이학석사 학위논문

진화적 도덕 반실재론 비판

2015년 8월

서울대학교 대학원

과학사 및 과학철학 협동과정

권 오 현

진화적 도덕 반실재론 비판

지도교수 장 대 익

이 논문을 이학석사 학위논문으로 제출함
2015년 5월

서울대학교 대학원
과학사 및 과학철학 협동과정
권 오 현

권오현의 석사 학위논문을 인준함
2015년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

국문초록

이 글은 두 가지 주장을 한다. 첫째, 도덕 판단을 정당화하는 도덕 사실은 마음 독립적이 아니라 반응 의존적이다. 둘째, 진화는 도덕 판단을 정당화하는 반응 의존적인 사실을 제공한다.

논의를 위해 진화적 도덕 반실재론자 리처드 조이스(R. Joyce)와 도덕 실재론자 제시 프린츠(J. J. Prinz) 사이에서 일어난 논쟁을 비교 검토한다. 조이스는 진화가 마음 독립적인 도덕 사실의 인식 가능성을 훼손한다는 '진화적 폭로 논증(evolutionary debunking arguments)'을 제기한다. 프린츠는 감정적 성향으로 반응 의존적인 도덕 사실이 구성된다는 '감수성 이론(sensibility theory)'으로 그에 맞선다. 이 글은 도덕 실재론을 구하는데 프린츠의 감수성 이론이 적격이라 생각한다. 따라서 프린츠의 주장을 비판적으로 보완한 '수정된 감수성 이론(modified sensibility theory)'을 제안하고자 한다. 수정된 감수성 이론의 핵심적인 논변은 다음과 같다. 첫째, 인간에게는 보편적인 도덕 감정을 일으키는 진화한 기초 가치가 있다. 둘째, 이상적 관찰자를 도입하면 감정 반응이 기초 가치로 수렴될 수 있다.

조이스는 프린츠의 감수성 이론이 도덕 실재론을 살리는 좋은 대안이 아니라며 세 가지 측면에서 비판했다. 첫째, 어떤 행동에 대해 느끼는 감정은 개인이나 집단마다 다를 수 있다. 히틀러와 나는 인종 청소를 보며 다른 감정을 느낀다. 곧 감수성 이론은 극단적인 상대주의를 함축한다. 둘째, 도덕 판단은 욕구나 감정과 무관하게 따라야 할 정언적 이유에 대한 믿음을 표현한다. 하여 감정 없이 행위 이유도 없다고 주장하는 감수성 이론은 도덕 요구의 정언성을 수용할 수 없다. 셋째, 감수성 이론이 그르다고 판단하는 행동과 상식 도덕관이 그르다고 판단하는 행동이 일치하지 않을 우려가 있다. 감수성 이론은 어떻게 행위자가 인종 청소를 불승인하리라 자신할 수 있는가?

수정된 감수성 이론은 도덕 판단의 객관성을 회복할 수 있다. 인간의 진화 역사에서 생존과 번식에 중요한 가치는 특정 감정 반응과 연결되었다. 성적으로 난잡한 사람에게 느끼는 역겨움이 그렇다. 이런 가치들이 우리가 도덕이라 부르는 심리, 행동 기제의 바탕이다. 인간 종은 기초 가치를 공유한다. 우리는 진화적 탐구로 승인과 불승인의 감정을 불러온 전형적인 가치들을 범주화할 수 있다. 따라서 다른 모든 조건이 동등하다면, 도덕 판단은 기초 가치에 느끼는 보편적인 감정으로 수렴된다. 이에 도덕 판단에 필요한 도덕 외적 사실을 모두 알고, 실천적으로 합리적인 이상적 관찰자를 상정한다. 현실의 관찰자가 느끼는 감정은 이상적 관찰자가 느낄 수 있는 감정으로 제한된다. 현실의 관찰자는 이상적 관찰자가 나타내리라 예상되는 반응을 참고한다. 내가 느끼는 감정을 이상적 관찰자도 느낄까 생각하는 행위는 올바른 도덕 판단으로 가는 길잡이다. 그리하여 현실의 관찰자가 나타내는 반응은 인간 종이 동의하

는 객관적인 도덕 판단이 된다. 이상적 관찰자는 경험적으로 입증된다. 인간에게는 가상적인 눈으로 자신을 바라보는 능력과 감정을 평가하는 감정인 메타감정이 있다.

수정된 감수성 이론은 도덕의 불가피한 실천적 힘을 보여줄 수 있다. 도덕은 우리 조상들의 생존과 번식에 매우 중요했다. 도덕적으로 행동하려는 욕구는 인간 본성의 일부이다. 그런데 현실의 관찰자는 무지로 말미암아 자기에게 좋은 욕구가 무엇인지 모를 수 있다. 행위자는 이상적 관찰자의 관점에 설 때 자기에게 좋은 욕구가 무엇인지 이해하고 그에 승인의 감정을 느낄 수 있다. 좋음을 알고 승인하는 행위는 곧 좋음을 추구하게 만든다. 감정과 동기는 필연적으로 연결되어 있기 때문이다.

수정된 감수성 이론은 도덕 내용의 일치를 이룰 수 있다. 이상적 관찰자는 현실의 관찰자와 동일하게 진화한 기초 가치와 감정 반응을 가진 존재로 정의된다. 그렇다면 현실의 관찰자가 내리는 도덕 판단이 이상적 관찰자의 관점에서 설명되는 만큼, 둘의 판단이 일치할 거라 보는 건 자연스럽다.

이상적 관찰자를 도입함으로써 프린츠의 감수성 이론이 내재한 문제점은 해결된다. 이상적 관찰자는 직관에 잘 부합한다. 우리는 흔히 자신의 도덕 판단이 이상적 관점에서 나온 판단이거나 거의 근접하다고 생각한다. 이상적 관찰자는 현실의 관찰자와 독립된 별개의 실체가 아니다. 우리는 경험과 학습으로 도덕 외적 정보를 모으고, 편견과 관습에서 벗어나 자신의 관점을 객관화하려고 한다. 그제야 감정은 올바른 반응이 된다. 가상적인 관찰자를 상정하는 능력은 자연선택의 산물이다. 사회적 존재로서 인간은 매우 쉽게 타인의 눈을 소환하는 방법을 획득한다. 따라서 도덕 사실은 반응 의존적이며 진화는 도덕 판단을 참으로 정당화한다.

