부담을 진다고 본다. 그리고 다시금 기능주의적인 내적 감각 이론을 옹호함으로써 내적 감각 이론이 자연과학에 친화적인 내성에 관한 통일적 이론이면서도 환원적 물리주의에 대해 열려있을 수 있다고 주장한다.

기능주의적 내적 감각 이론을 옹호하기 위해서는 기능적 속성에 관하여 골드만을 비롯한 학자들이 제기한 비판들에 대해 답변을 내놓아야 한다. 내적 감각이 심적 상태의 기능적 속성을 감지한다는 견해에 대해서는 크게 네 가지의 비판이 제기되었다. 이는 각각 기능적 속성이 관계적 속성이라는 점, 기능적 속성이 경향적 속성이라는 점, 기능적 속성이 연언적 속성이라는 점, 그리고 기능적 속성의 인과적 효력과 관련하여 제기되었다. 본 논문에서는 이러한 비판들을 살펴보고, 기능주의적인 입장에서 이에 대한 답변을 제시할 것이다. 이때 제시할 대부분의 답변들은 분석적 기능주의와 심리적 기능주의 양측에서 활용할 수 있는 것들이나, 심리적 기능주의가 조금 더 유리한 측면이 존재한다. 따라서 본 논문은 심리적 기능주의를 받아들이는 내적 감각 이론이 가장 최선의 내적 감각 이론이라고 주장할 것이다.

목 차

기능주의적 내적 감각 이론에 관한 소고

1. 서론

내적 감각 이론이란 내성에 관한 이론으로, 우리가 스스로의 내부적인 심적 상태에 관한 지식을 얻는 방식이 외적 감각을 통해 외부 세계에 대한 지식을 얻는 방식과 유사하며, 심적 상태들을 감지하는 일종의 내적 감각이 존재한다는 입장이다. 내적 감각 이론의 철학사적 기원은 흔히 로크에게서 찾는데, 현대의 분석철학적 전통에서는 암스트롱, 라이칸, 니콜스와 스티치, 그리고 골드만이 각자 내적 감각 이론을 이어받아 발전시켰다. 암스트롱과 라이칸까지만 하더라도 내적 감각 이론은 주로 내성이 외부 지각과 비슷한 방식으로 작동한다는 주장 자체에 더 치중되어 있었고, 내성을 수행하는 내적 감각 메커니즘이 구체적으로 어떤 구조를 갖고 어떠한 방식으로 다양한 심적 상태들을 감지해 그에 관한 메타 표상을 산출하는지에 대해서는 니콜스와 스티치, 그리고 골드만에 이르러서야 진전이 이루어졌다. 현대적인 내적 감각 이론의 역사는 그리 길지 않고, 이를 본격적으로 다룬 학자의 수도 그렇게 많지는 않으나, 그럼에도 내적 감각 이론은 자기 지식 이론의 한 갈래이자 내성 이론의 한 종류로서 분석 철학적 맥락 내에서 자리매김하고 있고, 내적 감각 이론을 비판의 대상으로 삼거나 여러 내적 감각 이론들을 정리 및 비교해 소개하는 이차 문헌들의 수는 그리 적지는 않다.

내적 감각 이론의 발전 과정을 살펴볼 때 가장 흥미로운 지점 중 하나는 이 이론의 지지자들이 기능주의에 대해 가진 입장이 저마다 다를뿐더러, 심지어는 극명하게 서로 대립한다는 것이다. 암스트롱, 라이칸, 니콜스와 스티치는 모두 어떤 방식으로든지 기능주의를 받아들이고 있는 반면에, 골드만은 기능주의를 맹렬하게 비판한다. 사실 조금만 생각해보면 내적 감각 이론과 관련된 논의에서 기능주의가 화두로 떠오른다는 사실은 그리 놀라울 만한 것이 아니다. 내적 감각 이론은 심적 상태에 대한 내성을 설명하고자 하는 이론이고, 기능주의는 심적 상태의 정의에 관한 이론이다. 심적 상태가 어떻게 정의되는지에 따라 심적 상태를 내성하는 방식도 영향을 받을 것임은 어느 정도

당연하다고도 할 수 있겠다. 하지만 의외로 내적 감각 이론의 지지자들이 각자 기능주의에 대하여 어떤 입장을 가지고 있었는지는 내적 감각 이론에 관한 이차 문헌들에서 크게 다루어진 바가 없다. 그렇기에 본 논문에서는 기존의 내적 감각 이론들을 살펴보며, 각 학자가 기능주의에 대해 취한 입장이 그들의 내적 감각 이론과 어떻게 연관되고 이로 인해 어떠한 설명적 차이나 이론적 장단점이 발생하는지에 주목해보고자 한다. 특히 가장 최근에 내적 감각 이론을 발전시켰던 골드만이 기능주의적인 내적 감각 이론을 비판했다는 점에 착안하여, 내적 감각 이론이 과연 기능주의를 거부하고 내성이 심적 상태의 기능적 속성이 아닌 다른 속성에 의해 성립한다고 주장해야 하는지를 검토해 볼 예정이다.

정리하자면, 본 논문의 목표는 첫째로는 내성에 관한 이론인 내적 감각 이론을 소개하고 기존에 전개되었던 내적 감각 이론들의 주요 이론적 쟁점들을 검토하면서 동시에 각 학자가 가진 기능주의에 대한 관점이 이들의 내적 감각 이론에 어떤 영향을 미치는지를 확인하는 것이다. 이를 위해서 간략하게나마 기능주의의 여러 갈래가 어떻게 구분되는지도 짚고 넘어갈 필요가 있겠다. 그다음 목표는 각각의 내적 감각 이론이 갖는 설명적 및 이론적 장점과 한계점을 알아보고, 기능주의에 대하여 어떤 입장을 취했을 때 가장 내적 감각 이론이 바람직한 형태가 되는지를 검토하는 것이다. 이후 마지막으로는 기능주의 및 기능적 속성을 통한 내성에 대한 비판을 살펴보고, 이에 대하여 가능한 대응을 제시함으로써 기능주의적인 내적 감각 이론을 다시금 옹호해볼 것이다. 그리고 결론적으로 심적 상태에 관한 심리적 역할 기능주의를 받아들이고 내적 감각의 입력값을 심적 상태의 기능적 속성으로 이해할 때 내적 감각 이론이 모든 종류의 심적 상태에 대한 내성을 설명할 수 있을 뿐만 아니라, 자연과학에 친화적이면서도 환원적 물리주의에 대해 열려있어서 형이상학적으로도 가장 개방적인 형태가 된다고 주장할 것이다.

2. 내적 감각 이론

2. 1. 자기 지식 및 내성의 문제

자기 지식(Self-knowledge)이란 무엇인가? 간단히 말해 자기 지식은 인간이 자기 자신의 심적 상태에 관하여 갖는 지식이다. 일상적인 의미에서 자기 지식이란 인간이 자기 자신에 대하여 갖는 모든 지식을 의미하겠지만, 심리철학에서 자기 지식은 일상적 의미에서보다는 좁은 범위의 지식을 일컫는다. 즉, 자기 자신에 대한 지식이라도 자신의 내적인 심적 상태에 관한 지식이 아닌, 자기 신체에 대한 지식을 비롯한 각종 지식은 철학에서 논의하는 자기 지식의 범위에는 속하지 않는다. 예컨대 내가 나의 머리카락이 검은색임을 아는 것이나, 내가 대학원생임을 아는 것은 철학에서 말하는 자기 지식에는 포함되지 않는 것이다. 그렇다면 지금 논의 대상이 되는 종류의 자기 지식이란 어떤 것들이 있는지 몇 가지 예시를 살펴보자. '나는 대한민국의 수도가 서울이라고 생각한다', '나는 내 동생에게 화가 난 상태다', '나는 지금 왼쪽 팔꿈치가 아프다', '나는 지금 차가운 콜라가 매우 마시고 싶다', '나는 파인애플이 올라간 피자를 싫어한다', '나는 어제 길에서 들었던 노래를 지금 머릿속에서 흥얼거리고 있다' 등과 같은 믿음들을 내가 가지고 있고, 이 믿음들이 정당화된 참인 믿음들이라면 나는 자기 지식을 가지고 있다고 말할 수 있을 것이다.

그런데 이러한 자기 지식은 '내 머리카락은 검은색이다.' 또는 '나는 서울대학교 철학과 석사과정에 재학 중이다.'라는 자기 지식과 어떤 점에서 다르기에 철학적 논의에서 전자와 후자가 구분될 필요가 있는 것일까? 우선 내 머리카락의 색깔에 대한 지식을 살펴보면, 나는 내 머리카락을 두 눈으로 직접 봄으로써 내 머리카락이 검은색임을 알 수 있다. 이러한 지식 획득 과정은 내가 직접 봄으로써 내 동생의 머리카락도 검은색임을 알게 되거나, 내가 신은 신발이 검은색이라는 것을 알게 되는 것과 다르지 않다. 또한 다른 사람이 내 머리카락을 보고 내 머리카락이 검은색임을 알게 되는 것과 내가 내 머리카락이 검은색임을 알게 되는 것과도 큰 차이가 없다. 이러한 종류의 자기 지식에서는 단지 지식의 대상이 나 자신, 혹은 나에게 속하는 것일 뿐, 지식 자체가 다른 지식 일반과 특별히 구분되는 것은 아니라 할 수 있다. 내가 대학원생임을 아는 것 또한 마찬가지다. 나는 재학증명서나 등록금 납부 기록 등 여러 외부적 사실들을 통해 내가 대학원생이라는 사실을 알 수 있고, 다른 사람들 또한 같은 방식으로 여러 외부적 사실들을 통해 내가 대학원생이라는 것을 알 수 있다.

그렇다면 이제 앞서 철학적 의미에서 자기 지식에 속하는 것으로 소개되었던 사례들을 살펴보자. 다른 사람들은 내가 '대한민국의 수도는 서울이다'라고 생각한다는 것을 어떻게 알 수 있는가? 여기에는 다양한 근거들이 활용될 수 있을 것이다. 예컨대 내가 직접 "대한민국의 수도는 서울이다"라고 말하는 것을 들었을 수도 있고, 대한민국의 수도가 어디인지 묻는 설문지에 내가 서울이라고 답한 것을 보았을 수도 있다. 혹은 내가 상식적인 대한민국 국적의 성인이라면 대한민국의 수도가 서울이라는 것쯤은 당연하게 알고 있을 것이라 추론했을 수도 있을 것이다. 그러나 이는 내가 나 스스로에 관해 '나는 대한민국의 수도가 서울이라고 생각한다'는 자기 지식을 획득하는 방식과는 사뭇 다른 것처럼 보인다. 나는 굳이 입 밖으로 내가 말하는 것을 듣거나 다른 어떤 외적인 행동을 보이는 것을 관찰하지 않고서도 내가 그렇게 생각한다는 것을 알 수 있는 것 같기 때문이다. 또한, 일견 어떠한 추론적 과정을 거치지 않고서도 이러한 자기 지식을 획득할 수 있어 보인다. 다른 예시를 보자. 타인은 내가 지금 차가운 콜라가 매우 마시고 싶다는 것을 어떻게 알 수 있는가? 내가 차가운 콜라가 매우 마시고 싶다고 말하는 것을 들었을 수도 있고, 편의점에 진열된 차가운 콜라를 보며 눈을 떼지 못하거나 침을 꼴깍 삼키는 것을 보았을 수도 있다. 그리고 이러한 외적으로 드러난 행동적 증거들로부터 내가 차가운 콜라가 매우 마시고 싶음을 추론할 수 있다. 그러나 나는 비록 지금 차가운 콜라가 매우 마시고 싶음에도 이러한 외적인 행동들을 통해 나의 욕구를 드러내는 것을 절제할 수 있고, 그럴 경우 타인이 나의 욕구에 대해 알 수 있는 수단은 상당히 제한된다. 반면, 나는 아무런 외적인 행동들을 나타내지 않으면서도 스스로 차가운 콜라가 매우 마시고 싶음을 당연하게 알 수 있다. 심지어 겉보기에는 콜라가 전혀 마시고 싶지 않은 것처럼 행동하면서 타인이 나의 욕구에 대해 그릇된 믿음을 형성하게 유도하는 와중에도 나는 사실 내가 콜라가 마시고 싶다는 것을 알 수 있다.

위의 논의를 보면 자신의 믿음, 욕구, 감정, 고통 등 다양한 심적 상태에 대해 갖는 자기 지식은 모종의 특수성을 지니는 것처럼 보인다. 바로 이 특수성이 정확히 어떤 것인가에 관해서는 이미 다양한 철학적 논의가 진행되었고, 그 중에서 가장 활발히 논의된 주제 두 가지를 간략하게 소개하자면 다음과 같이 정리할 수 있다. 첫째는 자기 지식이 특수한 인식적 확실성을 갖는

지에 대한 논의다. 자기 지식이 특수한 인식적 확실성을 갖는다는 입장의 극단에는 인간이 자신의 심적 상태에 대해 갖는 자기 지식은 아예 틀릴 수 없다는 오류 불가능성 논제가 있다. 외부 사실에 대한 나의 믿음은 때때로 잘못될 수 있으나, 과연 내가 나의 내적인 심리 상태나 나 자신의 생각에 대해서 잘못된 믿음을 가질 수 있는가? 오류 불가능성 논제를 받아들이는 이들은 이 질문에 대해 그럴 수 없다는 답변을 내놓는다. 흔히 이러한 입장의 대표적인 지지자로 데카르트가 꼽히기는 하지만, 사실 이는 적절하지 않다. 데카르트의 코기토는 내가 생각하고 있다는 사실 자체만은 확실하다고 보지만, 그 생각의 내용 전체가 항상 오류 불가능하다고 말하는 것은 아니다. 예컨대 내가 대한민국의 수도가 서울이라고 생각한다는 자기 믿음을 형성할 때, 내가 생각하고 있다는 것만은 확실하지만, 내가 알지 못하는 모종의 이유로 인해 실제로는 내가 대한민국의 수도가 부산이라고 생각하고 있었고, 결과적으로 내가 대한민국의 수도가 서울이라고 생각한다는 믿음은 잘못되었을 수도 있는 것이다. 이러한 점에서 오류 불가능성 논제는 데카르트의 코기토보다 인간의 자기 지식에 관해 훨씬 더 강한 주장을 하고 있다. 그러나 철학 외에도 인지과학 및 심리학에서 자신의 심적 상태에 대한 믿음이 틀릴 수 있음을 보여주는 사례들이 다양하게 제시되고 있는 만큼, 현재 완전한 오류 불가능성 논제는 과거만큼 지지받고 있지는 못하고 있다. 그럼에도 여전히 자기 지식은 다른 지식 일반과 비교하면 더 안전하고 확실하다는 보다 약한 주장을 옹호하거나 오류 불가능성을 자기 지식 중에서도 더욱 제한적인 범위에는 적용할 수 있다는 주장을 옹호하는 견해들은 존재한다.

둘째로는 자기 지식이 지식 일반과는 구분되는 특수한 방식을 통해서 획득되는지에 관한 논의가 있다. 이는 앞서 언급한 자기 지식의 인식적 확실성과도 관련될 수 있다. 자기 지식에만 적용이 가능한 특수한 지식 획득 방식이 존재한다고 했을 때, 이 방식이 다른 지식 획득 방식들보다 인식적으로 더 안전하고 확실할 수 있기 때문이다. 그리고 이러한 방식으로 접근할 수 있는 것은 자기 자신의 심적 상태 뿐이기에, 인간은 자신의 심적 상태에 대한 특권적 접근이 가능하다고 할 수 있다. 이러한 특권적 접근에 대한 한 가지 해명으로서 제시할 수 있는 것이 인간이 외부가 아닌 자신의 내부를 들여다보는 방법으로서의 내성(Introspection)이다. 자기 지식에 관한 내성 이론을 지지하

는 이들은 내성을 올바르게 진행함으로써 인간은 자신의 내부에 있는 심리 상태들에 대한 지식을 획득할 수 있다고 본다. 내성이 정확히 어떤 능력 혹은 절차인가에 대해서는 다시 의견들이 나뉘는데, 크게는 대면(Acquaintance) 이론을 지지하는 입장과 내적 감각(Inner Sense) 이론을 지지하는 입장으로 나뉜다. 대면 이론은 인간이 스스로의 안에 있는 심적 상태들을 직접적으로 마주침으로써 그에 관한 지식을 획득할 수 있다는 입장이다. 이때의 마주침은 형이상학적으로나 인식론적으로나 직접적이며, 추론이나 다른 인식적 절차를 통해 매개될 필요가 없다. 그런 점에서 대면은 기초적이며 근본적인 것으로 이해된다. 반면, 내적 감각 이론은 인간이 자기 내부의 심적 상태들을 내성하는 방식이 외부 대상을 지각하는 방식과 크게 다르지 않다고 보는 입장이다. 외부 대상을 지각하는 감각과 내성이 유사하다고 본다는 점에서 해당 이론은 내적 감각 이론이라 불린다. 이때 내적 감각은 뇌 내에 물리적 기반을 가지는 메커니즘으로, 내성은 인식적이고 인과적인 절차를 통해 진행되는 것으로 이해된다. 눈이 외부 사물에 반사된 빛과 인과적으로 상호작용하여 시각적 표상을 형성하는 것과 같이, 내적 감각은 내부의 심적 상태와 인과적으로 상호작용하여 심적 상태의 표상을 형성한다는 것이다.

내성 이론이 우리가 내부에 있는 심적 상태를 들여다봄으로써 자기 지식을 획득한다고 주장한다면, 자기 지식에 관한 이론으로서 내성 이론과는 경쟁 관계에 있는 투명성(Transparency) 이론은 우리가 마치 투명한 유리창처럼 내부의 심적 상태를 통과해서 그 심적 상태의 내용을 살펴봄으로써 자기 지식을 획득한다고 본다. 내가 P라는 믿음을 가지고 있다고 할 때, 내가 해당 믿음을 가지고 있음을 나는 어떻게 알 수 있는가? 투명성 이론의 지지자들은 내가 내성을 통해 나 자신의 내부를 들여다본 후에 P라는 믿음이 나의 안에 있음을 발견한다고 설명하기보다는, 그저 나에게 P라는 내용이 참으로 받아들여지는지를 살펴보면 된다고 말한다. P가 나에게 참이라면, 나는 P라는 믿음을 가지고 있다. 앞서 사용한 예시를 다시 가져와 보자. 나는 대한민국의 수도가 서울이라고 믿는다. 이때 나는 내가 이러한 믿음을 가졌는지를 어떻게 아는가? 투명성 이론에 따르면, 나는 내성을 통해 나의 내부를 들여다볼 필요 없이, 단순히 대한민국의 수도가 서울인지만 답해보면 된다.

서론에서 밝힌 것과 같이, 본 논문에서는 자기 지식에 관한 여러 이론

가운데서도 내적 감각 이론에 집중할 것이다. 여기서 내적 감각 이론이 다른 이론들에 비해 가지는 장점이나 다른 이론들에 대한 비판을 자세하게 다루지는 않을 것이나, 간략하게나마 내적 감각 이론이 주목을 받을만한 이유를 소개해보도록 하겠다. 우선 내성 이론과 경쟁 이론인 투명성 이론을 살펴보자. 투명성 이론은 믿음에 관해서는 꽤나 설득력 있는 설명을 제시하는 것처럼 보인다. 그러나 과연 동일한 설명 방식이 모든 종류의 심적 상태에 대해서도 적용이 가능할 것인가? 앞서 사용했던 여러 심적 상태들의 예시를 다시 가져와 보자. '나는 대한민국의 수도가 서울이라고 생각한다', '나는 내 동생에게 화가 난 상태다', '나는 지금 왼쪽 팔꿈치가 아프다.''나는 지금 차가운 콜라가 매우 마시고 싶다', '나는 파인애플이 올라간 피자를 싫어한다', '나는 어제 길에서 들었던 노래를 지금 머릿속에서 흥얼거리고 있다' 이 여섯 가지 예시들 가운데서 첫 번째 예시만이 나의 믿음 상태에 해당하고, 나머지 다섯 가지는 믿음이 아닌 다른 심적 상태들을 설명하고 있다. 그런데 나머지 다섯 가지의 예시들에서 나의 자기 지식의 근거로 삼을만한 외부 사실이 존재하는지는 의문스럽다. '나는 지금 차가운 콜라가 매우 마시고 싶다'를 예로 들면, 내 앞에 차가운 콜라가 놓여져 있다거나, 내가 콜라를 바라보면서 침을 삼키는 것과 같은 외적인 사실들은 내가 정말 콜라를 마시고 싶은 강한 욕구를 지금 가졌는지에 대한 근거가 전혀 되지 못한다. 그럼에도 나는 당연하고도 정당하게 내가 콜라를 마시고 싶다는 것을 알 수 있다. 우리가 가질 수 있는 자기 지식은 믿음에 대해서만 국한되어 있지 않으며, 완전한 자기 지식 이론이라면 다양한 심적 상태들에 대해 우리가 갖는 광범위한 자기 지식을 설명할 수 있어야 한다. 그렇기에 투명성 이론은 완전한 자기 지식 이론으로서는 한계가 있고, 그에 비해 내성 이론의 설명에서는 내성이 믿음뿐 아니라 모든 혹은 최소한 더 많은 종류의 심적 상태에 대해 가능하다는 측면에서 내성 이론은 설명적 포괄성에서 앞선다고 할 수 있다.