주요어 : 진화론, 도덕 실재론, 도덕 반실재론, 진화적 도덕 반실재론, 진화적 폭로 논증, 감수성 이론, 리처드 조이스, 제시 프린츠
학 번 : 2013-20201

목 차

1. 서론

메타윤리학은 도덕 판단이란 무엇인가, 도덕 판단은 어떻게 정당화되는가, 도덕 용어는 실질적으로 어떤 의미를 지니는가? 등 도덕의 기원과 본성을 묻는다. 여기 안락사는 도덕적으로 옳지 못한 행동이라고 믿는 사람이 있다. 그는 생명은 무엇보다도 소중한 가치이므로 아무리 삶이 고통스러워도 인위적 방법을 통한 죽음은 용납할 수 없다고 주장한다. 그러나 반대편에서는 삶을 지속할 가치가 없을 경우, 인간이 지닌 품위를 지키기 위해 스스로 생을 포기할 권리가 있다고 응수한다. 이때 두 사람이 벌이는 논쟁은 '어떻게 살아야 하는가'에 답하는 규범적 성격을 띤다. 한편 다른 사람이 다가와 이렇게 말한다. "당신들의 주장을 뒷받침하는 객관적인 도덕 사실이 있는가? 없다면 이 모든 말은 단지 의견에 불과하다." 지금 이 사람이 묻는 질문은 어떤 행동이 도덕적으로 올바르냐가 아니다. 특정 행동을 옳게 해주는 근거가 무엇인가이다. 이에 답해야만 도덕은 인간 삶을 규제하는 실천 원리로서 자리매김할 수 있다.

따라서 우리가 행하는 도덕 판단의 정당성을 보증하는 도덕 속성이나 그 속성을 나타내는 사실의 실재는 메타윤리학의 핵심 주제이다. 이를 '도덕 실재론'이라 한다. "살인은 도덕적으로 옳지 못하다."라는 판단은 사실을 진술하는가 혹은 감정을 표현하는가, 도덕 판단의 참과 거짓을 가릴 수 있는가, 진리를 판별하는 도덕 사실이 있다면 그 사실을 어떻게 알 수 있는가, 도덕 사실의 인식은 우리 행위를 인도하는가? 도덕 실재론자는 이러한 문제들에 답하면서 객관적인 도덕 사실이 있다고 주장한다.

"살인은 도덕적으로 그르다."라는 판단은 매우 자명해 보인다. 평상시 우리는 이 판단의 진실을 의심하지 않는다. 살인은 문화를 막론하고 보편적으로 옳지 못하다. 살인을 도덕적으로 용인하는 사람이 있다면 그를 같은 공동체의 일원으로 여기지 않으리라. 일상 경험은 도덕 사실의 존재를 믿도록 호소한다. 우리 마음 밖에 살인을 그르게 만드는 객관적인 도덕 사실이 있다. 그렇지 않고서 어떻게 우리가 모두 살인마를 비난하고, 죄 없이 희생당한 사람에게 공감할 수 있겠는가.

그러나 문제는 그렇게 간단하지 않다. 도덕 사실이나 속성은 어디에 있고 어떻게 알 수 있을까? 도덕과 무관한 속성, 이를테면 책상의 길이나 두께 등은 눈에 보이며 측정할 수 있다. 도덕과 무관한 속성들은 자연 세계 내에 위치하며 과학적인 방법으로 측정, 조작 가능하다. 반면 도덕적 옳음이나 그름 등의 도덕 속성은 눈에 보이지 않는다. 우리는 어떻게 도덕 속성을 자연주의적인 세계관 내에 수용할 수 있을까, 현대 과학의 방법으로 도덕 속성을 기술하고 정의할 수 있을까? 이는 도덕 실재론이 마주한 과업이다. 일부 도덕 실재론자는 도덕 속성을 자연 속성으로 설명한다. 한 예로 도덕적 그름이란 우리에게 고통을 일으키는 뇌의 특

정한 물리적 상태일 수 있다. 도덕적 그름이 고통과 동일하거나 환원 가능하다면 도덕적인 옳고 그름은 어떤 행동이 쾌나 불쾌를 산출하느냐와 연결된다. 이런 입장을 '도덕 자연주의'라 한다. 도덕 자연주의의 성패는 과연 그 자연 속성이 무엇이며, 우리가 날마다 반복하는 도덕 행동을 고무하는지에 달려있다.

상술한 물음 덕에 도덕 실재론 문제는 경험과학적 탐구에 열려있다. 심리학, 뇌신경과학, 비교문화인류학, 진화생물학, 진화심리학, 실험경제학 등은 도덕 판단을 할 때 실제로 무슨 일이 일어나는지 밝힘으로써 도덕 실재론 논쟁의 경험적 근거를 제공한다. 도덕 사실을 신비로운 존재로 만들지 않으려면 이러한 연구 분야의 자료들을 적절히 활용해야 한다.

그 중에서도 다윈(C. Darwin)의 진화론은 도덕의 본성을 자연화하는 유망한 방법이다. 진화론은 한 종의 개체가 나타내는 정교한 신체 구조와 심리 기제 등이 왜 발생했는지 묻는다. 개체가 보유한 몇몇 형질은 수많은 시간을 거쳐 자연선택으로 설계된 적응이다. 적응적 특징들은 그 종의 선조들이 지속적으로 맞닥뜨린, 생존과 번식에 관련된 문제들을 효과적으로 해결했다. 인간 종이 발휘하는 보편적 행동인 도덕 또한 예외가 아니다. 도덕이 왜 있는지 알려면 그 진화적 기능을 물어야 한다. 과학적 사실로서 진화는 도덕의 기원을 밝힘과 동시에 도덕 판단의 정당성까지 확보해 주리라 기대된다. 시작을 알면 존재 이유도 알 수 있으리라. 그렇다면 먼저 진화론이 도덕의 기원을 어떻게 설명하는지 간단히 살펴보자.

다윈은 인간의 도덕성이 자연선택으로 진화한 본능이라고 생각했다. 다윈이 제시한 근거는 다음과 같다. 첫째, 인간을 포함한 많은 동물은 집단의 이익을 고려하는 사회적 본능을 지닌다. 둘째, 인간의 사회적 본능을 구성하는 사랑, 공감, 동정심 등의 감정은 보편적 기질이다. 셋째, 특정한 상황에서 인간은 전형적인 도덕 판단을 한다. 넷째, 인간의 도덕 판단은 즉각적인 반응인 경우가 흔하다. 다윈은 강한 사회적 본능을 가진 인간들이 모인 집단은 다른 집단보다 더 번성하여 이러한 습성이 후세대로 유전되었다고 추측했다. "높은 수준의 애국심, 충실성, 복종심, 용기, 동정심이 있어서 남을 도울 준비가 항상 되어 있고 공동의 이익을 위해 자신을 희생할 준비가 되어 있는 사람들이 많은 부족은 다른 부족에 비해 성공을 거둘 것이다. 이것이 바로 자연선택이다(Darwin 1998/2006: 215)." 그리하여 다윈은 마찬가지, 자연선택으로 발달한 인간의 지적 능력과 사회적 본능이 결합하여 인간만이 가진 특유의 도덕감(moral sense), 양심이 생겨났다고 주장했다.[1)]

하지만 도덕의 기원에 대한 다윈의 설명은 일정한 한계가 있다. 먼저 다윈이 제시한 근거는 문헌 연구와 직접 관찰에만 의존하여 다소 추정적이다. 또한 다윈은 사회적 본능이 집단을 위해 개체가 스스로를 희생하는 집단선택 과정으로 획득된다고 보았다. 그러나 집단선택은 집단 내에 혼자만 이득을 취하는 이기주의자가 한 명만 있어도 쉽게 무너질 위험이 있다.

---

1) 도덕의 기원에 대한 진화론적 설명과 그 윤리학적 함의의 자세한 내용은 김성한(2001)을 참고하라.

다윈이 해결하지 못한 어려움은 집단유전학과 진화생물학의 근대적 종합으로 새로운 전기를 맞는다. 멘델의 입자 유전에 기반을 둔 집단유전학은 자연선택의 주요한 기제인 유전 현상이 유전자라는 물리적 단위로 이루어짐을 밝힘으로써 진화를 바라보는 새로운 시각을 제공했다.

이제 진화 과정을 '유전자의 관점'으로 보는 신다윈주의자들은 사회생물학, 진화심리학이라는 이름 아래 도덕의 진화에 대한 더욱 엄밀한 그림을 제시했다. 소규모 공동체에서 살았던 조상들은 환경의 갖가지 위협으로부터 자신과 가족을 보호해야 했다. 포식자의 공격 방어하기, 성공적인 사냥을 위한 좋은 동료 구하기, 음식물 나누기, 짝을 얻고 지키기 등은 생존과 직결되는 문제였다. 이러한 상황에서 타인과 세계를 도덕적으로 판단하는 경향은 자신의 유전자 사본을 효과적으로 전달하게 도왔으리라. 1960년대 이후 다윈의 수수께끼를 설명하는 구체적인 모델이 제시되었고 진화생물학은 혁명적으로 진보했다.

해밀턴(W. Hamilton)의 '포괄 적합도 이론(inclusive fitness theory)'과 트리버스(R. Trivers)의 '호혜적 이타주의 이론(reciprocal altruism theory)'은 인간을 포함한 동물의 협동 행동과 이타성의 진화를 단순하고 명쾌하게 설명했다. 포괄 적합도 이론의 요지는 이렇다. 개체는 자신의 직접적 적합도[2]와 함께 유전자를 공유하는 다른 혈연의 간접적 적합도까지 고려한다. 따라서 내가 직접 자손을 낳는가와 상관없이 혈연관계인 사람을 도와 내 유전자가 전달된다면 혈연에게 사랑과 자원을 베푸는 이타적 행동이 진화할 수 있다. 타인에게 헌신하는 행동이 많은 비용을 요해도 유전자의 관점에서는 희생이 아니다. 나와 유전자를 공유한 사람의 번식 성공은 결국 내 유전자의 번식 성공이기 때문이다. 이러한 이유로 포괄 적합도 이론은 '혈연선택 이론(kin selection theory)'으로도 불린다.[3] 그러나 한 가지 난점이 있다. 포괄 적합도 이론은 부모와 자식 간의 사랑, 친지들에게 베푸는 원조를 잘 설명하지만 인간은 자기와 혈연관계가 아닌 사람과도 협동한다. 이 문제는 호혜적 이타주의가 해결한다.

호혜적 이타주의는 상호 부조이다. 내가 타인에게 전달한 이득이 미래의 어느 시점에 되돌아온다면 비친족 간의 이타적 행동은 진화할 수 있다. 호혜적 이타주의는 이타적 형질 T를 위한 어떤 유전자 G가 있다고 할 때 유전자 G는 이득을 받는 수혜자가 미래에 그 이익을 되갚을 경우, 오직 그러한 경우에만 선택된다. 트리버스에 따르면 호혜적 이타주의의 진화에는 세 가지 조건이 필요하다. 첫째, 두 명의 개체가 호혜적 상호작용에 참여해야하며 이타적 행동이 초래하는 비용은 호혜자가 돌려받는 이득보다 적어야 한다. 둘째, 호혜적인 거래는 과거에 호의를 갚지 않은 사기꾼이 있을 때 철회된다. 셋째, 개체들은 반복적으로 상호작용하여

2) '적합도(fitness)'라는 개념은 논자마다 약간의 차이가 있지만 간단히 말해 개체군 내의 다른 구성원들과 비교하여 한 개체가 자신의 유전자를 다음 세대에 얼마나 성공적으로 전달하는가를 의미한다. 주로 자손의 생산으로 이를 측정한다.

3) 포괄 적합도 개념이 등장하는 해밀턴의 논문은 수학적 내용이 많아 그 의의를 파악하기 쉽지 않다. 대중적인 입문서로 Dawkins(1976)를 참고하라.

사기를 치는 행위보다 호혜적 관계를 유지하는 이익이 훨씬 더 크다. 우리 조상들의 진화 역사에서 이 세 조건을 만족하는 사회적 협력은 빈번하게 일어났다(Trivers 1971).[4] 결론적으로 도덕은 사회적 동물인 인간의 생존과 번식 성공을 좌우하는 적응 형질로서 진화했다.