그렇다면 같은 내성 이론의 일종인 대면 이론과 비교했을 때 내적 감각 이론은 어떠한 장점이 있는가? 대면 이론에 따르면 내성은 우리의 심적 상태를 직접 마주하게 해주는 것이며, 이렇게 직접 마주한 심적 상태는 현상성을 지닌다. 거틀러(Gertler)는 이와 관련해 대면 이론이 극복해야 할 난점으로서 모든 심적 상태가 고유한 현상성을 지녀서 그를 통해 심적 상태들의 개별

화가 가능한지를 보일 수 있어야 한다고 지적한다.[1] 감각 상태들은 고유한 현상성이 있다는 것이 비교적 널리 받아들여지고 있지만, 믿음과 같은 다른 심적 상태들도 고유한 현상성이 있는지는 아직 논쟁거리로 남아있다. 이에 비해 내적 감각 이론은 심적 상태의 현상적 속성에만 의존해서 내성이 이루어진다고 주장할 필요가 없다. 따라서 믿음을 비롯한 여러 심적 상태들이 현상적 속성을 갖지 않는다고 인정하더라도 여전히 해당 심적 상태들에 대한 내성을 설명할 수 있다. 그런 점에서 내적 감각 이론은 대면 이론에 비해 다양한 심적 상태들에 대한 내성을 설명하기에 더욱 유리하다.

또한 대면 이론에 비해 내적 감각 이론은 자연과학 및 물리주의에 더욱 친화적이다. 대면 이론에서 정신은 심적 상태와 형이상학적으로 직접 대면하며, 이러한 대면이 심적 상태를 수반하는 물리적 상태와 어떠한 연관성이 있는지를 설명하는 것은 쉽지 않다. 특히 대면 이론에서는 심적 상태의 현상성이 핵심적인데, 현상성이 물리적 기반을 갖는지부터가 이미 논쟁적인 사안이다. 물론 현상성에 관한 문제는 내적 감각 이론에서도 중요하게 다루어지며 이후 이에 관해 본 논문에서도 살펴볼 예정이지만, 내적 감각 이론은 대면 이론과 달리 내성을 인과적인 절차로 이해한다는 점에서 경험과학적 연구와 연결점을 찾기가 더욱 용이하다. 내적 감각 이론의 입장에서 물리주의를 전제했을 때, 인간의 내성이 갖는 물리적 기반은 심적 상태의 기반이 되는 물리적 상태와 인과적으로 연결되어 있는 기관이나 구조로 이해될 수 있고, 나아가 인간이 아닌 다른 유기체나 인공지능의 경우에도 내성의 물리적 기반을 동일한 방식으로 이해할 수 있다. 인공지능과 관련해 내적 감각 이론이 대면 이론에 비해 갖는 장점은 더 확실하다. 심적 상태들을 갖는 인공지능을 설계한다고 가정했을 때, 설계 과정에서 인공지능이 심적 상태들을 단순히 가질 뿐만 아니라 그에 대한 내성을 할 수 있게 만들기 위해선 어떤 조건을 만족하여야 하는가? 대면 이론의 입장에서 대체 어떤 물리적 기반이 존재해야 대면이 가능해지는지는 상당히 불분명해 보인다. 반면, 내적 감각 이론에 따르면 인공지능이 내성을 할 수 있기 위한 물리적 기반은 바로 인공지능의 심적 상태들과 적절하게 인과적으로 연결되어 자기 자신에 관한 믿음들을 산출하는 메커니즘이 존재하는 것으로 설명할 수 있다. 그렇다고 해서 내적 감각 이론이 물리주

1) Gertler, *Self-Knowledge*, The Stanford Encyclopedia of Philosophy (Winter 2021 Edition).

의를 전제해야만 하는 것은 아니다. 외부 지각은 분명 눈이나 코와 같은 신체 기관들과 물리적으로 연관되어 있는 것처럼 보이지만, 그렇다고 해서 외부 지각의 가능성이 곧 물리주의를 받아들여야 한다는 결론으로 이어지는 것은 아니며, 물리주의를 거부하는 이들이 외부 지각을 인정하지 않는 것도 아니다. 마찬가지로 내적 감각 이론 또한 물리주의와 독립적으로 성립할 수 있다. 그런 점에서 내적 감각 이론은 물리주의에 대해 열려있는 이론이라고 할 수 있다.

정리하자면 내적 감각 이론의 주요 장점은 크게 두 가지이다. 하나는 다양한 종류의 심적 상태들에 대한 내성을 모두 설명할 수 있는 통일적 이론이 될 수 있다는 점이고, 다른 하나는 자연과학과 물리주의에 친화적이면서도 물리주의에 형이상학적으로 반드시 개입하지는 않는다는 점이다.

2. 2. 내적 감각 이론과 기능주의

철학사적으로 내적 감각 이론의 기원은 로크에게서 찾아볼 수 있다.

> 이러한 관념의 원천은 모든 인간이 오롯이 자신 안에 가지고 있는 것이다. 그리고 이것이 비록 외부 대상과는 무관하다는 점에서 감각은 아닐지라도, 감각과 매우 유사하며, 내적 감각이라고 충분히 불릴 만하다.
>
> (Locke, *An Essay on Human Understanding*, Book II)

로크는 이러한 내적 감각을 통해 정신이 그 내부적 작용들에 대한 관념들을 얻을 수 있다고 보았고, 이러한 관념 획득 방법을 반성(Reflection)이라고 불렀다. 그는 인간이 가진 모든 관념이 외부 대상에 대한 감각과 내적 감각으로서의 반성의 두 가지 경로를 통해서만 획득될 수 있는 것으로 보았다. 로크는 내적 감각이 외적 감각과 유사하다고 보았고, 또한 내적 감각이 주의(Attention)를 요구한다고 보았다는 점에서 내적 감각 이론의 가장 초기 형태를 제시했다고 볼 수 있다. 이후 현대 분석철학적 전통에서는 대표적으로 데이비드 M. 암스트롱, 윌리엄 라이칸, 숀 니콜스, 스티븐 스티치, 알빈 골드만 등의 학자들이 내적 감각 이론을 이어받아 발전시켜왔다.

위의 학자들이 각자 제시한 내적 감각 이론들이 공유하는 내용은 다음과 같다. 내성이란 내적 감각이 일종의 내부 스캐너와 같이 심적 상태들을 감지한 다음, 해당 심적 상태들에 관한 메타-표상을 산출함으로써 이루어지는 인과적인 작용이다. 이는 마치 눈이 외부 대상들에 반사된 빛을 감지하여 해당 대상들에 관한 시각적 표상을 산출하는 것과 유사하며, 이와 같은 유사점이 바로 내성 메커니즘을 내적 감각이라고 부를 수 있는 이유라 할 수 있다. 반면 이들의 이론들이 가장 차별화되는 지점은 내적 감각이 구체적으로 어떻게 여러 다른 내용의 심적 상태들을 내성하는지이다. 심적 상태의 종류는 고통을 느끼는 것과 같은 감각적 상태, '대한민국의 수도는 서울이다'라고 믿는 것과 같이 표상적 내용을 갖는 상태, 혹은 신호등의 빨간불을 바라볼 때의 지각적 경험에 따라오는 현상적 상태 등 다양하다. 바람직한 내적 감각 이론은 이러한 다양한 심적 내용들에 대한 내성뿐만 아니라 심적 상태의 종류 및 강도에 대한 내성도 설명할 수 있어야 한다. 예컨대 내가 오늘 비가 올 것이라고 믿고 있으며 이러한 믿음 상태에 대해 내성한다고 하자. 내적 감각 이론이 설명해야 하는 것은 크게 세 가지다. 첫째는 나의 심적 상태의 내용이 '오늘 비가 올 것이다'라는 것을 내성할 수 있다는 사실이다. 둘째는 나의 심적 상태의 종류가 욕구나 다른 어떤 심적 상태가 아니라 믿음이라는 것을 내성할 수 있다는 사실이다. 셋째는 나의 믿음이 얼마나 강한지를 내성할 수 있다는 사실이다. 심적 상태의 종류, 내용, 강도에 대한 내성을 설명하는 방식에서 여러 내적 감각 이론들은 차이를 보이고 있다.

이러한 설명 방식의 차이는 내적 감각 이론의 지지자들이 저마다 심적 상태에 대해 각자 다른 형이상학적 입장을 가지고 있다는 점과 관련된다. 이는 사뭇 당연한 결과라고 할 수 있는데, 심적 상태가 정확히 어떤 상태인지에 따라 내적 감각이 심적 상태와 인과적으로 작용하는 방식도 달라져야 할 것은 분명하기 때문이다. 환원적 물리주의를 받아들여 심적 상태가 물리적 상태의 일종인 신경적 상태라고 주장한다면, 내적 감각은 신경적 상태와 인과적으로 작용하는 메커니즘이어야 할 것이다. 예컨대 고통이 C-섬유의 활성화라는 신경적 상태와 동일하다면, 고통을 내성한다는 것은 내적 감각이 C-섬유의 활성화라는 신경적 속성을 감지하여 그에 관한 표상을 산출하는 과정이 된다. 혹은 기능주의를 받아들여 심적 상태가 기능적 속성에 따라 개별화된다고 주

장한다면, 내적 감각은 기능적 속성을 감지할 수 있어야 할 것이다. 고통의 기능적 정의가 신체적 손상에 의해 야기되고 찡그림과 움찔거림을 야기하는 상태라고 한다면, 고통을 내성하는 것은 이와 같은 기능적 속성을 감지함으로써 고통에 관한 표상을 산출하는 과정일 것이다. 그렇다면 이제 각각의 학자들이 제시한 내적 감각 이론의 기본적인 내용들과 더불어 그들이 심적 상태에 대하여 어떤 형이상학적 입장을 갖고 있으며, 이것이 그들의 내적 감각 이론과 어떻게 연관되는지를 살펴보도록 하겠다.

본 논문에서는 특히 기존의 내적 감각 이론가들이 기능주의에 대해 어떤 입장을 갖고 있는지에 주목해보고자 한다. 기능주의란 심적 상태의 정체성이 심적 상태 자체의 내적인 구성이나 속성에 의해서가 아니라, 그것이 속하는 인지적 체계 내에서 갖는 기능에 따라 정해진다는 입장이다. 예를 들어 고통 상태, 즉 고통 속에 놓여있는 상태가 일반적인 인간에게서 어떤 기능을 갖는지 살펴보면, 고통이란 신체적 손상으로 인해 야기되며, 찡그리거나 신음하는 행동을 야기하고, 또한 해당 상태를 벗어나고자 하는 욕구를 야기하는 경향을 갖는다. 이런 방식으로 심적 상태를 기능적으로 정의할 때 하나의 심적 상태는 보통 생리학적 자극과의 인과적 관계, 신체적 행동과의 인과적 관계, 그리고 다른 심적 상태와의 인과적 관계를 통해 정의된다. 기능주의를 받아들이는지 아니면 거부하는지, 만약 받아들인다면 어떤 종류의 기능주의를 받아들이고 있는지는 기존의 내적 감각 이론들에 있어서 중요한 차이를 발생시킨다. 기능주의의 종류를 나누는 기준은 크게 두 가지가 있다. 하나는 심적 상태의 기능적으로 정의하는 방식에 관한 입장에 따라 구분하는 것이고, 다른 하나는 기능적 역할을 수행하는 물리적 상태와 기능적으로 정의된 심적 상태의 관계에 대한 입장에 따라 구분하는 것이다. 전자는 기능주의를 기계-상태(Machine-state) 기능주의, 분석적(Analytic) 기능주의[2], 심리적(Psycho-) 기능주의의 세 가지로 구분하고, 후자는 기능주의를 역할(Role) 기능주의와 실현자(Realizer) 기능주의의 두 가지로 구분하는데, 이 두 가지 기준은 서로 직교하는 기준이기에 이론적으로는 총 여섯 가지 종류의 기능주의적 입장이 구분될 수 있지만, 기계-상태 기능주의는 현재 철학자들 사이에서 그다지 지지받

2) 블록은 분석적 기능주의라는 명칭 대신 대문자를 사용한 기능주의(Functionalism)라는 명칭을 사용한다.

지 못하고 있을뿐더러 현재 논의와는 크게 관련이 없기 때문에 배제하도록 하겠다. 그렇다면 일단 이론적으로는 분석적 역할 기능주의, 분석적 실현자 기능주의, 심리적 역할 기능주의, 그리고 심리적 실현자 기능주의라는 총 네 가지의 기능주의적 입장이 나뉠 수 있다.

우선 분석적 기능주의와 심리적 기능주의는 심적 상태의 기능적 정의가 어떠한 종류의 심리학적 이론에 기반하고 있는지에 따라서 나뉘게 된다. 심적 상태들의 입력값, 출력값, 그리고 상호관계에 관한 여러 이론들 가운데는 다양한 심적 상태들에 관해 사람들이 일상적으로 가지고 있는 상식들을 분석해 얻어진 것들도 있고, 일반인이 쉽게 접근할 수 없는 뇌과학적 탐구의 결과물들과 같은 전문적이고 경험적인 지식으로 구성된 이론도 있을 것이다. 선험적이고 분석적인 이론적 지식을 통해 심적 상태를 기능적으로 정의하고자 하는 것이 분석적 기능주의고, 경험적이고 과학적인 이론을 통해 심적 상태를 기능적으로 정의하고자 하는 것이 심리적 기능주의다.

기능주의적 입장들을 구분하는 다른 한 가지 중요한 방식은 기능적 역할을 수행하는 물리적 상태인 실현자 상태와 기능적으로 정의된 심적 상태의 관계를 어떻게 이해하는지에 따라 구분하는 것이다. 우선 이러한 구분은 모든 기능적으로 정의된 심적 상태에 대하여 어떤 물리적 상태가 정의된 기능적 역할을 수행한다는 전제하에 이루어질 수 있다. 예를 들어 고통이라는 심적 상태를 신체적 손상에 의해 야기되고, 찡그림이나 신음을 야기하며, 해당 상태를 벗어나려는 욕구를 야기하는 역할을 수행하는 상태라고 기능적으로 정의한다고 하자. 그리고 인간에게서 C-섬유의 활성화라는 물리적 상태가 해당 역할을 수행한다고 하자. 이때 C-섬유의 활성화는 고통 상태를 실현하는 물리적 상태다. 그런데 여기서 더 나아가 고통은 C-섬유의 활성화라는 물리적 상태와 동일하다고 볼 수 있는가? 아니면 앞서 나열된 관계적인 고차 속성들의 연언으로서의 기능적 역할 자체가 고통인가? 역할 기능주의는 연언적 고차 속성으로서의 역할 자체가 심적 상태이며, 그것을 실현하는 물리적 상태는 단지 심적 상태와 실현 관계 혹은 수반 관계에 놓여있을 뿐이라고 본다. 반면, 실현자 기능주의는 실현하는 물리적 상태가 곧 심적 상태와 동일하며, 기능적 역할은 물리적 상태인 실현자 상태에 대한 한정적 기술을 제공하는 것이라고 본다.[3)]

이러한 구분하에서 내적 감각 이론을 지지한 학자들이 기능주의에 대해 어떤 입장을 취하고 있는지를 나눠보면, 암스트롱은 분석적 실현자 기능주의를 받아들이고, 라이칸은 심리적 역할 기능주의를 받아들이며, 니콜스와 스티치는 심리적 기능주의를 받아들이면서 역할 기능주의와 실현자 기능주의 양쪽에 대해 열려있다고 할 수 있다. 한편 골드만은 기능주의를 거부하며, 내적 감각 이론이 기능주의와 맞지 않는다고 본다.[4] 이렇듯 내적 감각 이론의 지지자들 사이에서도 기능주의에 대한 입장은 서로 엇갈리는데, 바로 이러한 입장 차이가 그들의 내적 감각 이론의 설명력 및 이론적 강점에도 영향을 미치고 있다. 이제 각 학자의 내적 감각 이론의 내용과 그들이 기능주의에 대해 취하는 입장을 살펴보고, 각각의 이론이 내성 이론으로서 충실한 설명을 제시하고 있는지, 그리고 앞서 소개한 내적 감각 이론의 장점을 잘 갖추고 있는지 확인해보도록 하겠다.

2. 3. 기존 내적 감각 이론들

2. 3. 1. 암스트롱

암스트롱은 지각이 외부에 있는 물리적 세계의 상황을 대상으로 갖는 심적 사건이라면, 내성은 정신 내부에서 벌어지는 일을 대상으로 갖는 심적 사건이라고 구분한다. 그는 지각과 내성의 차이가 대상의 차이에 불과하기에 내적 감각이란 표현은 정당하다고 말하며, 내성이란 곧 뇌 내의 자기-스캔 과정이라고 말한다.[5] 이제 암스트롱의 내적 감각 이론의 주요 논점들을 살펴보겠다.

첫째로, 내적 감각을 통해 내성되는 심적 상태와 내성하는 심적 상태는 별개의 심적 상태이다. 내가 P라는 믿음을 갖는 것과, 내가 P라는 믿음을 가졌다는 것을 아는 것은 서로 다른 심적 상태라는 것이다. 이는 심적 상태들

3) Levin, *Functionalism*, The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)
4) 골드만이 명시적으로 거부하는 것은 분석적 기능주의지만, 그가 기능주의를 거부하는 이유를 살펴보면 심리적 기능주의에도 적용될 수 있는 이유들로 보인다.
5) 이때 암스트롱은 환원적 물리주의를 전제하고 있다. 심적 상태가 뇌의 물리적 상태와 동일하다면 내적 감각은 뇌 내의 물리적 상태를 감지하는 감각이라고 할 수 있다.

이 자기-시사적(Self-intimating)이거나 발광한다(Luminous)는, 즉 하나의 심적 상태 자체에 그 심적 상태에 대한 자기 지식을 구성하는 심적 상태가 포함된다는 주장을 거부하는 것이라 할 수 있다. 그리고 내성하는 심적 상태가 하나의 독립적인 심적 상태인 만큼 이는 다시 다른 내성하는 심적 상태의 내성 대상이 될 수 있다. 나는 P라는 믿음을 가질 수 있고, 해당 믿음을 내성해 다시 '나는 P라고 믿는다'는 믿음을 가질 수 있고, 이를 다시 내성해 '나는 '나는 P라고 믿는다'라고 믿는다'는 믿음을 가질 수 있는 식으로 말이다. 암스트롱은 이러한 고차 믿음의 생성은 자동적이지 않기에 믿음이 무한 증식하는 악순환은 발생하지 않으며, 내성된 심적 상태를 다시 내성하는 식의 연쇄는 논리적으로는 뇌의 물리적 한계를 고려했을 때 어딘가에서 끝이 나야 하지만, 연쇄가 필연적으로 끝나야 하는 지점이 정해져 있는 것은 아니라고 본다.

둘째로, 모든 심적 상태가 반드시 내성되는 것은 아니다. 현재 우리가 갖고 있는 심적 상태들 가운데 내성되고 있는 것들은 의식적인 심적 상태들이고, 내성되지 못하고 있는 것들은 무의식적인 심적 상태들이다. 이는 지각의 대상이 되는 외부 사태들을 지각되고 있는 사태들과 지각되지 않고 있는 사태들로 구분하는 것과 마찬가지이다. 앞서 암스트롱이 고차 믿음의 무한 증식의 악순환이 발생하지 않는다고 본 이유 또한 모든 심적 상태가 반드시 내성되는 것은 아니기 때문이다.

셋째로, 내성은 지각과 달리 특정한 신체 기관을 통해서 이루어지는 것이 아니다. 시각은 눈, 청각은 귀 등의 신체 기관을 통해서 이루어지지만, 모든 지각이 특정 신체 기관을 통해서 이루어지는 것은 아니다. 마찬가지로 내적 감각을 담당하는 단일한 감각 기관이 존재해야 하는 것은 아니다. 대신 내적 감각은 특정한 기관 대신 중추 신경계를 구성하는 구조 등을 통해서 심적 상태를 내성한다.

넷째로, 내성은 언어에 의존하지 않으나, 언어와 내성을 완전히 분리할 필요도 없다. 언어 능력이 없더라도 내성이 가능하다는 것은 언어를 습득하지 않은 어린아이들조차도 자신이 고통을 느낄 때 그 고통을 의식한다는 것에서부터 확인할 수 있다. 그러나 이는 내성이 언어와 무관하게 이루어진다는 것을 의미하는 것은 아니다. 언어 사용 능력이 발달할수록 자기 자신의 심적 상태에 대해 더욱 정교하고 복잡한 믿음을 가질 수 있고, 언어적 내용을 가진

심적 상태에 대한 내성도 가능해진다는 점에서 언어 능력과 내성 능력은 연결되어 있다는 것은 경험적으로 알려진 바이다.

다섯째로, 내성은 심적 상태와의 직접적인 대면이 아니며, 그저 대상에 관해 옳을 수도, 틀릴 수도 있는 정보를 얻는 과정에 불과하다. 그렇기에 내성은 오류 불가능하지 않다. 외부 지각이 잘못된 믿음으로 이어질 수 있는 것과 마찬가지로, 내적 감각 또한 잘못된 자기 믿음의 형성으로 이어질 수 있다. 그러나 지각이 오류를 범할 수 있다고 해서 우리가 지각을 통해 얻은 모든 믿음의 인식론적 지위를 부정하지 않듯이, 내적 감각이 오류 가능성을 내포한다고 해서 내적 감각을 통해 획득된 믿음이 지식으로 성립하지 못하는 것은 아니다. 내적 감각 이론에 친화적인 인식론적 이론으로는 신빙론이 있는데, 신빙론에 따르면 내적 감각이 대상에 관하여 오류를 범할 가능성이 있더라도 여전히 신빙성 있는 믿음 형성 과정이라면 내적 감각을 통해 형성된 믿음은 정당화된다.