도덕의 기원에 대한 진화적 설명은 도덕 판단의 정당성 또한 확보하는가? 진화론을 받아들인 초창기의 철학자들은 그렇다고 생각했다. 허버트 스펜서(H. Spencer)는 자연선택의 발견은 곧 도덕의 발견이며 진화가 보증하는 자연 사실이 도덕적 좋음이나 옳음과 동일하다고 주장했다(H. Spencer 1879/2004). 진화는 인간 종의 사회적 조화를 이끈다. 사회적 조화는 인간 삶을 더 탁월하고 행복하게 만든다. 따라서 사회적 조화를 목적으로 하는 행위 원리는 그 자체로 옳다. 스펜서의 견해는 진화에 대한 오해와 심각한 철학적 난점('자연주의의 오류'라 불리는 이 문제는 뒤에서 자세히 다루겠다)이 있지만 일견 상식 도덕에 호응하는 면이 있다.

다른 한편, 진화가 도덕 판단의 정당성을 보증하기는커녕 되레 객관적인 도덕 사실의 존재를 훼손한다고 주장하는 논자들이 있다. 이들은 도덕이라는 현상이 한낱 허구로 꾸며진 신화에 지나지 않는다고 말한다. 왜 그런가? 우리는 보통 어떤 사물이 실재한다고 말할 때 그 사물이 인식 주관과 독립적이라고 기대한다. 그런데 도덕이 진화의 산물이라면 도덕 사실은 우리 인식과 무관할 수 없다. 영아살해를 나쁘다고 판단하는 이유는 무엇인가? 영아살해를 나쁘게 만드는 사실이 우리 밖에 있어서가 아니다. 단지 자연선택이 그러한 행동을 나쁘게 판단하도록 만들었을 뿐이다. 진화는 개체의 생존과 번식에 도움이 되면 외부 사실과 상관없이 개체를 기만할 수 있다. 요컨대 자연선택의 목적은 개체의 유전자 사본을 복제하는 데 있지 세계의 사태를 반영하는 데 있지 않다. 따라서 이들은 진화가 도덕적 참이라는 허구를 폭로하는 강력한 무기라고 주장한다.

이 글은 도덕의 기원에서 진화가 수행한 역할을 받아들인다. 또한 진화론을 받아들이면 도덕 사실이 우리의 평가적 태도와 독립적일 수 없다는 전제도 인정한다. 이런 바탕에서 이글이 답하려는 핵심 질문은 다음과 같다. 왜 도덕 사실은 우리 마음과 독립적이어야 하는가, 왜 진화론은 도덕 판단을 정당화해 줄 수 없는가? 결론적으로 말해 진화는 도덕 판단의 참을 보증할 수 있으며 도덕 사실이 반드시 마음 독립적일 필요는 없다. 따라서 이 글은 진화가 객

4) 생물학자들이 사용하는 이타주의 개념은 혼동의 여지가 있다. 엄밀히 말해 이타주의는 행동의 결과가 아니라 개체의 동기와 관련된다. 어떤 행위가 이타적이라면 첫째, 행위자가 특정 동기를 갖고 둘째, 그 동기가 타인을 향한 마음이어야 한다. 따라서 행위자의 동기가 이타적이라면 어떤 행위가 타인을 돕지 못했더라도 이타적이다. 반면 행위자의 동기가 이기적이라면 어떤 행위가 타인을 도왔더라도 이기적이다. 혈연선택 이론과 호혜적 이타주의 이론은 심리적 이타주의가 아니라 타인을 도와 적합도를 증진하는 결과를 낳는 협동 행동이 어떻게 진화되었느냐에 방점이 있다. 그럼에도 혈연선택 이론과 호혜적 이타주의 이론은 자연선택이 친족과 비친족에 대한 협동을 규제하는 근접 기제로서 이타적 동기와 사랑과 혐오 등의 감정을 선호했고, 이는 도덕적 동기와 행동이 진화하는 데 중요하게 작용했음을 아울러 설명한다. 사회생물학과 진화심리학의 태동 이후 전개된 진화윤리학의 주요 개념과 쟁점은 James(2011)을 참고하라.

관적인 도덕 속성이나 사실의 토대를 훼손한다고 주장하는 진화적 도덕 반실재론을 비판하고, 진화를 사용하여 우리 판단에 의존하면서도 객관적인 도덕 사실을 구하려 한다.

주된 비판의 대상은 리처드 조이스(R. Joyce)가 제기한 진화적 도덕 반실재론이다. 조이스는 매우 적극적인 도덕 반실재론자이다. 그는 『도덕의 진화(*The Evolution of Morality*)』에서 도덕이 자연선택의 산물이라면 객관적인 도덕 사실의 인식과 존재 가능성은 정당화되지 못한다고 주장한다. 도덕의 진화적 기원은 도덕 판단의 정당성을 훼손한다. 도덕의 진화는 세계에 존재하는(또는 존재한다고 생각되는) 도덕 사실과 무관한 과정이다. 따라서 도덕 판단의 참을 보증하는 마음 독립적인 사실이 실재하는지 알 수 없다. 이를 '진화적 폭로 논증(evolutionary debunking arguments)'이라 한다. 위에서도 썼듯이, 폭로 논증은 진화적 도덕 반실재론자가 사용하는 일반적인 전략이다. 조이스는 한 걸음 더 나간다. 그는 진화한 도덕의 본성이 도덕 사실을 자연 사실로 설명하려는 모든 도덕 자연주의의 시도를 무력하게 만든다고 말한다.[5] 조이스가 보기에 도덕 판단의 핵심은 불가피한 권위에 있다. 이는 도덕 판단이 사적 욕구나 득실과 무관하게 따라야 할 정언적 이유에 대한 믿음을 표현한다는 뜻이다. 도덕 사실이란 게 있다면, 그러한 존재는 정언적 이유를 주어야 한다. 한데 도덕 판단의 정언성은 자연선택의 소산이다. 정언적 이유를 주는 사실은 없다. 정언적 이유가 없다면 도덕 사실은 없다. 반면 도덕 자연주의는 가언적인 체계이다. 이는 도덕 판단에 행위자의 욕구가 반드시 필요하다는 뜻이다. 적절한 욕구 없이 행위를 위한 이유는 생기지 않는다. 그렇다면 어떤 도덕 자연주의도 도덕의 정언성을 수용할 수 없다. 정언성 없는 도덕 자연주의는 거짓이다. 도덕 사실은 없다. 이처럼 정언성을 무기로 도덕 자연주의를 공격하는 조이스의 논증은 해결해야 할 중대한 도전이다.

프린츠(J. J. Prinz)는 『도덕의 감정적 구성(*The Emotional Construction of Morals*)』에서 도덕 판단을 특정한 반응과 결부지어 설명하는 '감수성 이론(sensibility theory)'[6]이 도덕 사실의 존재를 구할 수 있다고 반박한다. 도덕 사실이란 우리 마음과 독립적인 불가사의한 무엇이 아니다. 도덕은 인간 종만이 가진 독특한 행동 양식으로 우리 반응과 밀접하게 연결된다. 도덕 판단을 정당화하는 도덕 사실은 반응의 구성물이다. 프린츠가 상정하는 그 반응은

---

5) 여기서 두 가지 의미의 자연주의를 구분해야 한다. 하나는 초자연주의와 반대되는 의미로서 이 세계와 세계 내의 모든 존재자는 자연과학의 법칙들에 따른다는 입장이다. 따라서 세계의 모든 사실들은 자연의 사실이다. 이를 '포괄적 자연주의'라 한다. 다른 하나는 도덕 사실은 자연 사실로 설명될 수 있다는 입장이다. 이를 '도덕 자연주의'라 한다. 포괄적 자연주의는 도덕 자연주의를 포함하는 일반적 개념이다. 포괄적 자연주의자는 도덕 자연주의를 거부할 수 있다. 도덕을 자연과학적 방법으로 연구하면서 어떠한 도덕 사실도 존재하지 않는다는 결론이 가능하기 때문이다. 종교를 생각해보라. 종교를 과학적으로 탐구하면서 어떠한 종교적 사실도 없다고 말할 수 있다. 조이스가 바로 그런 경우이다. 조이스는 도덕을 진화론으로 설명하지만 진화가 도덕 사실의 실재를 논박한다고 본다. 조이스는 도덕 자연주의를 부정하는 포괄적 자연주의자이자 도덕 회의주의자이다.

6) 'sensibility theory'의 번역어인 '감수성 이론'은 윤화영(2009)을 따랐다.

감정을 느끼는 성향이다. 도덕 판단은 무엇보다도 감정 표현이다. 우리는 비도덕적인 행위를 보며 분노하고 분노는 그 행위를 금지하도록 추동한다. 현대 도덕 심리학의 경험 연구는 도덕 판단과 감정의 연결을 강하게 지지한다. 감정은 괜히 생기지 않았다. 감정은 개체의 생존과 번식에 도움이 되는 행위를 일으키는 진화적 적응이다. 예를 들어 혐오는 상한 음식이나 독을 가진 포식자를 회피하는 반응을 만들어 개체를 보호한다. 이렇게 감정은 우리가 외부 세계에 잘 대응하도록 정보를 주는 신뢰할 만한 탐지 장치이다. 어떠한 감정 없이 행위 이유나 동기는 발생하지 않는다. 하지만 프린츠는 도덕이 완전한 진화의 산물임은 부정한다. 도덕은 선천적인 감정과 다양한 문화가 조합된 독특한 규범 체계이다. 한 지역에서 부정적인 감정을 느끼는 행동을, 다른 지역에서는 용인할 수도 있다. 그렇더라도 감정에 기반을 둔 감수성 이론은 생물학적 사실과 도덕 판단이 무관하지 않음을 보여준다.

프린츠는 자신의 감수성 이론으로 조이스의 진화적 폭로 논증에 대항했고(Prinz 2008a), 조이스는 프린츠의 감수성 이론이 도덕 사실의 실재를 보증하는 좋은 대안이 되지 못한다며 비판한 바 있다(Joyce 2008; 2009). 두 사람이 벌인 대결은 흥미로운 대척점에 서 있다. 우리는 그들이 전개하는 이론을 비교 검토함으로써 진화론에 근거한 도덕 실재론 논쟁의 진면목을 이해하고 그 해결 가능성을 탐색할 수 있다. 이 글은 도덕 사실의 실재를 규명하는데 감수성 이론이 유효하다고 생각한다. 이 글은 조이스의 비판을 수용하고 프린츠의 이론과 조이스의 이론을 절충하여 감수성 이론을 택하더라도 객관적인 도덕 사실의 실재를 구할 수 있다고 주장하고자 한다. 그리하여 이 글은 도덕 판단의 참을 정당화하는 '수정된 감수성 이론(modified sensibility theory)'을 제안하고자 한다.

이를 위해 2장에서는 먼저 고전적인 메타윤리학의 도덕 실재론 논쟁과 조이스 이전, 초기 진화적 도덕 반실재론을 둘러싼 주요 개념과 쟁점들을 일별한다. 초기 진화적 도덕 반실재론을 대표하는 주요 논자는 마이클 루즈(M. Ruse)이다. 루즈는 조이스와 마찬가지로 도덕 판단을 정당화하는 사실은 자연선택이 만든 환상이라고 주장한다(조이스는 루즈의 폭로 논증을 계승하면서 나아가 도덕 자연주의 전체를 공박한다). 루즈에 맞서 초기 진화적 도덕 반실재론에 대응한 선행 연구도 살펴본다. 이로써 앞으로의 논의 전개를 이해하는 기초를 다지고 답해야 할 문제가 무엇인지 명확히 할 수 있다.