이와 같은 암스트롱의 논의는 현대 내적 감각 이론의 기초를 제공했고, 이후 다른 내적 감각 이론가들도 위에 소개된 논점들을 대체로 이견 없이 받아들이고 있다. 암스트롱의 내적 감각 이론은 모든 종류의 심적 상태에 대한 내성을 설명하는 완전한 자기 지식 이론으로서 제시되었는데, 이는 암스트롱이 심적 상태가 뇌 혹은 중추 신경계의 물리적 상태와 동일하다고 보았기 때문이라 할 수 있다. 내적 감각은 일종의 신체 내부에 존재하는 스캐너로서 물리적 상태인 심적 상태들을 스캔하고, 심적 상태의 종류가 달라지더라도 물리적 상태라는 점은 변하지 않기 때문에 내적 감각은 모든 심적 상태에 대해 내성할 수 있는 것이다. 그러나 동시에 암스트롱은 내성이 심적 상태로서의 심적 상태(Mental states qua mental states)에 대해 이루어져야 한다고 주장한다. 이때 암스트롱에게 심적 상태란 특정한 신체적 자극이나 다른 심적 상태에 의해 야기되고, 또 특정한 종류의 행동이나 다른 심적 상태를 야기하기에 적합한 상태이며, 이때 적합하다는 것은 경향적(Dispositional)이다.[6] 심적 상태를 이와 같은 방식으로 경향적인 인과적 관계들을 통해 정의한다는 점에서 암스트롱은 심적 상태에 대한 기능주의를 채택하고 있다. 더 구체적으로 들어가면, 암스트롱은 심적 상태가 물리적 상태와 동일하다고 본다는 점에서

6) Armstrong, *The Nature of Mind* (1981)

실현자 기능주의를 따르고 있으며, 심적 상태들의 기능적 정의는 일상적인 심리학적 지식들의 분석을 통해 얻어진다고 본다는 점에서 분석적 기능주의를 따르고 있다. 즉, 그는 분석적 실현자 기능주의를 받아들인다.

그렇다면 암스트롱의 이론에서 내적 감각이 내성하는 것은 심적 상태의 어떤 속성이어야 하는가? 암스트롱 본인은 내적 감각이 어떤 종류의 속성을 통해 심적 상태를 내성하는가에 관해 명시적으로 논의한 바는 없으나, 그가 견지하고 있는 기능주의적 관점과 그의 다른 발언들을 통해 그의 입장을 유추해볼 수 있다. 암스트롱은 심적 상태가 신경적 상태와 같은 물리적 상태와 동일하다고 보고 있기에 그의 이론에서 심적 상태는 신경적 속성과 같은 물리적 속성도 지니게 된다. 그렇다면 이러한 속성도 내적 감각의 감지 대상이 될 수 있는가? 예컨대, 고통을 내적 감각이 감지할 때 내적 감각이 C-섬유의 활성화와 관련된 신경적 속성을 감지함으로써 내성을 진행할 수 있는 것인가? 그렇지는 않은 것으로 보인다. 암스트롱의 말처럼 '심적 상태로서의 심적 상태'에 대해 내성이 이루어지기 위해서는 내성이 심적 상태의 기능적 속성을 통해서 이루어져야 할 것이다. 어떠한 물리적 상태가 심적 상태와 동일할 수는 있지만, 해당 상태가 그저 물리적이기만 한 상태가 아니라 특정 심적 상태와 동일하게 만들어주는 것은 그 물리적 상태가 특정한 기능적 속성을 갖고 있기 때문이다. 그리고 암스트롱이 분석적 기능주의를 받아들이는 만큼, 이때 심적 상태의 기능적 속성이란 상식적 심리학(Folk psychology)의 분석으로 알려지는 종류여야 하고, 특수 과학에 의해 알려지는 것이어서는 안된다. 상식적 심리학에 기반한 고통을 대략 기능적으로 정의해보면, 고통이란 신체적 손상에 의해 야기되는 경향이 있고, 찡그림을 유발하는 경향이 있고, 해당 상태를 벗어나려는 욕구를 야기하는 경향이 있는 상태이다. 물론 이러한 기능적 조건들을 만족시키는 물리적 상태가 C-섬유의 활성화라는 신경적 상태일 수는 있다. 그러나 C-섬유의 활성화라는 신경적 상태가 갖는 물리적 속성과 해당 신경적 상태와 동일한 심적 상태가 갖는 기능적 속성은 구분될 필요가 있으며, 이러한 구분 하에서 암스트롱은 내적 감각의 내성에 관여하는 속성이 물리적 속성이 아니라 기능적 속성이라고 볼 것이다. 즉 고통을 내성한다는 것은 내적 감각이 C-섬유의 활성화를 감지하는 것이 아니라, 신체적 손상에 의해 야기되고 찡그림을 유발하는 등 고통이 갖는 기능적 속성을 감지함으로써 이루

어진다.

그렇다면 이제 앞서 언급했던 내적 감각 이론의 두 가지 주요 장점을 암스트롱의 이론이 잘 살리고 있는지 확인해보자. 우선 암스트롱의 내적 감각 이론은 믿음이나 감각과 같은 특정한 심적 상태에 대한 내성에 관해서만이 아니라, 모든 심적 상태에 대한 내성에 관한 통일적인 설명을 제공하고자 한다는 점에서 내적 감각 이론의 첫 번째 장점을 잘 살리고 있다고 할 수 있다. 또한 암스트롱의 이론은 실현자 기능주의를 받아들여 환원적 물리주의 또한 받아들이고 있으면서도, 내적 감각이 심적 상태의 물리적 속성을 감지함으로써 심적 상태를 내성하는 것이라고 보고 있지는 않으며, 오히려 내성이 심적 상태의 기능적 속성에 기반해 이루어진다고 보고 있다. 그렇다면 암스트롱의 내적 감각 이론은 이원론자 등 물리주의를 받아들이지 않는 측에서도 심적 상태가 기능적 속성을 통해 개별화될 수 있다는 것에 동의할 수 있다면 충분히 받아들일 수 있다.

암스트롱의 내적 감각 이론에 대해서는 다음과 같은 문제가 제기될 수 있다. 바로 암스트롱이 제시한 내적 감각 모델이 다양한 심적 상태의 내성을 설명해낼 수 없다는 것이다. 암스트롱은 자신의 내적 감각 이론을 자기 지식에 관한 통일적인 이론으로서 제시했지만, 사실 그는 내성이 내부 스캐너와 같은 내적 감각을 통해 이루어진다는 주장 외에는 구체적으로 심적 상태들의 내용 및 종류가 어떻게 감지되는지에 대해서 크게 논한 바가 없다. 대표적으로 내적 감각이 감지할 수 있어야 하는 심적 상태의 내용에는 현상적 내용이 있다. 그런데 현상적 내용에 대한 내성이 과연 심적 상태의 기능적 속성을 감지하는 내부 스캐너에 의해서 가능할까? 이것이 가능하기 위해서는 심적 상태의 현상적 속성이 기능적 속성으로 환원되거나 설명될 수 있어야 할 터이다. 여러 학자들은 현상적 속성이 기능적 속성으로 환원될 수 없다고 보고 있다는 사실은 분석적 실현자 기능주의를 받아들이는 암스트롱의 내적 감각 이론에 부담으로 작용한다. 이후 라이칸의 논의에서 살펴볼 것처럼 이러한 문제에 대해 기능주의적 입장에서 대응할 수 있는 여지가 없는 것은 아니지만, 암스트롱은 본인은 해당 문제에 관해 추가로 설명해주는 바가 없다.

2. 3. 2. 라이칸

암스트롱이 현대 분석철학적 논의 맥락 내로 내적 감각 이론을 다시 가져온 이후, 내적 감각 이론을 이어받아 본격적으로 발전시킨 것은 라이칸이었다고 할 수 있다. 암스트롱의 이론에서도 이미 내적 감각은 심적 상태들이 의식적이냐, 무의식적이냐를 구분해주는 역할을 수행하는 것으로 이해되었고, 라이칸의 이론적 초점 또한 자기 지식의 문제보다는 의식의 문제에 맞추어져 있다.[7] 내적 감각에 의해 포착된 심적 상태는 의식적 심적 상태가 되며, 그렇지 못한 심적 상태는 무의식적 심적 상태가 된다.

라이칸이 제시한 내적 감각 이론의 주요 논점들을 정리하자면 다음과 같다. 첫째로, 우리의 뇌 혹은 중추 신경계 내에는 우리의 심적 내부 작용들에 대해 주목할 수 있는 메커니즘이 존재한다. 이러한 메커니즘의 존재는 인지과학이나 뇌과학적 발견들을 통해서 뒷받침되기도 하지만, 경험적 증거 이상으로 우리가 내성의 주목을 자발적으로 통제할 수 있다는 사실이 해당 메커니즘의 존재를 뒷받침한다.

둘째로, 내적 감각은 물리적 메커니즘으로서 오류를 범하거나 오작동할 수 있는 가능성을 지닌다. 이는 크게는 두 가지 인식적 문제를 발생시킬 수 있는데, 하나는 인식적 주체가 어떤 심적 상태를 갖고 있을 때 내적 감각이 오작동하여 해당 심적 상태에 대해 적절한 믿음을 형성하지 못하는 것이고, 다른 하나는 내성의 대상이 되는 심적 상태가 부재하는데도 내적 감각이 오작동하여 실제로는 갖고있지 않은 심적 상태에 대한 위양성을 결과로 내놓는 것이다. 라이칸은 양쪽 사례 모두가 경험적으로도 존재할뿐더러, 형이상학적으로도 문제가 되지 않는다고 진단한다.[8]

---

7) "... the inner-sense account was offered as a theory of the conscious/unconscious distinction in our original sense, not as a contribution to epistemology." Lycan, *Consciousness and Experience* (1996)

8) 슈메이커는 두 가지 문제들 중 전자와 관련해 자기맹(Self-blindness)과 합리성이 상충한다고 주장하며 자기맹의 가능성을 함축하는 내적 감각 이론을 비판한 바 있다. 슈메이커의 비판의 구조를 간략하게 소개하면 다음과 같다.
(1) 내적 감각 이론에 따르면, 어떤 심적 상태와 관련해서도 해당 심적 상태에 놓여있으면서 해당 심적 상태에 자신이 놓여있다는 것을 의식하지 못하는 자기맹적인 합리적 주체가 존재할 수 있다.
(2) 고통과 믿음을 비롯한 일부 심적 상태들과 관련해서는 자기맹적인 합리적 주체가 존재

셋째로, 내적 감각은 외부 지각과 반드시 완벽하게 닮아있을 필요는 없다. 지각적 경험은 현상성을 동반하지만, 그렇다고 해서 내적 감각을 통한 내성도 현상성을 꼭 가져야만 하는 것은 아니다. 물론 내성을 비롯한 모든 심적 작용이 고유한 현상성을 가진다고 주장하는 측도 존재하지만, 이는 내적 감각 이론 자체와는 독립적인 주제로서 내적 감각 이론의 지지자가 반드시 수용해야 할 주장은 아니다. 앞서 암스트롱이 제시한 내적 감각 이론에서도 내적 감각은 특정한 신체 기관에 의존하지 않는 등, 외부 지각과 차이를 보이는 지점들이 존재했다.

넷째로, 내적 감각은 심적 상태의 기능적 속성을 감지한다. 라이칸은 심적 상태에 대한 기능주의를 받아들이는데, 이는 그의 내적 감각 이론에서 내적 감각이 여러 심적 상태들을 구분할 수 있게 만들어주는 기반이 된다. 라이칸은 내적 감각의 출력값에 대해서는 비교적 분명하게 명시하고 있다. 내적 감각은 심적 상태에 대한 표상을 산출한다. 그리고 이때 내적 감각은 단순히 일차적 심적 상태의 내용을 그대로 재활용하는 것이 아니라, 추가적으로 그것을 특정한 방식으로 있는 것(Being a certain way)으로서 표상하는데, 이때의 특정한 방식은 해당 심적 상태가 갖는 내적인 기능적 역할에 의해 결정되는 것이다. 따라서 내적 감각에 의해 산출된 이차적 표상은 일차적인 심적 상태가 가지고 있지 않았던 내용이 포함된다. 그리고 바로 이러한 기능적 역할의 차이로 인해 내적 감각은 서로 다른 종류의 심적 상태들을 구분하여 표상할 수 있게 되고, 해당 표상은 내성의 결과로 획득된 심적 상태에 관한 믿음에서 어떤 심적 동사가 사용되는지를 정당화해준다.

특히 라이칸은 심적 상태의 고유한 현상성조차 기능적으로 설명할 수 있다고 보는데, 그렇기에 암스트롱의 이론과 달리 라이칸의 내적 감각 이론에서는 심적 상태가 갖는 현상적 내용에 대한 내성 또한 내적 감각이 심적 상태의 기능적 속성을 감지함으로써 이루어질 수 있다.[9] 여러 학자들은 심적 상태

할 수 없다.

(3) (1)과 (2)는 모순이다.

슈메이커가 제기한 비판에 대해서는 다양한 답변이 제기되었지만, 필자가 보기에 가장 효과적인 답변은 거틀러가 제시한 답변이다. 거틀러에 따르면, 내적 감각 이론가는 슈메이커가 주장하는 합리성 개념을 받아들이고서, 내적 감각이 자주 오류를 범하는 주체는 합리성의 기준에 미달한다고 답변할 수 있다. 그렇다면 슈메이커의 비판에서 (1)은 참이 아니다. 내적 감각 이론에 따르면 자기맹적인 주체는 존재할 수 있지만 해당 주체는 합리성의 기준을 만족하지 못하며, 자기맹적인 합리적 주체는 존재할 수 없다.

의 현상적 내용이 기능적으로 설명될 수 없다고 비판해 왔으며, 이때 자주 등장하는 것이 '전도된 스펙트럼' 사례이다. 해당 사례를 간단하게 설명하자면 다음과 같다. 색의 스펙트럼에서 초록색과 빨간색은 서로 반대되는 색들이다. 그런데 어떤 이에게는 보통 사람에게 초록색으로 보이는 것이 빨간색으로 보이고, 반대로 보통 사람에게 빨간색으로 보이는 것이 초록색으로 보이는 식으로 색이 시각적으로 전도되어 느껴진다고 하자. 하지만 이 사람은 태어날 때부터 이런 상태였기에, 남들에게 초록색으로 보이는 것을 빨간색으로 보면서도 그것을 초록색이라고 부른다. 문제는 이러한 사람이 서로 다른 물체들의 색의 동일함이나 유사성을 보통 사람과 마찬가지로 구분할 수 있는 등, 보통 사람과 기능적으로는 동일할 수 있는 것처럼 보인다는 점이다. 즉, 이렇게 전도된 스펙트럼을 가진 이가 신호등의 빨간불을 볼 때 갖는 심적 상태는 보통 사람이 신호등의 빨간불을 볼 때 갖는 심적 상태와 기능적으로는 동일하지만, 현상적으로는 다를 수 있다는 것이다. 이러한 가능성을 인정한다면 심적 상태의 기능적 속성은 현상적 내용을 포함하지 못하며, 또한 내적 감각이 현상적 상태에 대한 내성을 수행하는 것은 기능적 속성을 통해서일 수가 없게 된다.

라이칸이 '전도된 스펙트럼'에 대응하는 방식은 전도된 스펙트럼을 가진 사람의 심적 상태가 일반적인 사람의 심적 상태와 기능적으로 동일할 수 있다는 주장을 정면으로 부정하는 것이다. 이를 주장하기 위해 라이칸은 기능과 그것을 실현하는 구조의 단순한 이분법을 거부하며, 한 단계에서는 구조적인 것으로 간주되었던 것이 더 하위 단계에서는 기능적인 것으로 이해될 수 있다고 말한다. 기능적 구성은 이분법적인 것이 아니라 연속적 단계들로 이루어지며, 가장 상위 단계의 기능들은 일상적인 상식 심리학에서도 파악될 수 있는 것들이지만 더욱 하위 단계로 갈수록 기능들은 신경적이고 화학적인 단계에 가까워진다. 이러한 구도하에서 라이칸은 전도된 스펙트럼을 가진 이가 행동으로 드러나는 상식 심리학의 단계에서만 보통 사람과 기능적으로 동일해 보이며, 더욱 하위 단계로 내려가면 보통 사람과 기능적인 차이를 가진다고 주장한다. 이와 같은 라이칸의 답변을 그의 내적 감각 이론에 적용해보면 여러 다른 종류의 심적 상태들이 모두 동일한 기능적 단계에 있지는 않으며, 현상적인 내용을 갖는 심적 상태는 비교적 낮은 기능적 단계에 있다고 말할 수

9) Lycan, *Consciousness and Experience* (1996)

있을 것이다. 그러나 낮은 단계라고 하더라도 여전히 기능적으로 정의되기 때문에 현상적 상태에 대한 내성도 기능적 속성을 통해 이루어질 수 있다.

여기서 드러나는 것은 암스트롱은 분석적 기능주의를 받아들였던 반면에, 라이칸은 심리적 기능주의를 받아들이고 있다는 점이다.[10] 현상적 상태와 같이 낮은 기능적 단계에 있는 심적 상태는 일상적인 상식 심리학의 분석을 통해서 기능적으로 정의될 수는 없으며, 점점 더 낮은 기능적 단계로 갈수록 신경적이고 화학적인 단계에 근접해지는 만큼 더 낮은 기능적 단계에 속하는 심적 상태들의 기능적 정의는 전문적인 과학적 탐구의 결과로만 얻어질 수 있게 된다. 또한 라이칸은 심적 상태가 물리적 상태와 동일하다고 말하지 않으며, 기능적 역할이 곧 심적 상태라고 보고 있다는 점에서 심리적 역할 기능주의를 받아들인다고 할 수 있다.

그렇다면 심리적 역할 기능주의를 받아들이는 라이칸의 내적 감각 이론은 앞서 소개한 내적 감각 이론의 장점들을 잘 가지고 있는지 확인해보자. 우선 라이칸의 내적 감각 이론 또한 암스트롱의 이론과 마찬가지로 모든 심적 상태에 대한 내성을 설명하고자 할 뿐만 아니라, 현상적 내용을 갖는 심적 상태들의 내성에 관해서는 분석적 기능주의를 받아들였던 암스트롱의 이론보다 더욱 잘 설명해낸다고 할 수 있다.

2. 3. 3. 니콜스와 스티치

니콜스와 스티치는 자기 지식을 설명하기 위해 모니터링 메커니즘 이론(약칭 MM이론)을 제시한다. 그들의 이론에 따르면, 인간의 인지적 구조에는 스스로의 심적 상태들을 감지하고 해당 심적 상태에 관한 자기 믿음을 산출하는 모니터링 메커니즘이 포함되어 있다. 바로 이 모니터링 메커니즘이 그들의 MM이론에서 내적 감각의 역할을 수행하는 것이다. 해당 메커니즘은 명제 태도들의 내용이 되는 표상들을 입력값으로서 갖고, 출력값으로서 믿음들을 산출한다.

---

10) 라이칸 본인은 자신의 기능주의적 입장을 목적론적 기능주의(Teleological functionalism), 또는 호문쿨루스적 기능주의(Homuncular functionalism, or homunctionalism)라고 부르고 있지만, 앞서 제시한 기능주의적 입장의 분류에 따르면 그의 기능주의는 심리적 기능주의에 속한다고 말할 수 있다.

이때 입력값이 될 수 있는 표상들은 명제 태도별로 구분된 일종의 '상자'에 담겨있다. 이와 관련해 이들은 우리가 가진 모든 믿음의 표상적 내용이 담겨있는 '믿음 상자'와 같은 심적 메커니즘을 상정한다. 믿음 상자에 P라는 내용을 갖는 표상이 담겨있다고 할 때, 해당 표상이 모니터링 메커니즘에 입력되면 해당 메커니즘은 자동으로 '나는 P라고 믿는다'라는 내용의 표상을 출력값으로 산출하고, 산출된 표상은 또 하나의 믿음의 내용이 되어 다시금 믿음 상자에 들어가게 된다.[11] 이제 내적 감각 이론으로서의 MM이론의 주요 특징을 살펴보도록 하겠다.

니콜스와 스티치의 이론의 특징 중 하나는 다양한 심적 상태의 종류에 대해 저마다 별개의 모니터링 메커니즘을 배정한다는 점이다. 즉, 내적 감각이 단일한 메커니즘이 아니라 여러 개의 메커니즘으로 구성된다고 보는 것이다. 우선 이들은 '믿음 상자' 외에도 '욕구 상자' 등 표상적 내용을 갖는 심적 상태마다 그 표상적 내용을 처리하는 별개의 메커니즘이 있다고 상정한다. 그리고 각각의 메커니즘에는 각각 연결된 모니터링 메커니즘이 있다. 여러 모니터링 메커니즘들은 저마다 연결된 심적 메커니즘으로부터 표상적 내용을 입력값으로 받고, 그에 맞추어 '나는 P라고 믿는다', '나는 P를 욕구한다' 등의 내용을 가진 표상을 출력값으로 내놓는 것이다. 이때 입력값들은 다양한 종류의 심적 상태의 표상적 내용이지만, 모든 출력값들은 공통적으로 믿음의 내용을 구성하는 표상으로서 믿음 상자에 들어가게 된다. 이는 심적 상태가 가진 표상적 내용이 일종의 재배치(Redeployment) 과정을 거치는 것이라 할 수 있다. 이러한 재배치 과정을 통해 니콜스와 스티치가 제시한 내적 감각인 모니터링 메커니즘은 우리 스스로의 심적 상태에 관한 믿음을 생성한다.

둘째로, 내적 감각의 입력값에 관해서 니콜스와 스티치는 오직 표상만이 모니터링 메커니즘에 입력될 수 있다고 본다. 즉, 내적 감각은 심적 상태의 표상적 속성을 통해서 심적 상태를 내성한다. 그렇기 때문에 그들은 자신들이 제시한 MM이론이 모든 종류의 심적 상태에 대한 내성을 설명하지는 못한다는 한계를 인정하며, 오직 표상적 내용을 갖는 심적 상태에만 한정된 설명을 제시한다. 이때 가장 문제가 되는 것은 지각적 경험과 관련된 심적 상태들인

11) Nichols and Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds* (2003)

데, 지각적 경험에는 현상적 내용이 포함되어 있기 때문이다. 내가 신호등의 빨간 불을 바라보고 있다고 하자. 이때 나의 시각적 경험은 신호등에 관한 표상적 내용 외에도 고유한 현상적 내용을 가질 것이고, 관련된 나의 심적 상태는 현상적 속성을 갖는다. 그러나 현상적 속성이 표상적 속성으로 환원되지 않는 한, 모니터링 메커니즘은 현상적 속성을 입력값으로 갖지 못하므로 MM 이론은 해당 심적 상태의 현상적 내용에 대한 내성을 설명하지 못한다. 직관적으로 우리는 현상적 내용에 대해서도 내성을 진행할 수 있는 것으로 보인다. 비록 그 내성에 관해 언어적으로 표현할 수 있는 방법은“나는 콜라병을 바라볼 때 이러이러한 느낌을 갖는다”에 불과할 수 있지만, 이때 ‘이러이러한 느낌’이 지시하는 바는 타인에게는 전달되지 않더라도 발화자 자신에게는 알려진다. 현상적 내용을 갖는 심적 상태에 대한 내성을 설명하기 위해서는 MM이론 외에 추가적인 이론적 자원이 필요한 것으로 보인다. 물론 현상적 속성이 온전히 표상적 속성으로 환원된다는 입장을 취한다면 이는 니콜스와 스티치의 이론에 큰 문제가 되지 않겠지만, 이들은 현상적 속성이 표상적 속성으로 환원될 수 있다는 견해에 대해 유보적이며, 따라서 지각적 상태에 관한 내성을 MM이론이 설명하기 위해서는 추가적인 이론적 장치가 필요하다는 점을 인정한다. 이들이 제시하는 해결책은 지각-믿음 매개 메커니즘을 추가로 상정하는 것이다. 해당 메커니즘은 현상적 내용을 입력값으로 갖고 그 내용과 적절히 관련된 내용의 표상적 믿음을 산출함으로써 지각적 상태의 현상적 내용을 표상적 내용으로 변환해준다. 그리고 이렇게 변환된 표상적 내용은 내적 감각의 입력값이 될 수 있다는 것이 니콜스와 스티치의 해결책이다.

셋째로, MM이론에 따르면 내적 감각의 출력값은 그저 심적 상태에 대한 표상이 아니라 그 자체로 이미 하나의 고차-믿음이다. 그러나 이러한 방식으로 어떠한 표상이 모니터링 메커니즘에 의해 생성되어 믿음 상자에 넣어졌다는 사실로 인해 우리가 해당 표상을 내용으로 갖는 믿음을 가진다고 설명하는 것이 과연 인식론적으로 적절한 설명인가 의문을 가질 수 있다. 내적 감각 이론은 지각과 내성의 유사성에 주목하는 이론이니만큼, 지각적 메커니즘에 의해 산출된 표상이 우리의 지각적 믿음을 정당화해줌으로써 지각을 통해 외부 세계에 대한 지식을 획득하는 과정과 내적 감각을 통해 자기 지식을 획득하는 과정은 유사해야 한다. 그러나 MM이론이 제시한 구도에 따르면, 모니

터링 메커니즘에 의해 산출된 표상은 우리의 믿음과 정당화 관계에 놓여있기보다는, '믿음 상자'에 집어넣어짐으로써 믿음이 된다. 이를 지각의 경우와 유비해보면, 지각적 메커니즘이 산출한 표상 또한 우리의 지각적 믿음을 정당화시켜주기보다는 '믿음 상자'에 집어넣어짐으로써 자동적으로 믿음을 형성한다는 식으로 설명되어야 한다.

니콜스와 스티치의 내적 감각 이론에서 설명이 필요한 채로 남아있는 주요 쟁점은 다음과 같이 정리할 수 있다. 첫째로는 우리가 심적 상태들을 내성할 때, 심적 상태들의 내용 외에도 그 심적 상태의 종류 및 강도 또한 표상된다는 사실을 내적 감각이 어떻게 설명할 것인지이다. MM이론에서 이에 대한 설명을 찾아보자면, 각 종류의 심적 상태를 관장하는 메커니즘마다 별도의 모니터링 메커니즘이 연결되어 있다는 점에서 실마리를 발견할 수 있다. 한 MM이 산출하는 자기-믿음에 어떤 심적 동사가 포함되는지는 그 모니터링 메커니즘이 어떤 심적 메커니즘과 연결되어 있는지에 의해 결정되고, 따라서 개별 모니터링 메커니즘만으로는 우리가 내성을 통해 현재 내성의 대상이 되는 심적 상태가 어떤 종류의 심적 상태인지를 파악할 수 있다는 사실을 설명할 수 없더라도, 각 모니터링 메커니즘이 여러 심적 상태에 연결되어 있는 총체적 구조를 통해서 이를 설명해 낼 수 있다고 주장할 수 있다. 그러나 이는 여전히 내성이 심적 상태에 대한 종류-표상(Type-representation)을 할 수 있다는 사실에 대한 설명을 내적 감각 자체의 입력값을 통해서 설명하고 있지 않으며, 심적 메커니즘들의 특정한 총체적 구조를 상정하는데 기대어 설명하고 있다는 한계를 지닌다고 할 수 있다. 이와 관련해 니콜스와 스티치는 내적 감각은 여러 모니터링 메커니즘의 총체로서 이해되어야 하며, 따라서 내적 감각 자체는 심적 상태에 대한 종류-표상을 충분히 설명해낼 수 있다는 반론을 제기할 수 있다. 그럼에도 여전히 내성을 위해서 모든 심적 상태 종류마다 별도의 메커니즘이 요구되는 것은 과도한 요구라는 비판은 피할 수 없다. 그리고 이와 같은 해결책을 받아들인다 하더라도, 여전히 심적 상태의 강도에 대한 표상이 어떻게 가능한지는 설명되지 않은 채로 남아있다. 예컨대 내가 '오늘은 비가 올 것이다'라고 매우 강하게 믿고 있고, 또한 '내일도 비가 올 것이다'라고 약하게 믿고 있다고 하자. 이러한 두 믿음들이 MM에 입력된 후 생성된 출력값들은 두 믿음의 강도의 차이를 어떻게 표상하는가? 믿음의 내용은 단순

한 재배치를 통해 설명될 수 있을지언정, 믿음의 강도는 재배치로 설명될 수는 없어 보인다. 내가 '내일도 비가 올 것이다'라고 약하게 믿고있다고 해서, '나는 내일도 비가 올 것이라고 믿는다'는 자기 믿음도 동일한 정도로 약할 필요는 없어 보이기 때문이다.

둘째로는 우리가 심적 상태의 현상적 내용에 관한 내성을 할 수 있다는 사실을 어떻게 설명할 것인지가 과제로 남아있다. 니콜스와 스티치는 현상적 내용을 표상적 내용으로 변환하는 추가적 메커니즘을 상정함으로써 이를 설명하고자 했는데, 과연 이러한 해결책이 정말로 모니터링 메커니즘을 통한 지각적 상태의 내성 문제를 해결해주는가는 의심스럽다. 앞서 말한 것처럼, 일상적 의미에서 우리가 스스로의 지각적 경험을 내성할 때 우리는 그 현상적 특성에 대해서도 주목할 수 있다. 내가 신호등의 빨간불을 보고 있음을 내성할 때, 나는 바로 그 빨간 불빛을 보는 것이 어떠한 느낌인지에 대해 내성할 수 있는 것이다. 즉, 우리는 심적 상태의 현상적 내용에 대한 내성이 가능하다. 그러나 위에 제시된 해결책에 따르면, 모니터링 메커니즘을 통한 지각적 상태의 내성은 현상적 내용 자체에 대해서는 전혀 이루어지지 않으며, 이미 표상적으로 변환된 내용에 대해서만 이루어진다. 따라서 현상적 내용을 갖는 지각적 상태에 대한 내성에 관한 설명이 제시되지 못한다는 본질적인 문제점은 실질적으로 전혀 개선된 바가 없다. 또한 MM이론은 표상이 '믿음 상자'에 집어넣어짐으로써 믿음이 생성된다는 식으로 설명하는데, 이런 점에서 피치우토와 카러더스는 니콜스와 스티치의 이론이 내적 감각 이론이라고 말하기에는 그들이 제시한 모니터링 메커니즘이 출력값을 비롯해 지각적 메커니즘과 유사하거나 공통된 점이 지나치게 부족하다고 비판한 바 있다.[12] 이런 점에서 이들의 내적 감각 이론은 모든 심적 상태에 대한 내성을 설명하는 통일적 이론이라는 내적 감각 이론의 장점을 잃어버린다. 한편 이들의 내적 감각 이론은 라이칸과 마찬가지로 심리적 기능주의를 받아들인다고 할 수 있다. 한 심적 상태를 바로 그 심적 상태로 만들어주는 것은 표상적 내용을 담은 토큰이 특정한 박스, 즉 특정한 인지적 메커니즘과 어떠한 관계에 놓여있는지이고, 심적 상태의 정체성이 이와 같은 인과적 관계들을 통해 정해진다는 점에서 MM이론은 기능주의적이다. 이때 인지적 메커니즘들은 어떤 경우에는 상식 심리학

12) Piccutio and Carruthers, *Inner Sense* (2014)

이 제공하는 풍부한 정보들을 활용하기도 하지만, 니콜스와 스티치는 내성을 담당하는 메커니즘은 상식 심리학적 정보를 활용하지 않는다고 보고 있다. 그런 점에서 이들의 내적 감각 이론은 심리적 기능주의를 따르지만, 역할 기능주의와 실현자 기능주의 중 어느 쪽을 분명하게 지지하고 있지는 않다.[13]

2. 3. 4. 골드만

골드만의 내적 감각 이론은 여러 내적 감각 이론들 중 가장 최근에 등장하였으며 또한 가장 발전된 형태라는 평가를 받는다.[14] 그의 이론에 따르면 내성은 심적 상태가 표상하는 내용을 메타-표상하는 과정으로, 이를 수행하는 내적 감각은 심적 상태를 인식하고, 그 내용을 재배치하며, 그리고 때로는 심적 상태의 내용을 한가지 형식으로부터 다른 형식으로 변환한다.

먼저 골드만은 라이칸과 마찬가지로 내적 감각의 출력값이 표상, 즉 심적 상태에 관한 메타-표상이어야 한다고 주장한다. 이때 심적 상태에 관한 메타-표상이 포함해야 할 핵심적인 내용으로 골드만은 세 가지를 꼽는데, 이는 바로 일차적 심적 상태의 내용와 종류(Type), 그리고 강도(Intensity)다. 흥미롭게도 골드만이 제시한 내적 감각 이론은 라이칸의 이론과 니콜스와 스티치의 이론 양쪽을 일부 수용하고 있다. 우선 일차적 심적 상태의 내용을 내적 감각이 재표상하는 방식과 관련해서는 골드만은 니콜스와 스티치가 주장한 것처럼 내용의 재배치가 이루어진다고 본다. 니콜스와 스티치의 이론에서 내용의 단순한 재배치는 표상적 내용을 갖는 심적 상태들에서는 문제없이 이루어졌으나, 현상적 내용을 갖는 지각적 상태가 내성을 통해 표상적 내용을 갖는 자기 믿음으로 이어지기 위해서는 일종의 변환 메커니즘이 추가적으로 필요했다. 골드만도 유사한 관점에서 내적 감각이 심적 상태의 내용을 변환하는 작업을 수행할 필요가 있다고 진단한다. 지각적 상태에 관한 내성 외에도 내성의 대상이 되는 일차적 심적 상태의 내용이 이차적인 자기 믿음의 내용과 종

13) "Positing a 'box' which represents a functionally characterized processing mechanism or a functionally characterized set of mental states does not commit a theorist to the claim that the mechanism or the states are spatially localized in the brain." Nichols and Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds* (2003)

14) Piccutio and Carruthers, *Inner Sense* (2014)

류가 다른 모든 경우에서 변환은 필요하다. 다른 한편으로 그는 라이칸이 주장한 바와 같이 내적 감각이 단순히 일차적인 심적 상태의 내용을 그대로 재활용하지 않고, 추가적으로 특정한 방식으로 표상한다는 점에 동의한다. 여기서 골드만은 이 추가적인 특정한 메타-표상 방식에는 대표적으로 심적 상태의 종류와 강도가 포함되어야 한다고 본다. 예컨대 일차적인 심적 상태인 믿음이 있고, 그 내용이 P라고 하자. 이때 내적 감각을 통해 해당 믿음을 내성한 결과로 생기는 자기-믿음은 '나는 P라고 믿는다'를 내용으로 갖는데, 골드만은 나아가 '나는 P라고 강하게/약하게 믿는다'라는 믿음 또한 내적 감각이 산출한 표상에 의해 설명될 수 있어야 한다고 보는 것이다. 골드만은 내용적 변환이 정확히 어떻게 이루어지는지, 그리고 어떻게 내적 감각이 일차적 심적 상태가 갖는 속성들을 재배치 및 변환된 출력값과 적절히 묶는지(Bind)에 대해서는 아직 구체적인 설명을 제공하지 못한다는 점을 인정하며, 다만 변환의 과정에서 반드시 일차적 내용이 온전하게 변환될 필요는 없을 것이라고 말한다. 예컨대 일차적인 지각적 상태는 이차적인 자기 믿음보다 훨씬 풍부한 내용을 포함하고 있는데, 변환 과정에서 이 풍부함은 손실될 수밖에 없다는 것이다.[15]

골드만의 이론이 앞서 살펴본 다른 내적 감각 이론들과 가장 차이점을 드러내는 지점은 바로 내적 감각이 어떤 속성을 통해서 내성을 진행하는지

---

15) 니콜스와 스티치나 골드만이 주장한 것처럼 지각적 상태의 변환 과정에서 내용의 손실이 반드시 필수적인가에 관해서는 의문이 들 수 있다. 내가 콜라병을 바라보고 있을 때 갖는 시각적 상태의 내용에 비해 '나는 콜라병을 보고 있다'는 이차적인 자기 믿음의 내용은 일견 빈약해 보인다. 그러나 자세히 들여다보면, 이는 내성 자체의 한계 때문이라기보다는 콜라병이라는 개념적 표상을 사용해 믿음을 형성했기 때문으로 보인다. 당연하게도 내가 당장 콜라병을 바라볼 때의 현상성에는 콜라병의 구체적인 색감이나 병 표면에 맺힌 물방울의 형태가 주는 느낌 등 콜라병의 개념적 표상에는 포함되지 않는 것들이 많이 있을 것이다. 아무리 콜라병의 시각적 특징들에 대한 수식을 추가해 표상적 내용을 더하더라도 이토록 풍부한 내용을 손실 없이 자기 믿음으로 가져오기에는 부족하다.
이와 관련해 거틀러는 지시적 주목(Demonstrative attention)을 사용해 현상적 상태들에 관해서 내용의 손실 없이 내성할 수 있다고 본다. 내가 콜라병을 보고 있을 때 갖는 현상성에 주목하여 '나는 바로 이러한 현상적 상태에 놓여있다'는 믿음을 형성한다면 '나는 콜라병을 보고 있다'의 경우와 달리 내용의 손실이 발생하지 않는다. 이때 지시적 주목이 성립하기 위해 요구되는 것은 현상적 상태와 내성하는 상태가 동시에 존재하는 것이다. (거틀러는 두 심적 상태가 동시에 존재할 뿐만 아니라, 내성하는 상태가 현상적 상태를 내포해야 한다고 주장하였으나, 내적 감각 이론의 입장에서는 내성되는 상태와 내성하는 상태가 별개의 심적 상태라고 보는 만큼 두 상태의 동시 존재만이 필요 조건이며 내포 관계는 불필요할 것으로 보인다.) 비록 거틀러는 내적 감각 이론이 아니라 대면 이론을 지지하지만, 거틀러가 제시한 지시적 주목 자체는 내적 감각 이론에서도 충분히 활용할 수 있는 이론적 자원이라고 할 수 있다. Gertler, *Introspecting Phenomenal States* (2001)

와 관련해서이다. 골드만은 심적 상태가 가질 수 있는 총 네 가지 종류의 속성들을 후보로 놓는다. 이는 기능적 속성, 현상적 속성, 표상적 속성, 그리고 신경적 속성이다. 앞서 등장한 내적 감각 이론가들은 모두 심적 상태에 관한 기능주의를 어떤 형태로든 받아들이고 있었고, 그렇기에 이들이 제시한 내적 감각은 기능적 속성을 통해 심적 상태를 감지했다. 그러나 골드만은 이들과 다른 노선을 택하며, 오히려 기능적 속성을 통해서 심적 상태를 내성할 수 없다는 입장을 전개한다.

골드만이 처음 내적 감각의 내성이 활용하는 속성의 후보로 놓았던 것은 현상적 속성이다. 기존 철학적 논의에서 지각적인 상태들은 현상적 속성을 갖는다고 흔히 인정되지만, 그 외 다른 모든 심적 상태들이 현상적 속성을 갖는지는 직관적으로 확실하지 않다는 것이 중론이었다. 현상적 속성을 갖지 않는 심적 상태들이 존재한다면, 현상적 속성에 기반한 내성 이론은 모든 심적 상태들에 대한 설명을 제시할 수 없다. 그러나 골드만은 모든 심적 상태들이 현상적 속성을 가진다고 볼 여지가 있다고 말하며 이를 지지하는 두 가지 논거를 제시한다. 첫째로 그는 지각적이지 않은 심적 상태가 현상성을 가진다는 것을 한 가지 사례를 통해 보이고자 한다. 무언가를 말하려고 하지만 단어가 떠오르지 않을 때, 우리는 말하고자 하는 바가 있음을 분명 어떤 방식으로 느끼지만 아직 그 내용은 음운론적으로 구성되지 않았고 따라서 감각적 현상성은 갖지 못한 상태이다. 골드만은 해당 사례가 감각적 현상성 외에도 사유의 개념적 부분이 갖는 현상성이 존재한다는 것을 보여준다고 주장한다.

둘째로 그는 현상성에 관한 잘 알려진 사고 실험인 '과학자 메리' 사례를 변형시켜 모든 심적 상태가 현상성을 가짐을 보여주고자 한다. '과학자 메리' 사례를 간략히 소개하자면 다음과 같다. 과학자인 메리는 평생 무채색의 방에서 나와본 적이 없고 실제로 색을 경험한 적이 없다. 대신 그녀는 직접 경험 외의 방식들로 색과 관련된 모든 물리적 및 기능적 지식을 습득했다. 그럼에도 그녀가 처음 방을 나와서 색을 경험할 때, 그녀는 색을 경험한다는 것이 질적으로 어떤 느낌인지를 새롭게 배우게 된다. 해당 사례는 지각적 경험이 다른 속성들로 환원되지 않는 현상적 속성을 갖는다는 것을 보여주고자 구상된 것이었는데, 골드만은 이 사례를 변형시켜 모든 심적 상태가 현상적 속성을 갖는다는 것을 보여줄 수 있다고 주장한다. 골드만의 변형된 사례는 색

을 경험한 적 없는 사람 대신 특정한 심적 상태를 경험한 적 없는 사람을 상정한다. 해당 인물이 처음으로 특정 심적 상태를 가질 때 그는 그 심적 상태를 가지는 것이 어떤 느낌인지를 새로이 알게 될 것이라는 직관에 동의한다면, 모든 심적 상태가 현상적 속성을 갖는다는 주장에도 동의할 수 있을 것이다.[16] [17]

이후 골드만은 자신의 이론적 입장을 수정하는데, 그는 네 가지 후보 속성 중에서 신경적 속성을 제외한 나머지 세 가지 속성이 어째서 내적 감각이 내성에 활용하는 속성이 될 수 없는지를 각각 제시한 후, 마지막 후보인 신경적 속성을 통해서 내적 감각이 심적 상태를 내성한다고 주장한다. 내적 감각이 신경적 속성을 통해 내성한다는 주장은 몇 가지 장점을 지닌다. 먼저 심적 상태에 관한 물리주의를 지지하는 입장에서는 내적 감각이 물리적 속성의 일종인 신경적 속성을 통해 내성한다고 상정할 때 내성이 진행되는 과정에서 심적 상태가 어떻게 자기 지식의 형성에 인과적인 영향력을 행사할 수 있는지가 가장 자연스럽게 설명된다고 할 수 있다. 신경적 속성은 물리적 속성인 만큼 물리주의를 오랫동안 괴롭혀 온 정신의 인과적 효력에 관한 난제를 벗어날 수 있다. 또한 신경적 속성에는 어떤 신경 세포군 혹은 신경 회로가 활성화되었는지 및 어느 정도로 활성화되었는지도 포함될 수 있는데, 이는 내성의 대상이 된 심적 상태가 어떤 종류의 심적 상태인지에 관한 정보와 심적 상태의 강도에 관한 정보를 내적 감각 메커니즘에 제공하는 방법이 될 수 있다.