3장에서는 본격적으로 조이스와 프린츠의 메타윤리 이론을 비교한다. 두 사람의 주장은 세 가지 측면에서 극명하게 대립한다. 첫째는 도덕 판단의 본성이다. 조이스는 도덕 판단이 행위 이유에 대한 믿음을 표현한다는 인지적 성격을, 프린츠는 도덕 판단이 감정적 성향을 표현한다는 비인지적 성격을 강조한다. 둘째는 도덕 사실의 성격과 실재의 유무이다. 조이스는 진화가 마음 독립적인 도덕 사실의 인식 가능성을 훼손한다는 진화적 폭로 논증을, 프린츠는 도덕 사실이 감정 반응으로 구성된다는 감수성 이론을 주장한다. 나아가 조이스의 진화적 폭로

논증은 도덕 판단의 객관성과 실천적 힘이 유용한 허구라고 말한다. 도덕 판단은 거짓이지만 우리 삶에 도움이 되기 때문에 믿는다. 프린츠의 감수성 이론은 급진적이다. 감정은 도덕 행위를 위한 동기를 주지만 어떤 사안에 대한 행위자의 감정은 개인 또는 공동체마다 다를 수 있다. 프린츠는 이를 부정하지 않으며 도덕의 객관성을 부정하고 실재론에 바탕을 둔 도덕 상대주의를 옹호한다. 셋째는 도덕 선천주의 대 비선천주의이다. 조이스는 도덕 판단을 하는 경향이 진화적 적응이라는 선천주의를, 프린츠는 도덕 판단은 생득적인 감정과 문화가 결합하여 생긴 부산물이라는 비선천주의를 지지한다. 이 세 가지는 서로 별개로 다루어야 할 입장이 아니다. 각각의 논점은 두 사람의 반실재론적, 실재론적 입장과 긴밀히 연결되어 있다. 충실한 도덕 실재론을 짜기 위해서는 세 가지 주제가 함께 맞물려야 한다.

4장에서는 감수성 이론에 대한 조이스의 비판과 그 대응 방안을 모색한다. 주요 논점은 다음과 같다. 첫째, 도덕 판단의 객관성을 어떻게 확보하는가? 둘째, 도덕 판단의 정언적 성격을 어떻게 설명하는가? 셋째, 감수성 이론이 그르다고 판단하는 행동이 상식 도덕관이 그르다고 판단하는 행동과 일치하는가? 감수성 이론이 처한 이 같은 어려움은 수정된 감수성 이론으로 극복할 수 있다. 수정된 감수성 이론이 노정하는 핵심적인 대응 논변은 다음과 같다. 첫째, 인간에게는 보편적인 도덕 감정을 일으키는 진화한 기초 가치가 있다. 둘째, 이상적 관찰자의 관점에서 기초 가치 덕에 감정 반응이 수렴됨을 옹호한다. 그럼 이제 논의를 시작해 보자.

2. 도덕 실재론 논쟁과 초기 진화적 도덕 반실재론 고찰

도덕 실재론 논쟁의 역사는 깊고 넓어 그 면모를 전부 파악하기란 불가능하고 이 글의 목적도 아니다. 여기서는 이 글이 다루는 주제를 이해하는 데 필요한 기본적인 개념과 쟁점을 중심으로 서술한다. 편의상 도덕 실재론과 도덕 반실재론 사이의 대립을 '도덕 실재론 논쟁'이라고 부르겠다. 그 뒤 전통적인 도덕 반실재론의 영향을 받은 마이클 루즈의 초기 진화적 도덕 반실재론을 분석한다. 루즈는 진화적 도덕 반실재론자들이 사용하는 일반적인 전략을 정초했다. 조이스는 루즈의 논의를 발전적으로 계승하여 도덕 자연주의의 기획을 무너뜨리려 한다. 마지막으로 루즈의 진화적 도덕 반실재론에 대응한 선행 연구들을 살펴봄으로써 도덕 실재론 문제를 해결하는 단초를 찾는다.

2.1. 도덕 실재론 논쟁의 주요 개념과 쟁점

도덕 실재론자는 객관적인 도덕 사실이 존재한다고 생각한다. 도덕 사실이 있다는 형이상학적 입장은 특정한 인식론적 물음과 연결된다. 우리가 도덕 판단을 할 때 어떤 심성 상태가 나타나는가? 이때 도덕 실재론자는 "살인은 그르다."라는 판단이 믿음을 표현한다고 말한다. 믿음은 세계가 어떠한가라는 사태를 반영한다. 그러므로 참이거나 거짓일 수 있으며 서술문의 형태로 발화되는 주장이다. 즉 도덕 판단은 도덕 사실의 존재를 믿음으로 표상하며 그에 비추어 판단의 참과 거짓을 가릴 수 있다. 이렇게 도덕 판단이 진리치를 가진다는 입장을 '인지주의(cognitivism)'라 부른다. 도덕 실재론자는 모두 인지주의적 관점을 공유한다(Sayre-McCord 2006: 40).

도덕 반실재론자는 객관적인 도덕 사실은 존재하지 않는다고 생각한다. 일부 도덕 반실재론자들은 도덕 판단이 진리치를 가진다는 입장을 거부한다. 그들에게 도덕 판단은 믿음이 아니라 세계가 어떠해야 하느냐라는 화자의 주관적 욕구를 표현하는 데 불과하다. 욕구는 참이거나 거짓일 수 없으며 비서술문의 형태로 발화된다. 이러한 입장을 '비인지주의(noncognitivism)'라 부른다. 대표적인 비인지주의 도덕 반실재론자는 에이어(A. J. Ayer)이다. 비인지주의 도덕 반실재론이 태동하는데 영향을 끼친 철학적 배경은 현대 도덕 실재론 논쟁을 이해하는 단초다. 이를 살펴보자.

19세기와 20세기 초 흄(D. Hume)과 무어(G. E. Moore)는 도덕 언어에서 사실과 가치를 엄격히 구분해야 하며 사실로부터 가치를 도출하려는 어떠한 자연주의적 시도도 오류임을 주장했다. 먼저 흄은 당대의 도덕 논증이 '~이다'와 '~아니다'와 같은 사실 진술로부터 갑자기 '~해야 한다'나 '~해서는 안 된다'라는 당위 진술로 뛰어 넘는 경향이 있음을 발견했다. 흄은

이런 추론이 논리적으로 타당하지 않다고 생각했다. 전제에 없는 진술이 결론에 나올 수는 없기 때문이다(Hume 1980/1998: 40). 즉 도덕과 무관한 명제로부터 도덕에 관한 명제가 따라 나오는 일은 논리적으로 불가능하다. 흄의 이러한 통찰은 '흄의 법칙'으로 불리며 "존재로부터 당위를 도출하지 말라."는 명령으로 곧잘 사용한다.