그렇다면 골드만은 어째서 다른 세 가지 속성들을 거부하는가? 먼저 현상적 속성의 경우를 보자. 그는 첫째로 현상적 속성은 그 자체로는 인과력이 없고, 그것을 수반하는 물리적 기반만이 인과력을 가지기 때문에 인과적 메커니즘으로서 상정된 내적 감각이 현상적 속성을 통해서 내성을 할 수 없다고 말한다. 더 심각한 문제는 그 다음인데, 골드만은 앞서 본인의 이전 저작에

16) Goldman, *The Psychology of Folk Psychology* (1993)

17) 골드만이 제시한 변형 사례가 골드만이 생각하는 것처럼 원래 사례와 평행하지 않을 수 있다. 사례의 주인공은 특정 심적 상태를 한 번도 직접 경험한 적 없으면서도 그 심적 상태와 관련한 모든 객관적 지식을 갖고 있어야 할 텐데, 색을 경험한 적 없이 색과 관련된 객관적 지식을 갖는 것에 비해서 이는 훨씬 불가능해 보인다. 예컨대 명제적 믿음을 갖는 심적 상태를 경험한 적 없이 명제적 믿음을 갖는 심적 상태에 관한 모든 객관적 지식을 획득하는 것은 불가능할 수 있다.

서 주장한 바를 뒤집고 과연 모든 종류의 심적 상태들이 현상적 속성을 가지는지부터가 의심스럽다고 말하며, 설령 모든 심적 상태들이 현상적 속성을 가진다고 하더라도 서로 다른 종류의 심적 상태가 갖는 현상적 속성이 그것들의 종류를 구분할 수 있는 기반이 되어줄 만큼 서로 구별된다고 할 수 없다고 지적한다. 따라서 현상적 속성은 내적 감각이 활용하는 속성의 후보로서 제외된다.

다음으로 표상적 속성을 보자. 대부분의 철학적 논의에서 심적 상태들은 표상적 속성을 갖는 것으로서 인정받는 만큼, 골드만은 표상적 속성이 현상적 속성에 비해서는 나은 후보라고 인정하고 있다. 그러나 앞서 살펴본 다른 내적 감각 이론들에서 지적된 것처럼, 골드만은 현상적 상태들이 표상적 속성으로 온전히 포착될 수 있는지에 대해 회의적이다. 따라서 현상적 상태들에 관한 내성은 표상적 속성을 활용하는 내적 감각으로는 설명될 수 없다. 나아가 골드만은 현상적 속성이 표상적 속성으로 환원될 수 있다는 표상주의자들의 주장을 받아들이더라도, 표상적 속성은 여전히 심적 상태의 종류와 강도를 포착하기에 적합하지 않다고 본다. 예컨대 내가 '대한민국의 수도는 서울이다'라고 믿을 때, 내가 해당 믿음을 얼마나 강하게 갖고 있는지는 표상적인 속성을 통해서는 포착될 수 없다는 것이다. 그러나 내적 감각이 활용하는 속성은 심적 상태의 종류와 강도에 관한 메타-표상의 산출을 설명할 수 있어야 하기에 표상적 속성 또한 후보로서 거부된다.[18]

마지막은 기능적 속성이다. 골드만은 현상적 속성이나 표상적 속성에 가한 비판보다 훨씬 강하고 다양한 비판을 기능적 속성에 대해서 제기하는데, 그의 비판적 쟁점은 크게 세 가지로 정리해볼 수 있다. 앞서 기능주의를 소개하며 살펴본 것처럼, 기능적 속성은 경향적인 인과적 관계의 연언이다. 골드만은 이에 관해 기능적 속성이 경향적 속성이라는 점과 관계적 속성이라는 점 때문에 기능적 속성은 내적 감각의 내성 작용에 활용되기에 부적절해진다고 비판한다. 또한 그는 기능적 속성의 인과적 효력에 관해서도 의심이 제기될 수 있다고 본다. 이에 관해서는 다음 장에서 더 자세히 다루어보도록 하겠다.

우선 골드만의 내적 감각 이론은 모든 심적 상태의 내성에 대한 설명

18) Goldman, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (2006)

을 제공하고자 한다는 점에서 통일적이고 포괄적인 자기 지식 이론으로서의 장점은 지니고 있다. 또한 현상적 상태에 관한 내성을 설명하는 데도 기능주의적 이론들에 비해 유리한 입지에 있다. 현상적 상태가 신경적이거나 물리적인 기반을 갖지 않는다는 주장은 현상적 상태가 기능적으로 설명될 수 없다는 주장보다 더욱 강한 주장이기 때문이다. 예컨대 '전도된 스펙트럼' 사례에 대한 라이칸의 답변을 받아들이지 않고, 여전히 전도된 스펙트럼을 가진 사람의 심적 상태가 보통 사람의 심적 상태와 기능적으로는 동일할 수 있다고 보는 학자도 전도된 스펙트럼을 가진 사람의 심적 상태의 신경적 기반이 보통 사람의 심적 상태의 신경적 기반과 동일하다고까지는 주장하고 싶지 않을 수 있다. 한편 골드만은 내적 감각이 신경적 속성을 통해 내성한다고 주장함으로써 환원적 물리주의를 전제해야만 하는 것으로 보인다.[19] 심적 상태가 신경적 속성을 가지기 위해서는 심적 상태가 물리적 상태와 동일하다고 보는 것이 가장 자연스럽다. 그러나 이런 식으로 환원적 물리주의를 전제하는 것은 앞서 등장했던 내적 감각 이론들과 비교하면 형이상학적 개입이 강한 편이라 할 수 있고, 내적 감각 이론의 장점 중 하나였던 물리주의에 대한 개방성을 잃어버리는 것이 된다. 앞서도 설명한 것처럼 내적 감각 이론은 물리주의와 친화적인 이론으로 흔히 받아들여지곤 하지만, 엄밀히 말해 내적 감각 이론 자체는 굳이 물리주의를 전제할 필요가 없다. 내적 감각 이론의 핵심은 내성이 지각과 유사하다는 것이고, 보다 자세하게는 내적 감각이 심적 상태와 인과적으로 상호작용함으로써 일차적 심적 상태들에 관한 메타-표상을 산출한다는 것이다. 이와 같은 주장은 물리주의를 받아들이지 않더라도 충분히 성립할 수 있다. 그런 점에서 골드만의 내적 감각 이론은 앞서 소개된 이론들에 비해 물리주의에 깊게 개입해야 한다는 부담을 갖고 있다.

---

19) 환원적 물리주의를 전제하지 않고서 내적 감각이 심적 상태의 신경적 속성을 통해 내성을 진행한다고 보는 것은 매우 부자연스럽다. 우선 심적 상태가 물리적 상태와 동일하지 않으면서 신경적 속성을 가진다고 보는 것부터가 이상하다. 심적 상태의 신경적 속성이라는 표현을 최대한 호의적으로 해석하여 동일성을 전제하지 않고 심신 수반 관계만 전제한 상태에서 심적 상태를 수반하는 물리적 상태의 신경적 속성이라고 해당 표현을 이해할 가능성이 없지는 않다. 그러나 이런 경우 신경적 속성을 통한 내성은 사실상 심적 상태 자체에 대한 내성이 아니라 심적 상태를 수반하는 물리적 상태를 파악하는 작용에 불과해진다. 따라서 골드만과 같이 내적 감각이 심적 상태의 신경적 속성을 통해 내성을 진행한다고 주장하기 위해서는 환원적 물리주의가 전제되어야 한다고 보아야 한다.

3. 기능주의적 내적 감각 이론 옹호

3. 1. 골드만의 내적 감각 이론의 문제점

골드만의 이론은 가장 발전된 형태의 내적 감각 이론이라는 평가를 받고 있다고 했지만, 이러한 평가를 받은 것은 더 이전의 내적 감각 이론들이 주로 내성과 외부 지각의 유사성에 주목하고 구체적으로 다양한 심적 상태의 내용에 대한 내성이 어떻게 이루어지는지를 충분히 설명하지 않았던 것에 비해 골드만의 이론은 심적 상태들의 내용과 종류, 그리고 강도에 대한 내성을 내용의 재배치와 변환 등의 과정을 통해 이루어지는 것으로서 더 구체적으로 설명했기 때문이지, 골드만이 기능적 속성을 거부하고 신경적 속성을 내적 감각의 감지 대상으로 두었기 때문은 아니라고 할 수 있다. 오히려 신경적 속성을 채택함으로써 골드만의 이론은 더 발전된 내적 감각 이론임에도 불구하고 이전 이론들이 겪지 않았던 문제점을 겪게 된다. 우선 환원적 물리주의를 전제해야만 한다는 점에서 골드만의 내적 감각 이론은 이전 이론들에 비해 형이상학적 입장의 선택지가 좁다. 심신 문제에 관해 이원론이나 창발론 등의 입장을 갖고 있는 학자들은 골드만의 내적 감각 이론을 자기 지식이나 내성을 설명하는 이론으로서 수용할 수 없게 되는 것이다.

또한 환원적 물리주의를 받아들임으로써 골드만의 내적 감각 이론은 환원적 물리주의가 갖고 있는 문제점 중 하나인 심적 상태에 대한 배외주의(Chauvinism)를 떠안게 된다. 예컨대 사람들에게서 고통이라는 심적 상태는 공통적으로 C-섬유의 활성화라는 신경적 상태와 동일하다고 하자. 그렇다면 골드만의 이론에서 사람들이 고통을 내성하는 것은 C-섬유의 활성화라는 신경적 상태가 가진 신경적 속성을 내적 감각이 감지함으로써 이루어지는 것일 터이다. 그런데 사람이 아닌 다른 생명체들도 고통이라는 심적 상태를 가질 수 있는 것처럼 보이지만, 이들은 C-섬유를 갖고 있지 않아서 C-섬유의 활성화와 같은 신경적 상태를 갖지 못할 수도 있다. 심지어 특수하게는 사람이면서도 모종의 이유로 인해 C-섬유의 활성화라는 신경적 상태를 결여하는 경우도 상상해 볼 수 있겠다. 하지만 동일성에 대한 이행성 규칙을 받아들인다면, 심적 상태가 신경적 상태와 동일하면서 서로 다른 신경적 상태들이 같은 심적

상태일 수는 없을 것이다. 그렇기에 골드만의 내적 감각 이론은 C-섬유의 활성화를 갖지 못한 개체는 고통을 내성할 수 없게 된다는 배외주의적 결론을 도출하게 된다. 하지만 이러한 배외주의의 문제를 받아들이면서라도 환원적 물리주의를 받아들이고자 하는 입장에서는 골드만의 내적 감각 이론이 환원적 물리주의에 개입함으로써 형이상학적 입장의 선택지가 줄어들게 만든다는 것이 크게 문제로 느껴지지 않을 수 있다. 그렇다면 보다 심각한 문제는 어디에 있는가?

내적 감각 메커니즘이 신경적 속성을 통해 내성한다는 것은 사실 물리주의를 가정한다면 굳이 다른 논거를 들지 않더라도 자연스럽게 따라오는 결론일 수 있다. 그러나 현상적, 표상적, 기능적 속성들이 신경적 속성으로 환원된다고 하더라도 여전히 이러한 속성들을 통해서 심적 상태들과 내성에 관한 설명을 제공하는 것이 더욱 적절할 수 있다. 다른 모든 종류의 속성들을 통한 설명을 거부하고 오직 신경적 속성을 통해서 내적 감각을 설명하려고 한다는 측면에서 골드만의 후기 이론은 내성과 심적 상태들에 관한 기존 철학적 논의와 단절되어 버린다. 내적 감각이 정말 순수하게 신경적인 속성만을 감지하는 메커니즘이라면, 일상적 언어와 개념 수준에서 이루어지는 철학적 분석은 이 이상 내적 감각을 통한 내성에 대해 할 말이 없고, 후속 과제로 남는 것은 신경적 속성들에 관한 자연과학적인 탐구뿐이다. 유사한 맥락에서 피치우토와 카러더스는 여러 내적 감각 이론들을 정리해 평가하는 글에서 다음과 같은 우려를 표한 바 있다.

> … 각 인지적 체계의 입력값에 관한 기술이 적절하게 이루어지는 층위에 따라서 인지적 체계들 사이에 구분을 지을 수 있을지도 모른다. 한편으로 모든 심적 메커니즘들은 어떤 기술적 층위에서는 신경적 속성에 대해 민감해야 하지만, 많은 경우 심적 메커니즘이 오히려 내용적 혹은 계산주의적-구문론적 속성들에 대해 민감하다고 설명하는 것이 더욱 적절할 수 있다.[20]
>
> (Picciuto and Carruthers, *Inner Sense*, p.15)

20) 해당 부분은 원래는 피치우토와 카러더스의 논문에서 라이칸의 내적 감각 이론에 대한 비판으로서 작성되었고, 동일 논문 후반부에 골드만의 내적 감각 이론에 대해서도 동일하게 적용될 수 있다고 언급된 것이다. 그러나 이러한 비판은 골드만의 이론에 대해 더 강하게 적용될 수 있으며, 라이칸의 이론은 상대적으로 이러한 비판에 의해 덜 타격을 입는다는 것을 본 논문의 후반부에 보일 것이다.

설령 물리주의를 받아들여 심적 상태가 신경적 상태와 동일하다고 보더라도, 내적 감각이 기능적 속성으로는 포착되지 않는 식으로 신경적 속성을 감지한다고 볼 필요는 없을 것이다. 오히려 골드만 자신 또한 신경적 속성에 기반한 내적 감각 이론을 제시하기 이전에, 현상적 속성에 기반한 내성 이론을 고려하던 시기에는 유사한 비판을 제기한 바 있다.

> 그러나 문제는 이러한 신경적 사건들에 기입되어 있는 내용들이 신경적 속성들에 관한 것이냐이다. 다시 말하지만 이는 받아들이기 어렵다. 신경적 사건들은 시각적 정보를 처리하지만, 인지과학자들은 이러한 신경적 사건들에 신경적 내용을 부여하지는 않는다. … 물론 정보 처리는 상당 부분 유기체 내에서 인간-하위적 수준에서 발생한다. 그러나 정보 처리가 순수하게 인간-하위적이기만 하다면, "정신적"이라고 인식할 수 있는 언어적 표지는 생성되지 않는 듯하다. … 따라서 비록 "포도당 과다"와 같은 속성들이 인간-하위적 수준에서 감지되긴 하지만, 이러한 속성들은 정신적이고 언어적 표지를 갖는 체계가 이용하는 속성들이라고는 볼 수 없다.
>
> (Goldman, *The Psychology of Folk Psychology*, p.20)

그렇다면 물리주의적 관점에서 고려해볼 수 있는 다른 접근법은 앞서 배제되었던 다른 후보 속성들이 신경적 속성으로 환원된다는 것을 인정하면서도 여전히 어떤 설명적 층위에서는 유의미하게 내적 감각의 인과적 메커니즘이 활용하는 속성으로서 인정될 수 있다고 생각하는 것이다. 이런 접근법을 취하고 나면 곧바로 골드만이 잠정적으로 받아들이고 있던 한 가지 가정이 드러나게 된다. 바로 후보로 소개된 여러 가지 종류의 속성 중 한 가지만이 내적 감각에 의해 활용되어야 한다는 가정이다. 그가 현상적 속성이 내적 감각이 활용하는 속성이 될 수 없다고 주장했던 이유는 모든 심적 상태가 현상적 속성을 가지는 것은 아니기 때문이었으며, 마찬가지로 표상적 속성이 내적 감각이 활용하는 속성이 될 수 없다고 주장했던 이유도 현상적 상태들이 표상적 속성을 가지지 않기 때문이었다. 그는 모든 심적 상태들이 공통적으로 갖는 한 가지 종류의 속성만이 내적 감각에 의해 활용될 수 있다고 가정하였기에, 물리주의를 받아들일 경우 모든 심적 상태들이 가질 수밖에 없는 속성인 신경

적 속성이 활용되어야 한다고 주장한 것이다. 그러나 다양한 심적 속성들이 물리적 층위에서는 신경적 속성과 동일하다고 한다면, 앞서 거부된 여러 속성들이 모두 내적 감각에 의해 활용되지 못할 이유가 없다. 오히려 현상적 · 표상적 · 기능적 속성들과는 완전히 구분되는 다른 어떠한 신경적 속성만을 통해 내적 감각이 내성을 진행한다면, 그러한 메커니즘의 작용이 대체 어떤 의미에서 내성이라고 할 수 있는지가 불분명하다. 신경적 속성을 감지할 수 있는 메커니즘에는 위에 음식물이 들어온 것을 감지하면 위액을 분비하게 하는 메커니즘 등, 정신적인 영역과 전혀 관계 없는 것들도 포함될 수 있는데, 내적 감각이 신경적 속성만을 감지한다고 한다면 이런 비(非)정신적 메커니즘들과 어떻게 구분될 수 있겠는가? 내적 감각 이론이 내성에 대한 적절한 설명이 될 수 있으려면 내적 감각이 그저 신경적 연결망으로 이루어진 메커니즘으로만 이해되어서는 곤란하고, 오히려 심적 상태들의 현상적, 표상적, 기능적 속성과 인과적으로 상호작용하는 메커니즘으로 이해되어야 마땅하다.

나아가 내적 감각 이론이 물리주의와 독립적으로 성립할 수 있기 위해서도 내적 감각이 신경적 속성이 아닌 다른 속성을 통해 내성한다고 상정될 필요가 있다. 내적 감각과 심적 상태의 인과적 상호작용이 어떤 식으로 이루어지는지에 대한 이해 방식은 심신 문제에 관하여 수반론을 받아들이는지, 또는 환원적 물리주의와 비환원적 물리주의, 그리고 이원론 중 어느 쪽을 지지하는지에 따라 달라질 수 있는 여지가 있다. 그러나 내적 감각이 내성에 활용하는 속성을 신경적 속성으로 고정해두고 나면 내적 감각 이론은 물리주의로부터 더 이상 자유롭지 못하다. 불필요한 형이상학적 개입을 최소화한다는 측면에서도 신경적 속성이 아닌 다른 속성으로 눈길을 돌릴 필요가 있다.

이제 자연스럽게 다시 논의 대상으로 떠오르는 후보는 기능적 속성일 것이다. 골드만이 기능적 속성을 가장 강하게 비판한 것과는 반대로, 골드만 이전에 내적 감각 이론을 발전시켰던 학자들은 모두 기능주의자였고, 기능주의 자체는 현재까지도 철학 및 인지과학에서 심적 상태에 관한 주요 이론으로 자리매김하고 있다. 그리고 기능주의를 받아들인다면, 골드만이 고려했던 다른 후보들인 표상적 속성 및 현상적 속성은 기능적 속성을 통해서 충분히 설명될 가능성도 있다. 앞서 살펴본 것처럼, 라이칸의 심리적 기능주의는 현상적 속성을 기능적 속성으로 설명하고 있다. 기능주의적인 내적 감각 이론은 심적 상

태의 내용과 종류, 그리고 강도에 대한 표상을 설명하는 측면에서도 분명한 강점을 갖는다. 심적 상태들의 서로 다른 내용과 종류, 그리고 강도는 당연하게도 다른 기능을 가질 것이기 때문이다. 또한 기능주의는 물리주의를 전제하지 않기에 내적 감각이 활용하는 속성이 기능적 속성이라는 주장은 앞서 언급한 것처럼 내적 감각 이론이 심신 문제에 관해 더 다양한 형이상학적 입장을 포용할 수 있게 만들어 줄 수 있다. 이상과 같은 동기들은 다시 한번 기능주의적 내적 감각 이론을 검토할만한 충분한 동기를 제공하는 듯하다. 그렇다면 이제 내적 감각이 기능적 속성을 통해 심적 상태를 감지하는 것은 적절하지 못하다는 비판들을 자세히 살펴보고, 이에 대해 설득력 있게 답변을 제시할 수 있는지를 확인해보도록 하겠다.

3. 2. 기능주의의 종류와 내적 감각 이론의 관계

앞서 기능주의와 그 종류들에 대해 간략하게나마 알아본 이유는 내적 감각 이론에서 내적 감각이 기능적 속성을 입력값으로 갖는다고 할 때, 기능적 속성이 의미하는 바가 어떤 종류의 기능주의를 전제하는지에 따라서 달라질 수 있기 때문이다. 그런 의미에서 다시금 각 종류의 기능주의를 받아들인다는 것이 내적 감각 이론에 어떤 영향을 미치는지를 정리해보고 넘어가자.

분석적 기능주의를 전제한다면 심적 상태의 기능적 속성이란 심적 상태들에 관한 일상적 지식에 대한 분석을 통해 알려지는 속성일 것이고, 심리적 기능주의를 전제한다면 심적 상태의 기능적 속성은 최선의 과학 이론들을 통해 경험적으로 알려지는 속성일 것이다. 암스트롱은 분석적 기능주의를 받아들였던 만큼, 그의 내적 감각 이론에서 내적 감각이 감지하는 기능적 속성이란 심적 상태들의 일상적 지식을 분석해 얻어지는 인과 관계들을 연언으로 연결함으로써 제시될 수 있다. 예컨대 임의의 심적 상태 x가 고통이라는 심적 상태일 때 갖는 기능적 속성 F는 다음과 같을 수 있다: F = x는 신체적 손상에 의해 야기되려는 경향이 있음 ∧ x는 움찔거림을 야기하려는 경향이 있음 ∧ x는 x로부터 벗어나려는 심적 상태 y를 야기하려는 경향이 있음 ∧ …. 한편, 심리적 기능주의를 받아들이는 라이칸의 내적 감각 이론에서 내적 감각이 감지하는 기능적 속성에는 과학적 탐구를 통해서만 알려지는 인과 관계들도

포함될 수 있다. 예를 들어 라이칸에게는 앞서 언급한 기능적 속성 F에 다음과 같은 내용도 연언으로 포함될 수 있는 것이다: x는 히스타민 분비에 의해 야기되려는 경향이 있음 ∧ x는 혈당 수치가 70 mg/dL 이하로 떨어짐에 의해 야기되려는 경향이 있음 ∧ …. 라이칸의 기능주의는 무수히 많은 기능적 단계들이 연속적인 위계를 형성하고 있다고 보는 만큼, 그의 기능주의에서 심적 상태의 기능적 속성에는 일상적인 지식의 수준의 인과 관계들부터 신경화학적 수준에서 발생하는 인과 관계들까지 모두 포함될 수 있다.

또한 역할 기능주의를 전제한다면 내적 감각이 기능적 속성을 감지함으로써 내성의 대상으로 삼는 심적 상태란 물리적 자극, 행동, 그리고 다른 심적 상태들과 갖는 관계적인 인과적 특징들을 연언으로 연결한 것으로서의 고차 속성을 갖는 상태이자 오직 그러한 상태일 것이고, 실현자 기능주의를 전제한다면 심적 상태란 고차 속성을 갖는 상태 자체라기보단 기능적 조건을 만족시키는 물리적 상태일 것이다. 한 가지 주의할 것은 역할 기능주의와 달리 실현자 기능주의는 내적 감각이 내성에 활용하는 속성이 기능적 속성이 아니라 물리적 속성, 혹은 신경적 속성이라는 주장과 양립 가능하다는 점이다. 기능적 속성은 단지 심적 상태와 동일한 물리적 상태인 실현자를 특정하는 데에만 동원되고, 심적 상태가 실제로 인과적 영향력을 행사하는 것은 실현자가 같은 물리적 속성인 신경적 속성을 통해서 이루어진다고 볼 수 있기 때문이다. 그렇기에 실현자 기능주의는 기능적 속성이 내적 감각에 의해 감지되는 속성으로서 부적절하다는 비판으로부터 상대적으로 자유로울 수 있는 여지가 있다. 다만, 암스트롱의 분석적 실현자 기능주의를 소개할 때 논의되었던 것처럼 실현자 기능주의를 채택하더라도 물리적 속성과 구분되는 기능적 속성에 주목해야 할 필요성은 존재할 수 있다. 내성이 “심적 상태와 동일한 물리적 상태”를 감지함으로써 이루어지는 것이 아니라, “심적 상태로서의 심적 상태”에 대해 이루어진다고 주장하려면 내적 감각은 심적 상태의 기능적 속성을 감지한다고 보는 것이 타당하다. 한편 역할 기능주의를 받아들이는 내적 감각 이론의 경우에는 내적 감각이 기능적 속성을 감지함으로써 내성을 진행한다는 주장을 보다 적극적으로 옹호해야만 할 이유가 분명하다. 이와 같은 기능주의적 입장들의 구분은 기능적 속성이 내적 감각의 입력값이 될 수 없다는 비판에 대해 답변할 때 염두해야 할 것이다.

3. 3. 기능적 속성에 대한 비판과 답변

이제 기능적 속성이 내적 감각의 입력값이 될 수 없다는 비판들을 살펴보고, 이에 대한 가능한 답변들은 어떤 것들이 있는지 살펴보도록 하겠다. 내적 감각 이론의 입장에서 기능적 속성을 강하게 비판했던 골드만의 주장들을 중심적으로 다룰 것이고, 다른 학자들이 기능주의와 관련해 제기한 문제들도 알아볼 것이다. 또한 그 외에 기능주의적 내적 감각 이론에 대해 제기될 수 있을 법한 비판들도 예상해 답변해보겠다.

심적 상태의 기능적 속성이란 심적 상태를 기능적으로 정의해주는 관계적이고 경향적인 속성들의 연언이다. 골드만은 기능적 속성의 이러한 성질로 인해 기능적 속성이 내적 감각에 의해 감지되는 속성으로서 부적절해진다고 본다. 이때 주목할만한 점 한 가지는 골드만이 자신이 주로 비판의 대상으로 삼고자 하는 것은 분석적 기능주의이며, 심리적 기능주의에 대해서는 특별히 할 말이 없다고 언급한다는 점이다.[21] 그러나 실제로 그가 제기하는 비판들을 살펴보면 그의 비판들은 기능적인 속성 자체의 성질에 대해 제기되는 것들이다. 그리고 이러한 성질은 분석적 기능주의에 의해 정의된 기능적 속성과 심리적 기능주의에 의해 정의된 기능적 속성이 공통적으로 갖는 것들이다. 따라서 골드만의 비판에 대응해야 하는 것은 비단 분석적 기능주의를 받아들이는 내적 감각 이론만이 아니며, 심리적 기능주의를 받아들이는 내적 감각 이론 또한 같은 부담을 지고 있다. 그러나 분석적 기능주의와 심리적 기능주의의 대응 방식은 차이를 보일 수 있으며, 각각의 대응이 얼마나 성공적인지에 따라 내적 감각 이론이 어떤 종류의 기능주의를 더 선호할만한지가 밝혀질 수 있을 것이다. 이제 구체적으로 기능적 속성의 어떤 성질이 어떠한 문제를 일으킨다고 비판을 받는지 검토한 후, 이에 대한 반박을 제시할 수 있는지 확인해보도록 하겠다.

---

21) Goldman, *The Psychology of Folk Psychology* (1993)

3. 3. 1. 기능적 속성은 관계적 속성이라는 점에 대하여

골드만이 가장 먼저 문제 삼는 부분은 바로 기능적 속성이 관계적 속성이라는 점이다. 즉, 기능적 속성은 심적 상태가 외부 자극과 같은 입력값들, 행동적 출력값들, 그리고 다른 심적 상태들과 어떤 인과적 관계에 놓여있는지에 관한 속성이다. 그러나 내적 감각이 과연 이러한 관계적 속성을 감지할 수 있는가? 관계적 속성을 감지하기 위해서는 우선 관계항들을 감지할 수 있어야 할 것이다. 외부 자극과 행동과 같은 관계항들을 감지하는 것 자체는 그렇게 어려운 문제는 아닐 수 있다. 문제는 하나의 심적 상태가 다시금 다른 심적 상태들을 관계항으로 가진다는 점이다. 기능적으로 정의된 다른 심적 상태들은 다시 저마다 입력값, 출력값, 그리고 또 다른 심적 상태들을 관계항으로 가질 것이고, 결국 하나의 심적 상태의 기능적 속성을 감지하기 위해서 내적 감각이 감지해야 할 관계항의 수는 폭발적으로 증가하게 된다. 해당 문제는 기능주의가 전체론적으로 심적 상태들을 정의하기 때문에 발생한다고 할 수 있다. 골드만이 이러한 감지 대상의 증가가 무한히 늘어나는 악순환으로 이어진다고 말하고자 하는 것은 아니지만, 분명 어떤 심적 상태들은 비교적 많은 수의 다른 심적 상태들과 관계를 맺는 것으로 이해된다. 대표적으로 믿음이나 욕구 등의 심적 상태는 다른 많은 믿음 및 욕구들과 관련된다. 그렇기에 내적 감각이 관계적 속성을 통해 심적 상태들을 감지할 때 관계항의 수가 폭발적으로 증가할 가능성은 충분히 실질적으로 존재하며, 이는 기능적 속성에 기반한 내적 감각 이론에게는 위협이 된다고 골드만은 보고 있다.

내적 감각이 관계적 속성을 입력값으로 갖기 위해서는 관계항 외에도 관계 자체를 감지할 수 있어야 할 것이다. 특히 심적 상태의 기능적 정의에서는 하나의 심적 상태가 입력값, 출력값, 그리고 다른 심적 상태들과 갖는 인과적 관계가 핵심적이다. 그렇다면 내적 감각은 인과적 관계의 존재를 감지할 수 있어야 한다. 그러나 과연 내적 감각이 단순히 관계항들의 존재 여부를 넘어서 어떤 것이 원인이고, 어떤 것이 결과가 되는지를 파악할 수 있는가? 예컨대 고통의 기능적 속성에는 신체적 손상에 의해 야기된다는 것이 포함되는데, 내적 감각이 단순히 신체적 손상이 발생했음을 알아차리는 것을 넘어서 신체적 손상이 고통을 야기하고 있다는 인과성을 파악할 수 있는가? 최소한

인과에 대한 흄적인 관점에서는 이는 불가능할 것이다.

또한 골드만은 즉각적인 내성이 일어날 때는 때로 그 심적 상태의 원인에 관한 정보가 전혀 없이도 내성이 가능한 것처럼 보인다고 지적한다. 그는 아침에 의식이 들자마자 자신이 머리가 아프다는 것을 알게 되는 경우를 예로 들며, 막 일어난 상태라 두통의 원인에 대한 정보가 전혀 없는 상태에서도 여전히 두통에 대한 내성이 가능하다고 말한다. 만약 내적 감각이 기능적 속성을 통해 두통을 내성한다면, 내적 감각이 감지해야 하는 대상에는 두통을 야기한 원인도 포함되어야 할 것이고, 그렇다면 원인을 전혀 알지 못하는 상태에서 기능적 속성을 통해 두통을 내성하는 것은 불가능해진다. 그러나 두통의 실제 사례를 보면 분명 즉각적인 내성이 가능해보이는 만큼, 골드만은 인과적 관계에 대한 정보를 포함하는 기능적 속성이 내적 감각의 입력값이 될 수 없다고 주장한다.[22]

그러나 과연 위에 제기된 비판처럼 내적 감각은 관계적 속성을 입력값으로 가질 수 없는 것일까? 우선 관계항의 폭발적 증가 문제에 대해 답해보자. 첫째로, 관계항의 폭발적 증가는 골드만이 우려하는 것처럼 쉽게 발생하지 않을 수도 있다. 내적 감각이 심적 상태의 기능적 속성을 입력값으로 갖는다는 것은 반드시 각 심적 상태를 완전히 기능적으로 정의하기 위해 필요한 모든 연언들을 입력값으로 갖는다는 것을 의미하지 않을 수도 있기 때문이다. 대부분의 기능주의자들은 심적 상태를 기능적으로 정의하기 위해서 각 심적 상태가 갖는 모든 관계적 속성들을 열거하려고 하지는 않는다. 여기선 슈메이커의 예시를 가져와 보겠다. 어떤 개체에서는 고통을 전체 실현하는 상태와 특정한 생리학적 상태 P가 공존하면 체온이 높아진다고 하자. 그렇다면 고통이라는 심적 상태는 해당 개체에게서 상태 P와 공존하면 체온 상승을 야기한다는 관계적 속성을 필연적으로 가질 것이다. 그러나 이러한 관계적 속성이 고통의 기능적 정의에 포함되어야 한다고 주장할 기능주의자는 아마 없을 것이라고 슈메이커는 지적한다.[23] 유사한 관점에서, 내적 감각 또한 심적 상태를 식별하기 위해 해당 심적 상태를 완전히 기능적으로 정의할 수 있을 정도의 정보를 요구하지 않을 수 있다. 예를 들어 고통의 완전한 기능적 정의를 여기

22) Ibid.
23) Shoemaker, *Some Varieties of Functionalism* (1981)

서 제시하지는 못하더라도 그 정의는 신체적 손상에 의해 야기되는 경향이 있다는 것과 움찔거리는 행동을 야기한다는 두 가지 항보다는 많은 연언적 항들을 가질 것은 분명하다. 그러나 이 두 가지 항만 가지고도 내적 감각이 고통을 식별할 수 있을 가능성은 존재하는 듯하다. 이는 마치 우리가 길에서 강아지를 볼 때 굳이 강아지의 유전자까지 검사해보지 않아도 강아지의 외견과 냄새, 촉감 정도로도 우리가 본 것이 강아지라고 식별할 수 있는 것과 같다고 할 수 있다. 심적 상태의 기능적 속성에 포함된 연언들의 특정한 일부가 감지된다면 내적 감각은 해당 심적 상태의 표상을 생성할 수 있다는 부분적 매칭의 가능성을 인정한다면, 하나의 심적 상태를 식별하기 위해 감지되어야 할 관계항의 수를 줄일 수 있다.

이러한 부분적 매칭을 통한 식별을 통해 기능주의적인 내적 감각 이론이 해당 문제를 해결할 가능성은 골드만 본인도 일부 인정하였다. 또한 부분적 식별의 가능성은 골드만이 생각한 것 이상으로 내적 감각에 잘 들어맞는 설명이 될 추가적인 이유도 존재한다. 외부 지각을 통해 획득한 지식과 마찬가지로, 내성을 통해 획득한 자기 지식 또한 확실함의 정도가 저마다 다를 수 있다. 이는 내성을 진행할 때 얼마나 내적 주의를 기울였는지, 어떤 의식적 상태에 놓여있었는지 등의 요인에 영향을 받는다. 많은 내적 주의를 기울여서 내적 감각이 심적 상태의 완전한 기능적 정의에 가까운 정도의 기능적 속성을 입력값으로 갖는다면 그 결과로 생성된 자기 믿음은 더욱 확실하고 오류를 범할 가능성이 비교적 낮을 것이고, 큰 주의를 기울이지 않고 심적 상태를 간신히 식별할 수 있을 정도의 정보만을 활용해 생성된 자기 믿음은 훨씬 더 틀릴 가능성이 클 것이다. 앞서 기존의 내적 감각 이론들을 살펴볼 때 내적 감각 이론이 가졌던 특징 중 하나는 내적 감각을 외부 지각과 유사하다고 봄으로써 내성의 오류 가능성을 인정하는 것이었다. 이런 관점에서 부분적 매칭을 통한 식별은 내적 감각의 작동 방식에 매우 잘 들어맞는 설명이라 할 수 있다. 외부 지각도 대상을 식별하는데 대상이 가진 모든 지각적 특징을 활용하지 않는다는 점에서 내적 감각과 외부 지각의 유비도 잘 성립하며, 내성의 오류 가능성도 잘 설명하고 있기 때문이다. 그리고 심적 상태가 갖는 기능적 속성들 중 일부만을 입력값으로 사용해서도 내적 감각이 작동할 수 있다면, 관계항의 폭발적 증가의 위협은 상대적으로 줄어든다.

둘째로, 관계항 수의 폭발적 증가 자체가 그렇게 심각한 문제가 아닐 수 있다. 골드만은 부분적 매칭을 통한 해결책이 관계항의 폭발적 증가를 완전히 배제해주지는 않기 때문에 해결책으로서 충분치 않다고 볼 수 있다. 그러나 반대편에서는 한층 더 강수를 두어 관계항의 수가 기하급수적으로 늘어나더라도 내적 감각이 이를 충분히 감당할 수 있다고 주장할 수 있다. 인간의 인지적 능력이 엄청난 양의 정보를 다루기에 적합할뿐더러 실제로도 엄청난 양의 정보를 항상 다루고 있다는 것은 이제는 널리 인정될 만한 사실이다. 더군다나 내적 감각은 외부 지각과 달리 특정한 신체 기관에 국한되지 않고, 인간의 뇌와 중추신경계의 전체적인 신경적 연결망에 넓게 분산되어 작동하는 매우 포괄적인 메커니즘으로 이해되고 있다. 이러한 연결적 메커니즘이 엄청난 양의 정보를 다루기에 적합하다는 사실은 인간의 뇌의 신경적 구조를 모방한 연결주의적 기계들의 성공적인 개발에 의해서도 지속적으로 입증되고 있다. 그렇기에 관계항의 수가 폭발적으로 증가할 가능성이 존재하더라도, 내적 감각은 그 모든 관계항들을 처리함으로써 심적 상태를 기능적 속성을 통해 감지할 수 있다고 주장할 수 있다. 그렇다면 관계항의 폭발적 증가를 이유로 기능적 속성이 입력값이 되지 못한다는 골드만의 주장은 큰 설득력이 없을 것이다.

다음으로는 내적 감각이 인과적 관계를 감지할 수 없다는 주장을 살펴보자. 골드만은 인과적 관계가 직접 관찰될 수 없다는 흄의 주장을 받아들여 내적 감각 또한 인과적 관계를 직접 감지할 수 없다고 본다. 그러나 암스트롱은 굳이 흄적인 관점을 받아들일 필요가 없으며, 오히려 우리가 일상적으로 인과성을 직접 인지할 수 있다고 답변한다. 그는 예시로 어떤 물체가 우리 몸을 누를 때, 바로 그 물체가 우리의 신체에 인과적으로 작용해 압박을 가하고 있다는 사실을 직접적으로 인지할 수 있다고 설명한다. 또한 그는 이러한 인과성 감지 능력은 대부분의 생명체들에게 당연하고 필수적인 능력이라고 지적한다. 자신의 신체에 발생한 감각이 어떤 외부 요인에 의해 야기되었는지를 아는 것은 생존과 직결되는 문제기 때문이다.[24] 다만 이때 내적 감각이 감지하는 인과성이 정확히 어떤 것인지에 대해서는 추가적인 설명이 필요해 보인다. 예컨대 반사실적 인과 이론을 받아들이는 기능주의자의 입장에서는 심적

24) Armstrong, *Causes are Perceived and Introspected* (1993)

상태의 기능적 정의에 포함되는 인과 관계들도 반사실적 인과 관계들일 것이다. 그렇다면 이렇게 기능적으로 정의된 심적 상태를 내성하기 위해서는 내적 감각도 반사실적 인과 관계를 감지한다고 말해야 할 것으로 보인다. 그러나 내적 감각이 인간-하위적(Sub-personal)인 메커니즘으로 이해된다면, 심적 상태의 기능적 속성에 포함된 인과적 관계를 파악할 때 사용되는 인과성의 개념이 인지적 주체가 가진 인과성에 대한 이론적 개념과 꼭 일치한다는 보장은 없을 듯하다. 누군가가 인과적 관계를 반사실적 관계를 통해 이해한다고 해서 그의 내적 감각 또한 반사실적 인과성을 감지하지는 않을 수 있는 것이다. 정확히 어떤 인과성의 개념이 내적 감각에서 활용되는지는 아직 철학적으로도, 경험과학적으로도 확실히 말할 근거가 없어 보인다. 그러나 내적 감각 자체가 인과적 관계를 감지할 수 없다고 말할 이유도 충분치 않다고 할 수 있다. 따라서 기능적 속성이 관계적 속성이라는 점은 기능적 속성이 내적 감각의 입력값이 되지 못할만한 이유라 볼 수 없다.

다만 이와 같은 답변은 분석적 기능주의보다는 심리적 기능주의에 더욱 친화적일 수 있다. 내적 감각에 의해 감지된 인과적 관계가 우리의 일상적 관념들을 분석해서 얻어진 심적 상태의 기능적 정의에 포함되는 인과적 관계와 일치하지 않는다면, 분석적 기능주의의 입장에서는 감지된 대상이 분석적 기능주의에서 정의하는 심적 상태와 다르다고 봐야 하기 때문이다. 한편, 심리적 기능주의는 기능적 정의에 포함되는 인과적 관계가 최선의 이론에 의해 알려지는 것들이라고 보기 때문에, 만약 내적 감각이 활용하는 인과성의 개념이 알려진다면 그것을 심적 상태를 정의하는 방식에 가져다 사용할 수 있다. 그런 점에서 인과적 관계의 감지에 관한 비판에 대해서는 분석적 기능주의보다는 심리적 기능주의가 더 잘 답변할 수 있다. 물론 분석적 기능주의가 해당 답변을 아예 활용할 수 없는 것은 아니다. 반대로 내적 감각이 심적 상태들을 관계적 속성으로서 감지할 때 사용되는 인과성의 개념이 우리가 일상적으로 갖는 인과성의 개념과 어떤 식으로든 이미 연관되어 있다고 생각한다면, 내적 감각이 파악하는 인과 관계들은 분석적 기능주의가 추구하고자 하는 심적 상태의 기능적 정의에 잘 부합한다고 볼 수도 있다.

3. 3. 2. 기능적 속성은 경향적 속성이라는 점에 대하여

골드만은 기능적 속성이 경향적(Dispositional) 혹은 가정적(Subjunctive) 속성이라는 점도 내적 감각의 입력값으로서 부적절한 이유라고 주장한다. 그는 경향적 속성에 대한 정보가 범주적 속성에 대한 정보에 비해 훨씬 얻기 어렵다고 말하는데, 이는 경향적 속성이 항상 발휘되는 것이 아니며, 그렇기 때문에 직접적으로 알려지지 않고 오직 추론적으로만 알려질 수 있기 때문이다.[25] 설령 내적 감각이 인과적 관계를 감지할 수 있더라도, 이는 실제로 인과적 작용이 발생했을 경우에만 감지되는 것일 터이다. 발휘되지 않은 채 잠재적으로 있는 경향적 속성을 내적 감각이 무슨 수로 감지하겠는가? 그러나 기능적 속성은 경향적 속성들의 나열로 이루어져 있다. 예컨대 고통의 기능적 정의에 따르면 고통은 찡그림이나 움찔거림을 야기하려는 경향을 가질 뿐이지, 반드시 찡그림이나 움찔거림을 야기하는 것은 아니다. 그렇다면 찡그림을 야기하려는 경향이 발휘되지 않아서 찡그림이 실제로는 야기되지 않은 경우에 내적 감각은 이를 감지할 수 없을 것이다. 나아가 경향적 속성이 실제로 발휘되었더라도 내적 감각이 이를 단순한 범주적 속성과 구분되는 경향적 속성으로서 감지할 수 있을지도 의심스러울 수 있다. 골드만은 고통이 찡그림을 야기한 것을 내적 감각이 감지했을 때, 내적 감각이 이러한 인과 관계가 경향적인 것인지, 아니면 범주적인 것인지를 구분할 수 없다고 보고 있다.