무어는 좋음이라는 속성을 다른 자연 속성으로 정의하려는 어떤 시도도 실패한다고 주장했다. 무어가 제시한 근거는 '열린 질문 논증(open question argument)'이다. 어떤 속성 P가 자연 속성 N으로 정의된다면 "P는 N인가?"라는 물음은 동어반복이며 닫힌 질문이다. "총각은 결혼하지 않은 남자이다."를 떠올려보라. 이 물음은 더 이상의 질문이 필요하지 않은, 언제나 참인 진술이다. 따라서 어떤 대상의 속성을 아주 기본적인 용어로 정의하고자 할 때 열린 질문 논증을 적용해 볼 수 있다. 정의대상과 정의속성이 동일하면 질문은 끝난다.

그런데 무어가 보기에 도덕 속성을 정의하는 질문은 영원히 계속된다. 스펜서를 좇아 좋음이 사회적 조화라는 속성과 동일하다고 가정해보자. 그렇다면 "사회적 조화는 좋은가?" 이 질문은 동어반복이 아니다. 좋음이 사회적 조화와 동일한 속성이냐고 묻는 질문은 타당하며 의미 있는 물음이다. 따라서 이 질문은 열려 있다. 무어는 좋음을 정의하는 속성을 아무리 많이 나열한들 결코 닫힌 물음이 될 수 없다고 주장했다. 좋음은 정의될 수 없다. "좋음은 좋음이며, 그게 전부다(Moore 1903: 6)." 좋음은 정의될 수 없기에 좋음을 자연 속성으로 정의하려는 자연주의의 기획은 헛된 일이다. 이를 '자연주의의 오류'라 한다.

비인지주의자들은 흄과 무어가 제기한 존재-당위의 엄격한 구분과 자연주의의 오류를 받아들였다.[7] 더하여 비인지주의자들은 논리실증주의의 철학적 신념에 크게 영향 받았다. 논리실증주의는 철학의 문제는 경험적으로 검증 가능한 유의미한 명제를 분석하는 일에 국한돼야 한다고 주장했다. 삶의 목적이나 신 존재 증명과 같은 논의들은 검증 불가능한 무의미한 명제들로 이루어져 있으며 마땅히 철학에서 제거해야 할 사이비 문제다. 논리실증주의자에게 진리치를 가질 수 있는 명제는 수학과 논리학의 분석명제, 관찰 가능한 과학의 종합명제에 한정된다. 윤리학의 명제는 그에 대응하는 사실을 발견할 수 없으므로 무의미하다.

따라서 에이어를 위시한 비인지주의자들에게 도덕 영역은 사실과 철저히 구분되는 당위에 속하며 경험적 용어로 환원될 수 없다. 그렇다면 도덕 판단은 어떤 기능을 하는가? 에이어는 도덕적 발화는 단지 감정을 표현한다고 주장했다. 에이어의 설명을 보자.

> 이제 내가 지금까지 말한 것을 일반화해서 설명하면, "돈을 훔치는 것은 옳지 않다."고 말할 때 나는 어떤 사실적 의미도 갖지 않는 문장을 말하는 것이고 참이거나 거짓인 어떤 명제도 표현하

7) 그러나 무어는 비자연주의적인 도덕 속성이 실재하며 우리 직관으로 이를 알 수 있다고 생각하는 도덕 실재론자이다. 따라서 비인지주의자들은 자연주의에 대한 무어의 비판은 받아들이지만 비자연적인 도덕 속성의 존재는 거부한다.

> 고 있지 않다. 그것은 마치 "돈을 훔치다니!"라고 쓰는 것과 같고, 거기에서 느낌표의 형태와 강조는 적절한 규약에 의해서 특별한 종류의 도덕적 불찬성의 감정이 표현되고 있음을 보여주는 것이다. 여기에 참이거나 거짓이라고 말할 수 있는 것은 분명히 아무것도 없다(Ayer 1946/2006: 162).

"돈을 훔치다니!"로 해석되는 도덕적 발화는 세계의 사태를 기술하지 않는다. 다만 특정 상황에 대한 나의 감정을 표현하고 있을 뿐이다. 감정은 세계가 어떠해야 하느냐에 대한 욕구나 태도를 표현한다. 그래서 도덕 판단을 할 때 외부 세계에 대한 객관적 지식은 필요하지 않다. 에이어의 이런 입장을 '정서주의(emotivism)'라 부른다. 정서주의는 비인지주의를 대표하는 이론이다.

도덕 판단이 단지 감정 표현에 지나지 않는다고 해서 비인지주의가 곧바로 도덕 주관주의 혹은 상대주의로 귀결되지는 않는다. 도덕 주관주의는 "X는 옳다."는 판단은 "X는 나에게 옳다."는 판단과 같다고 주장한다. 도덕 상대주의는 특정 사회, 문화, 관습에 따라 무엇이 옳은지는 상대적이라고 주장한다. 그렇다면 도덕 주관주의와 상대주의는 도덕적 불일치 문제를 설명할 수 없다. 우리는 종종 도덕 문제에 대해 의견이 갈린다. 우리는 특정 판단을 두고 갈등하며 논쟁을 거쳐 합의에 이르기도 한다. 그러나 도덕 주관주의와 상대주의에 따르면 나도 옳고 너도 옳다. 따라서 도덕적 불일치도 없다. 문제는 도덕 주관주의와 상대주의 둘 다 도덕 언어가 진리치를 가졌다고 전제하기 때문에 생긴다. 각자의 생각이 모두 참을 반영하므로 어떤 불일치도 원천적으로 불가능하다. 도덕 언어의 진리치를 부정하는 비인지주의는 도덕적 불일치를 설명할 수 있다. 불일치는 느낌이나 태도의 불일치이다(Stevenson 1937). 이때 행위자는 다른 태도보다 더 나은 태도가 있다고 생각하며 사람들이 자신의 태도에 동조하게끔 설득한다. 도덕적 발화는 상대방의 태도를 변경하도록 요구한다. 불일치를 해결하려고 노력하는 비인지주의는 상대주의가 아니다.