지각과의 비유도 경향적 속성이 직접 파악될 수 없음을 보여줄 수 있다. 유리는 부딪혔을 때 깨지려는 경향적 속성을 갖고 있지만, 당장 유리가 깨지고 있는 것을 보고 있는 도중이 아니라면 해당 경향적 속성은 시각을 통해서 알려지지 않고, 지금 유리가 깨지는 것을 보고 있다고 하더라도 단일 사례만으로는 유리가 부딪혔을 때 깨지려는 속성이 경향적 속성인지, 아니면 범주적 속성인지를 구분할 수 없다. 예컨대 유리를 처음 보는 사람의 입장에서는 유리가 별다른 조건 없이 부딪히면 항상 깨지는 것인지, 때때로 특정한 조건이 만족되었을 경우에만 깨지는 것인지 구분할 수 없는 것이다. 마찬가지로 고통이 찡그림을 야기한다는 경향적 속성은 실제로 찡그림이 야기되고 있을 때에만 내적 감각에 인과적으로 작용할 수 있고, 그때마저도 내적 감각이 주

25) Goldman, *The Psychology of Folk Psychology* (1993)

어진 입력값을 경향적 속성으로 파악한다고 하기 어렵다. 따라서 내적 감각이 경향적 속성의 연언인 기능적 속성을 통해 심적 상태를 파악한다고는 볼 수 없다는 것이 골드만의 주장이다.

그렇다면 이에 대해서 어떤 답변이 가능한가? 경향적 속성에 대한 비판과 관련해서도 앞서 언급되었던 부분적 매칭을 통한 식별 가능성은 유용하게 활용될 수 있다. 부분적 매칭이 가능하다면 기능적 속성에 포함되는 어떤 개별 관계적 속성도 내적 감각이 심적 상태를 식별하기 위한 필요조건이 아닐 수 있기 때문이다. 만약 심적 상태의 기능적 속성에 포함되는 모든 인과 관계들을 파악함으로써 내성이 이루어진다면, 잠재적인 상태로만 남아있고 발휘되지 않은 경향적 속성의 존재는 기능적 속성을 통한 내성에 문제가 될 것이다. 예컨대 고통이 찡그림을 야기하지 않은 경우에 내적 감각은 고통을 기능적으로 정의하는 연언 중 하나가 만족되지 않았으므로 고통을 내성하지 못하는 것이다. 그런데 부분적 식별을 허용한다면, 잠재적인 상태로 있는 경향적 속성은 그저 해당 회차에는 내성에 사용되지 않으면 그만이다. 고통이 찡그림을 야기하지 않았다면, 내적 감각이 고통이 찡그림을 야기하려는 경향을 가진다는 속성이 아닌 다른 경향적 속성들을 파악함으로써 고통을 내성하면 된다는 것이다. 그렇다면 기능적 속성이 경향적이어서 실제로 인과적 작용이 발생하지 않을 수도 있다는 사실 자체는 심적 상태의 식별 자체를 불가능하게 만들 정도의 타격은 아니다. 그러나 이와 같은 답변은 한 가지 한계가 있다. 심적 상태의 기능적 정의에 포함되는 경향적 속성들이 전부 동시에 잠재적인 상태로 남아있고 발휘되지 않는 경우에는 내적 감각이 심적 상태를 내성하는 것이 불가능해질 수 있다. 즉, 기능적 속성을 통한 내성에서는 부분적 식별을 허용하더라도 최소한 식별이 가능할 만큼은 경향적 속성들이 실제로 발휘되어야만 한다. 다만 이와 같은 한계가 기능적 속성에 기반한 내적 감각 이론에 매우 심각한 타격이 되는 것 같지는 않다. 내적 감각은 외부 지각과 유사한 만큼 오류를 범하기도 하며, 내적 감각을 통한 내성은 스스로의 심적 상태에 관한 참인 믿음을 산출하는 데 때로는 실패할 수도 있더라도 전체적으로 신빙성 있는 과정이면 된다. 그렇기에 모든 경향적 속성이 발휘되지 않는 것과 같은 특수하고 예외적인 경우에 내성이 실패하는 것은 내적 감각 이론이 충분히 감수할 수 있는 한계로 보인다.

또한 골드만이 경향적 속성 자체를 너무 난해한 것으로 보고 있다는 지적도 있다. 스테렐니는 우리가 일상적으로 경향적 속성에 관한 정보를 접할 수 있으며, 경향적 속성에 대한 접근을 어떻게 해명하는가의 문제는 기능주의가 특별히 겪는 문제는 아니라고 지적한다.[26] 실제로 외부 지각을 통해서 경향적 속성을 감지하는 것은 흔한 일이다. 앞서 언급한 사례로 돌아가보면, 유리는 부딪혔을 때 깨지려는 경향적 속성을 갖고 있다. 그런데 대부분의 사람들은 유리가 깨지는 것을 볼 때, 이것이 유리의 범주적 속성이 아니라 경향적 속성이라는 것을 곧바로 알 수 있는 것처럼 보인다. 다시 말해, 사람들은 유리가 깨지는 것을 보면서도 유리는 부딪히면 항상 깨진다고 오해하지는 않는다는 것이다. 이는 학습을 통해 가능해지는 것이라 할 수 있다. 당연히 유리를 아예 처음 보는 사람은 유리가 부딪혀 깨지는 것을 보면서 부딪혔을 때 깨짐이 유리의 범주적 속성인지, 경향적 속성인지 구분하지 못할 것이다. 그러나 경험이 누적되고 유리가 때로는 부딪혔을 때 깨지고, 때로는 부딪혔을 때 깨지지 않는 것을 보게 되면서 해당 사람은 차츰 별도의 추론적 과정을 거치지 않고서도 유리가 부딪혀 깨지는 것을 보고서 이것이 유리의 경향적 속성이 발휘되고 있기 때문이라는 것을 곧바로 알 수 있게 된다. 그렇다면 이제 기능주의적 내적 감각 이론을 옹호하는 측에서는 내적 감각이 심적 상태의 경향적 속성을 파악하는 것도 유사한 방식으로 이루어질 것이라고 말할 여지가 있다. 막 태어난 아이의 내적 감각은 심적 상태를 내성할 때 경향적 속성과 범주적 속성을 구분하지 못할 것이지만, 경험이 누적되고 심적 상태의 기능적 정의에 속하는 인과 관계들이 항상 발휘되지는 않는다는 사실을 학습하는 과정을 거치게 되면서 차츰 아이의 내적 감각은 심적 상태를 경향적 속성들을 통해서 내성하기 시작할 수 있는 것이다. 이와 같은 지각과의 유비를 받아들인다면 기능적 속성이 경향적 속성이라는 점도 내적 감각의 입력값이 되지 못할 결정적 이유를 제공하지는 못한다고 할 수 있다. 이와 같은 답변은 분석적 기능주의적 입장과 심리적 기능주의적 입장 양쪽에서 모두 활용할 수 있는 답변이다.

---

26) Sterelny, *Categories, Categorisation and Development: Introspective Knowledge is No Threat to Functionalism* (1993)

3. 3. 3. 기능적 속성은 연언적 속성이라는 점에 대하여

기능적 속성은 관계적이고 경향적인 속성들의 연언이다. 골드만은 기능적 속성이 관계적 속성이라는 점과 경향적 속성이라는 점에 비판을 집중했지만, 혹자가 보기에는 연언적 속성이라는 사실도 문제가 될 수 있다. 일단 내적 감각이 심적 상태의 정의에 포함되는 다양한 경향적이고 관계적인 속성들을 감지할 수 있다고 인정했다고 하자. 그런데 어째서 내적 감각이 이러한 속성들을 하나의 심적 상태를 구성하는 속성들로서 감지할 수 있는가? 사람의 내부 구성은 매우 복잡하며, 내적인 인과 관계들의 수도 무수히 많다. 지루하더라도 고통의 예시로 다시 돌아와보면, 고통이라는 심적 상태의 기능적 정의를 구성하는 연언들에는 신체적 손상에 의해 야기하려는 경향적 속성, 찡그림 및 움찔거림을 야기하려는 경향적 속성, 불쾌감을 야기하려는 경향적 속성, 고통으로부터 벗어나려는 욕구를 야기하려는 경향적 속성 등 많은 경향적이고 관계적인 속성들이 포함된다. 그리고 고통은 기능적 정의에 포함되지 않는 무수히 많은 인과 관계들의 관계항이기도 하다. 경우에 따라서 고통은 체온 상승을 야기한다거나, 사람을 기절시키는 원인이 될 수도 있고, 때로는 쾌락을 발생시키기도 할 수 있다. 그런데 어째서 내적 감각은 더도 말고, 덜도 말고 특정한 관계적 속성들만을 하나의 심적 상태로 묶는가? 내적 감각이 연언적 속성인 기능적 속성을 감지함으로써 심적 상태를 내성한다고 주장하기 위해서는 이와 같은 속성의 묶음이 내적 감각에 의해 어떻게 수행되는지를 설명할 필요가 있다.

설명되어야 할 것은 크게 두 가지로 정리할 수 있다. 첫째로는 내적 감각이 애초에 어떻게 속성들을 하나의 묶음으로서 감지할 수 있는지이다. 내적 감각이 신경적 속성을 통해 심적 상태를 내성한다고 보는 골드만의 입장에서는 이러한 설명적 부담이 비교적 덜하다고 말할 여지가 있는데, 여러 가지 신경적 속성을 묶어서 감지함으로써 하나의 심적 상태를 내성하는 것이 아니라 그 심적 상태와 동일한 신경적 상태를 곧바로 감지하면 된다고 주장할 수 있기 때문이다. 둘째로는 내적 감각이 어떻게 여러 가능한 관계적 속성들의 묶음 중에서 오직 특정한 묶음만을 하나의 심적 상태를 구성하는 것으로서 파악할 수 있는지이다. 역시 이러한 설명적 부담은 신경적 속성을 옹호하는 입

장에서는 비교적 덜 느낄 수 있는데, 신경적인 연결 구조가 이미 하나의 단위를 형성하고 있다면 내적 감각은 이미 형성된 신경적 단위를 파악하기만 하면 된다고 주장할 수 있기 때문이다.

우선 속성들의 묶음에 관한 첫 번째 문제에 답변하는 방식은 어느 정도는 '원래 그렇다'는 원론적인 대답처럼 비칠 수 있다. 내적 감각 이론은 지각과의 유비를 중요시하는 만큼, 지각에서 여러 가지 속성들이 하나의 대상으로 묶일 수 있다는 사실은 내적 감각에서도 여러 가지 속성들이 하나의 대상으로 묶일 수 있다고 볼 만한 이유를 제공한다. 예컨대 내가 콜라 캔을 바라보고 있을 때, 콜라 캔은 검정색과 빨간색과 같은 색에 관한 속성들과 길쭉하고 원통형이라는 형태적 속성들을 포함해 여러 시각적 속성들을 가지고 있다. 그럼에도 나의 시각은 이 여러 시각적 속성들을 콜라 캔이라는 하나의 시각적 대상을 구성하는 것으로서 받아들인다. 과거에는 시각적 정보 자체는 대상으로 구분되지 않는 무차별적 내용만을 포함하며, 그 시각적 정보를 적절히 취합해 대상으로 구성하는 것은 더 상위의 인지적 과정을 통해서라고 이해되기도 했으나, 현대에 이루어지고 있는 지각과 관련된 많은 철학적 및 인지과학적 논의들은 지각 자체가 이미 어느 정도 대상 단위로 구분된 내용을 갖는다고 보고 있다. 내적 감각과 관련해서도 마찬가지로 대답을 해야 할 것이다. 내적 감각은 지각과 유사하게 여러 속성들을 하나의 대상으로 묶어서 감지할 수 있다는 근본적인 특성을 지니는 것이다.

묶음에 관한 두 번째 문제에 대한 답변은 크게 두 가지 방면에서 이루어질 수 있다. 하나는 우리의 내적 감각이 선학습적으로 어떤 관계적 속성들이 하나의 심적 상태를 구성하는지를 알고 있다고 주장하는 것이고, 다른 하나는 내적 감각이 학습을 통해서 심적 상태가 어떤 관계적 속성들의 연언으로 구성되는지를 알게 된다고 주장하는 것이다. 이 두 가지 답변은 서로 배타적이지 않고, 상호보완적이라 할 수 있다. 지각이 대상을 구분하는 방법은 일부는 선학습적으로 지각의 작동 방식 자체에 입력되어 있고, 일부는 학습된다. 태어난 지 얼마 되지 않은 아기가 눈을 처음 뜨고 시각적 정보를 받아들일 때, 아기가 보는 세상은 아무것도 구별되지 않은 색과 형태의 혼란스러운 덩어리가 아니라 이미 어느 정도 대상들로 구분되어 있는 상태일 것이다. 아기는 별다른 학습 없이도 어떤 색과 형태의 조합이 하나의 대상을 구성하고, 나

머지는 그 대상에 속하지 않는 배경인지를 대략적으로는 알 수 있다. 만약 아기의 시각적 내용이 아무것도 구별되지 않은 혼란스러운 상태였다면, 물체가 다가오거나 멀어지고 나타나는 것에 전혀 반응할 수 없었을 것이다. 그러나 눈을 뜬 지 얼마 되지 않은 아기도 어머니나 아버지가 다가오고 멀어지는 것에 반응하는 만큼, 인간의 시각은 내용을 묶어서 대상을 구성하는 방식에 대한 선학습적인 능력을 갖추고 있다. 또한 시각적 대상을 구성하는 방식은 학습되기도 한다. 별자리를 전혀 모르는 사람에게 별이 가득한 밤하늘은 그저 어지럽게 흩어진 별 무더기처럼 보이겠지만, 별들을 별자리로 묶어서 보는 것이 익숙한 천문학자의 눈에는 밤하늘을 바라볼 때의 시각적 정보가 이미 별자리로 어느 정도 구분되어 있을 수 있다.

내적 감각도 마찬가지다. 가장 기본적인 심적 상태들 몇 가지와 관련해서는 어떤 관계적 속성들의 결합이 어떤 심적 상태를 구성하는지가 내적 감각의 작동 방식에 이미 선학습적으로 입력되어 있을 수 있다. 또한 내적 감각은 학습을 통해서도 특정한 관계적 속성들을 하나의 심적 상태로 묶게 될 수 있다. 지각을 통해 포착된 여러 속성들이 하나의 지각적 대상으로 묶일 수 있다면, 내적 감각이 다양한 관계적이고 경향적인 속성들을 하나의 심적 상태로 묶는다는 사실도 특별히 문젯거리가 될 것은 없어 보인다. 따라서 기능적 속성이 연언적 속성이라는 점은 내적 감각의 입력값이 되지 못할만한 사유가 아니라고 말할 수 있겠다.

또한 이러한 묶음의 문제는 비단 기능적 속성에서만 발생하는 것이 아닐 수도 있다. 신경적 속성을 통해서 내적 감각이 심적 상태를 내성한다고 주장하는 경우에도 여러 가지 신경적 속성들이 하나의 심적 상태를 구성할 가능성은 있다. 이는 특히 복잡한 내용을 갖는 믿음이나 욕구와 같은 심적 상태의 경우에 그러할 것이다. 내적 감각이 수많은 신경적 연결 관계들 가운데 어떠한 신경적 연결 구조가 하나의 심적 상태를 구성하는 것으로 파악하는지를 설명하는 것은 내적 감각이 여러 가지 관계적 속성들 중에서 어떤 연언적 속성이 하나의 심적 상태를 구성하는 것으로 파악하는지를 설명하는 것과 크게 다르지 않으며, 따라서 기능적 속성이 연언적 속성이라는 점은 신경적 속성에 비해 내적 감각 이론에 특별히 불리하게 작용하지는 않는다고 할 수 있다.

3. 3. 4. 기능적 속성의 인과적 효력에 대하여

내적 감각은 심적 상태와 인과적으로 작용하는 메커니즘으로서, 내적 감각의 입력값은 내적 감각에 인과적인 영향을 끼칠 수 있어야 한다. 그러나 일부 학자들은 심적 상태가 기능주의적으로 정의된다면 심적 상태는 인과적 효력을 상실할 것이라는 우려를 표한 바 있고, 골드만도 간략하게나마 이러한 우려를 언급하며 신경적 속성은 인과적 효력을 의심받지는 않는다는 점에서 유리하다고 말하고 있다. 이러한 우려에 관한 논의는 크게 형이상학적으로 필연적인 효과의 문제와 배제 문제라는 이름을 달고 다루어지고 있다. 심신 인과와 관련된 이 두 문제에 관한 상세한 기능주의적 답변을 제시하는 것은 본 논문이 목표하는 범위를 벗어나지만, 간략하게나마 해당 문제들이 내적 감각 이론과 어떤 관련성이 있는지를 확인할 필요는 있을 것이다.

우선 형이상학적으로 필연적인 효과의 문제를 루퍼트의 논의를 바탕으로 내적 감각 이론에 맞게 적절히 변형시켜 소개해보겠다.[27] 심적 상태의 기능적 정의에는 심적 상태가 다른 심적 상태와 어떻게 인과적으로 상호작용하는지도 포함된다. 그렇다면 심적 상태가 내적 감각에 인과적으로 작용하여 자기 믿음이라는 또 다른 심적 상태를 산출하는 것 또한 심적 상태의 인과적 관계 속성 중 하나로서 그 기능적 정의에 포함될 것이다. 한편, 내적 감각은 심적 상태를 기능적 속성을 통해 감지한다. 그리고 감지되는 기능적 속성에는 내적 감각에 인과적으로 작용해 자기 믿음을 산출한다는 관계적 속성이 포함된다. 그렇다면 이제 다음과 같은 구도가 완성된다: 내적 감각에 인과적으로 작용해 자기 믿음을 산출하는 심적 상태가 내적 감각에 인과적으로 작용해 자기 믿음을 산출한다. 그런데 이는 형이상학적으로 필연적이다. 그런데 과연 형이상학적으로 필연적인 관계가 인과적 관계라고 할 수 있는가? 루퍼트는 인과에 대한 규칙성 이론과 반사실적 이론 모두 이와 같은 형이상학적으로 필연적인 관계에서는 인과성을 주장할 수 없다고 말하는데, 이를 받아들인다면 기능적 속성은 내적 감각을 통해 생성된 자기 믿음과 인과적 관계에 놓일 수 없고 내적 감각의 입력값으로서 부적절하다.

배제 문제는 기능적으로 정의된 심적 상태가 물리적 상태인 실현자

27) Rupert, *Functionalism, Mental Causation, and Necessary Effects* (2006)

상태에 의해 실현된다고 할 때, 실제로 인과적으로 관여하는 것은 실현자 상태의 물리적 속성들이고 기능적 속성들은 인과적으로는 잉여적인 속성들에 불과하다고 말한다. 심적 상태를 실현하는 실현자 상태는 물리적인 인과 법칙에 따라서 내적 감각을 실현하는 물리적 메커니즘에 인과력을 행사할 것이고, 물리적 메커니즘은 다시 물리적인 인과 법칙에 따라서 자기 지식을 실현하는 다른 실현자 상태에 인과력을 행사할 것이다. 그렇다면 기능적 상태로서의 심적 상태들에게는 특별히 남아있는 인과적 역할이 없고, 그저 실현자 상태에 의해 실현되거나 실현자 상태에 수반함으로써 내적 감각 메커니즘과 관계한다고 볼 수 있다. 이에 동의한다면 내적 감각을 통한 내성에 인과적으로 관여하는 속성 역시 골드만이 주장한 것처럼 물리적 속성의 일종인 신경적 속성이어야 하고, 기능적 속성은 인과적 메커니즘인 내적 감각의 입력값이 될 수 없다고 말해야 할 것이다.

먼저 형이상학적으로 필연적인 효과의 문제는 기능주의적 내적 감각 이론의 입장에서 비교적 더 답하기가 쉽다. 해당 문제는 인과적 관계를 형성해야 할 두 관계항인 원인과 결과가 형이상학적 혹은 논리적 필연성으로 묶여 있기 때문에 발생한다. 그러나 내적 감각의 입력값으로서 작용하는 심적 상태의 기능적 속성이 내적 감각을 통한 자기 믿음의 형성을 필연화하지 않는다면 문제는 더 이상 발생하지 않는다. 필연화 관계를 해소하는 방안으로는 크게 두 가지가 있다. 첫째 방안은 심적 상태의 기능적 속성이 경향적 속성이라는 점에 주목함으로써 제시될 수 있다. 앞서 살펴본 것처럼, 경향적 속성은 항상 발휘되지는 않는다는 특징을 지니며, 심적 상태들도 항상 관련된 행동으로 이어지지는 않는다. 콜라를 마시고 싶은 욕구는 콜라를 마시려는 행동을 야기하려는 경향을 갖지만, 어떤 사람이 해당 욕구를 가지고 있다고 해서 무조건 콜라를 마시려는 행동을 나타내지는 않는다. 그런데 경향적 속성에 대해서는 크게 두 가지 형이상학적 견해가 존재한다. 하나는 경향적 속성이 항상 발휘되지 않는 이유는 경향적 속성이 조건적이지만 필연적으로 발현하기 때문이라는 입장이고, 다른 하나는 경향적 속성이 온전히 우연적으로 발현하기 때문에 항상 발휘되지 않는다고 보는 입장이다. 만약 후자에 동의한다면, 기능주의는 형이상학적으로 필연적인 효과의 문제를 회피할 수 있게 된다. 내적 감각에 인과적으로 작용해 자기 믿음을 산출하려는 경향적 속성은 필연적으로 내적 감

각에 인과적으로 작용해 자기 믿음을 산출하지 않게 되기 때문이다. 그러나 이러한 답변은 경향적 속성에 대해 특정한 이해를 전제한다는 점에서 완전한 답변이라고 보기는 어려운 점이 있다.

둘째 방안은 다시 내적 감각이 부분적 매칭을 통해 심적 상태를 식별할 가능성에 기대어 이루어질 수 있다. 내적 감각이 어떤 심적 상태를 내성할 때 그 심적 상태의 기능적 정의에 포함된 모든 관계적 속성들이 전부 내적 감각에 의해 감지될 필요가 없다면, 심적 상태가 내적 감각에 인과적으로 작용해 자기 믿음을 산출한다는 속성도 내적 감각의 내성 작용에 실제로 관여하는 속성의 목록에서 제외될 수 있을 것이다. 내적 감각에 인과적으로 작용하는 인과적 속성들은 형이상학적으로 필연적인 효과의 문제를 일으키지 않는 다른 속성들뿐이라면 형이상학적으로 필연적인 효과의 문제는 더 이상 내적 감각의 내성 과정에서 발생하지 않게 된다. 예컨대 고통의 기능적 속성에는 앞서 여러 차례 언급한 관계적 속성들과 더불어 내적 감각에 인과력을 행사해 고통에 대한 내성을 야기한다는 관계적 속성도 포함된다고 하자. 그런데 내적 감각이 고통을 내성할 때 실제로 감지하는 관계적 속성은 내적 감각에 인과력을 행사해 고통에 대한 내성을 야기한다는 관계적 속성을 제외한 나머지 속성들이라고 한다면, 기능적으로 정의된 고통과 내적 감각 사이에는 어떠한 필연적 관계도 형성되지 않는다.

배제 문제와 관련해서는 일반적인 기능주의적 답변 외에 내적 감각 이론과 특별히 관련지어 논의할 것은 없어 보인다. 기능주의적 입장에서 배제 문제에 답변하는 가장 흔한 방법은 이른바 일반화 문제를 역으로 제기하는 것이다. 즉, 심적 상태의 인과적 효력이 배제되는 방식은 수많은 특수 과학 분과들의 탐구 대상들이 되는 속성들에도 일반적으로 적용되어 이들로부터 인과적 효력을 마찬가지로 빼앗게 될 것이라고 대답하는 것이다. 다양한 특수 과학 분과들은 저마다 다른 층위에서 과학적 설명을 제시하고 있다. 예컨대 화학은 분자 단위에서 일어나는 화학적 작용들을 탐구한다. 그런데 결국 분자들이 갖는 화학적 속성들이 결국 분자 단위보다 더 아래로 내려가서 양자역학적인 수준에서 논하는 물리적 속성들에 의해서 설명될 수 있다고 한다면, 화학적 속성들은 양자역학적인 물리적 속성들에 수반할 뿐이고 그 자체로는 인과적 영향력이 없게 되는 것인가? 만약 여기에 동의할 수 없다면, 심적 상태에 대한

배제 문제에 관해서도 심적 상태가 인과적 효력을 빼앗기지 않는다고 말해야 한다.

다만 기능주의적 입장의 구분이 해당 문제에 어떤 식으로 연관되는지는 짚고 넘어가는 것이 좋겠다. 일견 실현자 기능주의는 심적 상태가 물리적 상태인 그 실현자와 동일하다고 봄으로써 심적 상태의 인과적 효력을 보장하여 배제 문제를 겪지 않을 수 있는 것처럼 보인다. 그러나 내적 감각이 기능적 속성을 통해 심적 상태를 내성할 수 있는지와 관련해서는 심적 상태가 물리적 상태와 동일하다는 실현자 기능주의의 주장이 실상 도움이 되지 않는데, 이는 심적 상태 자체의 인과적 효력만을 옹호하고자 하는 것이 아니라 심적 상태가 기능적 속성을 통해 인과적으로 내적 감각에 작용한다고 주장하고자 하기 때문이다. 심적 상태의 실현자는 물리적 상태로서 물리적인 속성도 가지지만, 심적 상태의 기능적 정의를 만족시키는 기능적 속성도 가지고 있다. 그런데 어떤 대상이 인과적 관계를 가질 때 그 대상의 모든 속성들이 인과적으로 관여하는 것은 아니다. 예컨대 크리스탈 컵은 깨지기 쉽다는 속성과 투명하다는 속성을 모두 지니지만, 컵이 바닥에 떨어져서 깨질 때 투명하다는 속성은 관여하지 않는다. 그런 의미에서 현재 검토 대상이 되는 논제는 기능적으로 정의된 심적 상태가 단순히 인과적 효력을 갖는지만이 아니라, 심적 상태가 물리적 상태와 동일하건, 동일하지 않건 간에 그것이 가진 기능적 속성들이 내적 감각의 내성 작용에 관여한다는 것이다.

이상과 같이 기능적 속성에 대해 가해진 총 네 가지 비판을 살펴보았고, 네 가지 비판 각각에 대해 기능주의적 관점에서 답변을 제시해보았다. 대부분의 문제와 관련해 제시된 답변들은 분석적 기능주의와 심리적 기능주의 양측에서 동일하게 활용할 수 있는 답변들이었으나, 첫 번째로 살펴보았던 관계적 속성과 관련된 비판에 대한 답변은 분석적 기능주의보다는 심리적 기능주의를 받아들이는 측에서 더 잘 제시할 수 있는 것으로 보였다. 또한 내적 감각 이론은 역할 기능주의와 실현자 기능주의 중 어느 쪽을 받아들이는지와는 관계 없이 기능적 속성을 통한 내성을 옹호해야 한다는 것도 확인해보았다.

4. 정리 및 결론

본 논문에서는 새롭고 독자적인 내적 감각 이론을 제시하고자 시도하지는 않았으나, 기존에 제시되었던 여러 내적 감각 이론들을 기능주의라는 틀을 통해 살펴보며 최선의 내적 감각 이론이 설명해야 할 쟁점들과 살려야 할 이론적 장점들을 확인하였다. 내적 감각 이론과 기능주의와의 관련성은 개별적인 내적 감각 이론들이나 혹은 기존 내적 감각 이론들을 정리해 소개한 다른 글들에서는 아직 크게 주목받지 못하였던 만큼, 이와 같은 새로운 접근법을 통해 내적 감각 이론을 분석하는 작업은 나름의 이론적 가치를 지니는 것으로 판단된다. 기존 이론들 가운데 암스트롱과 라이칸, 그리고 니콜스와 스티치가 제시한 내적 감각 이론은 모두 심적 상태에 관한 기능주의적 관점을 받아들이고 있었다. 암스트롱은 분석적 실현자 기능주의를, 라이칸은 심리적 역할 기능주의를, 그리고 니콜스와 스티치는 심리적 기능주의를 받아들여 각자의 내적 감각 이론을 전개했던 반면, 가장 최근에 자신의 이론을 전개했던 골드만은 기능주의적 내성 이론을 비판하며 대신 환원적 물리주의를 전제하는 신경적 속성에 기반한 내적 감각 이론을 주장하였다.

먼저 설명적 쟁점들과 관련해서는 내적 감각 이론은 심적 상태의 내용, 종류, 그리고 강도에 대한 표상이 가능하다는 것을 설명해야 했다. 이 세 가지에 관하여 가장 자세하게 논의를 제공했던 기존 이론은 골드만의 내적 감각 이론이었으나, 기능주의적인 내적 감각 이론이 이러한 쟁점들에 대해 만족스러운 설명을 제시할 가능성이 없는 것은 아니었다. 오히려 심적 상태의 서로 다른 내용, 종류, 그리고 강도는 심적 상태의 서로 다른 기능을 통해서 충분히 포착될 수 있을 것으로 생각된다. 예컨대 내가 갖고 있는 '대한민국의 수도는 서울이다'는 믿음은 누군가 나에게 "대한민국의 수도가 어디에요?"라고 묻는다면 서울이라고 대답하게 만드는 등, 나의 행동 및 다른 믿음과 욕구 등의 심적 상태들과 관련해 특정한 기능적 역할을 수행한다. 또한 믿음이라는 심적 상태는 욕구라는 심적 상태와 다른 기능을 가지는 등, 심적 상태가 갖는 기능의 종류에 따라 심적 상태의 종류도 구분할 수 있다. 또한 심적 상태의 강도 또한 심적 상태의 기능을 통해서 이해할 수 있는데, 예컨대 강한 믿음은 약한 믿음에 비해 야기하는 행동의 종류나 방식 등에서 차이를 보일 수 있고,

이와 같이 인과 관계적 차이를 보인다는 점에서 믿음의 강도는 기능적 차이로 이해될 수 있을 것이다. 그런 점에서 기능주의적인 내적 감각 이론은 골드만의 내적 감각 이론에 비해 심적 상태의 내용, 종류, 그리고 강도에 대한 표상을 설명하는 데 있어 부족함이 없다고 할 수 있다.

내적 감각 이론이 가질 수 있는 두 가지 이론적 장점으로는 모든 종류의 심적 상태에 대한 내성을 설명할 수 있는 통일적인 자기 지식 이론이라는 점과 물리주의와 친화적이면서도 물리주의를 반드시 전제하지는 않으며 심신 문제와 관련해 다양한 형이상학적 입장들에 대해 열려있을 수 있다는 점을 꼽았다. 기존 이론들 중 니콜스와 스티치의 이론은 현상적 내용을 갖는 심적 상태들에 대한 내성을 설명하는 것을 포기했다는 점에서 통일적인 자기 지식 이론으로서의 강점은 잃어버렸다. 한편, 골드만의 이론은 신경적 속성을 통해 내성이 이루어진다고 주장하며 물리주의를 전제함으로써 다양한 형이상학적 입장들을 수용할 수 있다는 장점을 포기했다. 소개한 두 가지 장점을 모두 살릴 수 있는 후보들은 암스트롱의 분석적 실현자 기능주의를 받아들인 내적 감각 이론과 라이칸의 심리적 역할 기능주의를 받아들인 내적 감각 이론으로, 둘 다 기능주의적 내적 감각 이론이었다. 그런데 실현자 기능주의는 심적 상태가 물리적 상태와 동일하다고 보는 만큼, 가장 넓은 스펙트럼의 형이상학적 입장들을 수용할 수 있는 것은 역할 기능주의라고 사료된다. 한편, 골드만의 내적 감각 이론은 내적 감각이 신경적 속성을 통해 심적 상태를 내성한다고 주장하며, 심적 상태가 갖는 다른 속성들이 내성에 관여함을 거부함으로써 기존 철학적 논의와 단절되며, 내성에 관한 상세한 탐구를 자연과학의 영역으로 넘겨버린다는 단점도 갖고 있었다. 따라서 기능적 속성을 통한 내성에 대해 제기된 비판들에 대해 답변을 제공할 수만 있다면, 다시금 기능주의적 내적 감각 이론을 옹호할 만한 동기가 존재함을 본 논문에서는 확인하였다.

이후 기능적 속성에 대해 가해진 비판들과 그에 맞서 제시할 수 있는 답변들을 살펴본 결과, 총 네 가지의 비판에 대해 각각 답변을 제시할 수 있었다. 제기된 비판은 기능적 속성이 관계적 속성이라는 점, 기능적 속성이 경향적 속성이라는 점, 기능적 속성이 연언적 속성이라는 점, 그리고 기능적 속성이 인과적 효력이 없다는 점과 관련된 것들로, 이러한 비판들에 대해서 본 논문은 기능주의적 관점에서 답변을 제시해보았다. 이 중 기능적 속성이 관계

적 속성이라는 점과 관련해서는 분석적 기능주의보다는 심리적 기능주의의 관점에서 더욱 만족스러운 답변을 제기할 수 있었다. 따라서 앞서 논의한 내적 감각 이론의 설명적 쟁점들과 이론적 장점들과 함께 고려해보면, 심리적 역할 기능주의를 받아들였을 때 내적 감각 이론이 자기 지식에 관한 이론으로서 설명적으로도 가장 충실해지고, 심신 문제와 관련해서도 가장 다양한 형이상학적 입장들을 포용할 수 있게 되는 것으로 보인다.

다만 심리적 역할 기능주의를 받아들이는 것은 내적 감각 이론에 몇 가지 이론적 과제 및 부담을 부여할 수 있다. 우선 심리적 기능주의는 분석적 기능주의에 비해 심적 상태에 대한 배외주의적인 색채를 띤다. 심적 상태의 기능적 정의가 최선의 과학적 이론들에 의해 제공된다면, 기능적 정의에 사용되는 인과 관계들 가운데는 신경 과학에 의해서 알려지는 매우 특수하고 인간에게 고유한 인과 관계들도 포함될 수 있게 된다. 그렇다면 정상적인 인간이 내성하는 심적 상태들은 그러한 인과 관계들을 내적으로 갖지 못한 개체들이 내성하는 심적 상태들과 동일하다고 말하기 어려워지고, 심리적 기능주의를 받아들인 내적 감각 이론이 설명할 수 있는 내성의 범위도 정상적인 인간의 내성으로 상당히 좁아지게 된다. 그러나 한편으론 심리적 기능주의를 받아들이는 내적 감각 이론은 오히려 이러한 단점을 신경적 속성에 기반한 골드만의 내적 감각 이론에 비해서 덜 심각하게 겪는다. 라이칸이 주장하는 것처럼 기능적 단계들의 연속적 위계가 존재함을 받아들이고, 다양한 심적 상태들이 각자 다른 기능적 단계들에 속한다고 주장한다면, 일부 심적 상태들은 비교적 높은 추상적 정도를 갖는 기능적 단계에 속함으로써 인간 외에도 다양한 개체들이 공유할 수 있는 방식으로 정의될 수 있는 여지가 있다.

내적 감각 이론이 역할 기능주의를 받아들였을 때 발생하는 문제점으로는 앞서 언급했던 심적 상태의 인과적 효력에 관한 배제 문제를 꼽을 수 있다. 그동안 많은 학자들이 기능주의적 입장에서 일반화 대응 등 여러 답변을 제시하기는 했으나, 아직 배제 문제가 완전히 해소된 철학적 문제라고 보기는 어려울 듯하다. 따라서 역할 기능주의를 받아들인 내적 감각 이론도 배제 문제와 관련된 이론적 부담은 지고 가야만 한다. 그러나 물리주의를 전제하지 않으면서도 자연과학적 탐구에는 친화적이고, 또한 모든 종류의 심적 상태에 대한 내성을 설명할 수 있는 통일적인 자기 지식 이론을 찾고 있는 이들에게

분명 심리적 기능주의를 받아들이는 기능주의적인 내적 감각 이론은 하나의 매력적인 선택지라고 할 수 있겠다.

참 고 문 헌

Armstrong, D. M. (1968), *A Materialist Theory of the Mind*, Routledge.

———————— (1981), *The Nature of Mind*, The Harvester Press.

———————— (1993), "Causes are Percieved and Introspected", [Peer commentary on "The Psychology of Folk Psychology" by A. Goldman]. *Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 29.

Block, N. (1978), "Troubles with Functionalism". *Minnesota Studies in the Philosophy of Science*, vol. 9, pp. 261-325.

Fetzer, J. (1993), "Goldman has not Defeated Folk Functionalism", [Peer commentary on "The Psychology of Folk Psychology" by A. Goldman]. *Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 42-43.

Gertler, B. (2001), "Introspecting Phenomenal States". *Philosophy and Phenomenological Research*, vol. 63, no. 2, pp. 305-328.

———————— (2003), *Privileged Access: Philosophical Accounts of Self-Knowledge,* Ashgate.

———————— (2010), *Self-Knowledge,* Routledge.

———————— (2012), "Renewed Acquaintance", in *Introspection and Consciousness* by D. Smithies and D. Stoljar. Oxford University Press.

Goldman, A. (1993), "The Psychology of Folk Psychology", *Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 15-28.

———————— (2006), *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Jackson, F. (1993), "Qualia for Propositional Attitudes?", [Peer commentary on "The Psychology of Folk Psychology" by A. Goldman]. *Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 52.

Levin, J. (2021), "Functionalism", in Stanford Encyclopedia of Philosophy, Winter 2021 Edition.

Locke, J. (1689), *An Essay Concerning Human Understanding*. Oxford University Press(1999).

Lycan, W. (1981), “Form, Function, and Feel”. *The Journal of Philosophy*, Vol. 78, no. 1, pp. 24-50.

———————— (1987), *Consciousness*. The MIT Press.

———————— (1996), *Consciousness and Experience*. The MIT Press.

Nichols, S. and Stich, S. (2003), *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press.

Picciuto, V. and Carruthers, P. (2014), “Inner Sense”, in *Perception and Its Modalities* by D. Stokes et al. Oxford University Press.

Rupert, R. (2006), “Functionalism, Mental Causation, and the Problem of Metaphysically Necessary Effects”. *Nous*, vol. 40, no. 2, pp. 256-283.

Shoemaker, S. (1981), “Some Varieties of Functionalism”. *Philosophical Topics*, vol. 12, no. 1, pp. 93-119.

Sterelny, K. (1993), “Categories, Categorisation and Development: Introspective Knowledge is No Threat to Functionalism”, [Peer commentary on “The Psychology of Folk Psychology” by A. Goldman]. *Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 81-83.

Abstract

On Functionalistic Inner Sense Theory

Hyunchae Kim

Department of Philosophy, Major in Western Philosophy

The Graduate School

Seoul National University

Keywords : Self-knowledge, Introspection, Inner sense, Psychofunctionalism

Student Number : 2020-27285

This paper argues that when the inner sense theory, as a theory of self-knowledge and introspection, can not only offer the most appropriate explanation about the introspection of various types of mental states, but can also be open to more various metaphysical positions regarding the mind-body problem by avoiding the supposition of reductive physicalism. The inner sense theory states that we acquire knowledge about our own mental states through a process of looking inward called introspection, and that introspection is similar to perception in important ways. The scholars who have contributed to the inner sense theory in the tradition of contemporary analytic philosophy include Armstrong, Lycan, Nichols & Stich, and Goldman. This paper examines the contents of each of their inner sense theories in relation to their respective positions regarding

functionalism about mental states. As the inner sense theory is a theory that aims to explain the introspection of mental states, the acceptance of functionalism implies that the mental states that are introspected by inner sense are to be defined in functionalistic terms. Then it can be said that for functionalistic inner sense theories, inner sense detects mental states through their functional properties.

Among the supporters of the inner sense theory, Armstrong, Lycan, Nichols and Stich all accept functionalism. More specifically, Armstrong adopts realizer analytic functionalism, and Lycan supports role psychofunctionalism, while Nichols & Stich accept psychofunctionalism but does not take a specific stance between role functionalism and realizer functionalism. Meanwhile, Goldman, who has been the most recent supporter of the inner sense theory, rejects functionalism and strongly criticizes the idea that inner sense takes the functional properties of mental states as inputs. He instead proposes a neural inner sense theory, which claims that inner sense takes the neural properties of mental states in order to introspect them. This paper argues that while Goldman's theory does have certain advantages, it can be seen as too detached from previous philosophical discussions on introspection as it attempts to explain introspection purely through neural properties, and it also bears a stronger metaphysical burden compared to the earlier inner sense theories because it presupposes reductive physicalism. This paper argues that by returning to the functionalistic version of the inner sense theory, the theory can be a unified theory of introspection that fits well with natural sciences, while also staying metaphysically open about reductive physicalism.

In order to defend the functionalistic inner sense theory, one must offer responses to the criticisms that Goldman and other authors have launched against functional properties. Four main criticisms have been given against the idea that inner sense introspects mental states by detecting their functional properties. They are respectively about the fact

that functional properties are relational properties, the fact that functional properties are dispositional properties, the fact that functional properties are conjunctive properties, and the causal efficacy of functional properties. This paper examines each of the criticisms, and attempts to offer functionalistic responses to them. While most of the responses can be used by both analytic functionalism and psychofunctionalism, there are certain points where the discussion is more favorable toward psychofunctionalism. Therefore, this paper argues that the best version of the inner sense theory accepts psychofunctionalism.

감사의 글

항상 학문적 열정을 보여주시며 학자로서의 귀감이 되어주시고 부족한 제자를 늘 인내와 정성으로 이끌어주신 한성일 교수님과, 진솔하고 날카로운 피드백을 통해 논문 작성에 도움을 주신 김기현 교수님, 이우람 교수님께 감사함을 표합니다.

언제나 사랑과 믿음, 그리고 기도로 저를 응원해주시고 지탱해주시는 저의 할머니, 송정자 여사님께 이 글을 바칩니다.