비인지주의는 도덕 판단과 동기에 대한 특정한 관점과 연결된다. 우리는 보통 도덕 판단이 그 판단을 실행하려는 동기를 동반한다고 생각한다. "그 일을 해서는 안 돼!"라고 판단한다면 그 일을 하지 않는 행위를 정당화하는 이유와 그 일을 하지 않으려는 동기를 갖는다. 이렇게 도덕 판단과 동기의 필연적 관계를 인정하는 입장을 '동기 내재주의(motivational internalism)'라 부른다. 반면 도덕 판단과 동기가 맺는 관계를 우연적이라 보는 입장도 있다. 이를 '동기 외재주의(motivational externalism)'라 부른다. 옳다고 판단한 일을 실천에 옮기지 않는 사람도 있다. 도덕 동기는 판단 외부에 있는 여러 요인들에 영향 받을 수 있다.

동기 외재주의는 도덕 판단의 실천적 권위와 자율성을 해친다. 도덕적으로 행동하도록 동기화되는 현상은 도덕 판단을 다른 판단과 구분해주는 매우 중요한 요소이다. 내재주의는 우리 직관에 잘 들어맞는다. 그래서 내재주의자는 외재주의자가 말하는 도덕 판단을 진정한 도

덕 판단으로 보지 않거나, 행위자가 어떤 이유로 실천적 합리성이 결여되어 동기를 갖지 못했다고 주장한다.

비인지주의자는 내재주의자로서 도덕의 실천적 본성을 매우 잘 설명한다. 감정, 느낌, 욕구와 같은 비인지적 심성 상태는 직접적으로 동기를 일으킨다. 욕구는 그 자체 동기적 상태이다. "살인은 나쁘다."라는 판단은 "살인이라니!", "살인 우우!"와 같은 경멸과 분개의 감정을 나타내며 그럼으로써 나 혹은 타인이 살인을 하지 않도록 요구한다. 비인지주의자에게 중요한 도덕의 핵심은 세계의 참모습을 보고하는 게 아니라 우리를 움직이는 실천적 힘이다.

그러나 비인지주의에는 약점이 있다. 도덕의 실천적 성격 못지않게 도덕이 객관적 지식이라는 인지주의의 주장도 직관에 잘 부합한다. 우리는 흔히 도덕적 질문에 자신의 평가적 태도와 독립적인 올바른 답변이 있다고 생각한다. 이런 답변은 객관적인 도덕 사실 덕분에 옳게 보인다. 우리는 때로 도덕적 고찰 덕에 이전의 의견을 고치고 도덕적 관점에 대한 합의에 이르기 때문이다(Smith 1993/2006: 14-15). 비인지주의는 도덕 판단의 객관성을 설명할 수 없다.

또한 비인지주의의 주장대로 도덕 판단이 참이거나 거짓일 수 없다면 도덕적 진술을 포함한 복합문장이나 논증을 분석할 수 없다. 예를 들어보자. '아동학대는 나쁘다. 아동학대가 나쁘다면 어린이집 교사의 아동학대는 나쁘다. 그러므로 어린이집 교사의 아동학대는 나쁘다.' 이 논증은 타당하다. 전제가 참이라면 결론도 반드시 참이다. 하나 비인지주의에 따르면 첫 번째 전제는 "아동학대라니!"로 해석되어 진리치가 사라진다. 그렇다면 이 논증은 타당한 논증이 아니게 된다. 논증에 사용된 문장은 진리치가 있어야 하기 때문이다. 또한 두 번째 전제에 사용된 조건문, '아동학대가 나쁘다면'은 "아동학대라니!"로 해석될 수 없다. 조건문을 감정 표현으로 치환하면 첫 번째 전제와 두 번째 전제에 등장하는 '나쁨'의 의미가 다르다는 애매어의 오류를 범하고 만다. 그러나 이 논증은 타당하다. 첫 번째 전제와 두 번째 전제에 사용된 '나쁨'의 의미는 동일하다(Geach 1965/1972). 기치(P. Geach)는 어떤 문장의 의미는 단독으로 사용되건 복합문장의 일부분으로 사용되건 늘 동일하다는 프레게(G. Frege) 논제로 비인지주의가 맞이한 이 같은 맹점을 비판했다. 일상적인 도덕적 담화나 논증에서 도덕 판단을 구성하는 진술은 동일하게 사용된다. 이를 '프레게-기치 문제'라 부른다. 프레게-기치 문제는 도덕적 진술에 참과 거짓을 가릴 수 있는 인지적 요소가 있음을 보여준다.

모든 반실재론자가 비인지주의자는 아니다. 인지주의 도덕 반실재론자는 비인지주의가 처한 어려움을 피하려 도덕 판단에 진리치가 있음을 인정한다. 철학자 맥키(J. L. Mackie)가 대표적이다. 맥키는 자신의 도덕 반실재론을 '오류 이론(error theory)'이라 칭한다. 도덕 판단은 세계에 대한 믿음을 나타내며 참과 거짓을 가릴 수 있다. 그러나 도덕 판단의 참을 보증하는 객관적인 도덕 사실이나 속성은 없다. 모든 도덕 믿음은 거짓이다. 도덕 영역에서 우리

는 체계적인 오류에 빠져있다(Mackie 1977/1990). 맥키의 오류 이론은 진화적 도덕 반실재론과 매우 깊은 관련을 맺는다. 이에 대해 다음 절에서 자세히 다루겠다.

비인지주의가 처한 어려움에서 알 수 있듯이, 객관적인 도덕 사실이 있다는 믿음은 쉽게 논박되지 않는다. 도덕 실재론은 보통 사람들이 품은 직관에 강력히 뿌리내리고 있다. 그래서 오랜 옛날부터 철학자들은 도덕 사실이 무엇이고 어디에 있는지 설명하려고 노력했다. 플라톤은 '좋음의 이데아'라는 초월적 속성이 도덕적 올바름을 이끈다고 생각했다. 현실은 이데아의 그림자일 뿐이다. 한데 이데아는 눈에 보이지 않는다. 좋음의 이데아를 향해 올라가려면 사물 너머를 보려는 끊임없는 훈련과 관조가 필요하다. 플라톤은 철학의 임무가 이데아의 모조품인 일상 세계로부터 사람들의 눈을 위로 올리는 데 있다고 말했다.

오늘날 도덕 실재론자들은 더 이상 초월적 속성에 의지하지 않는다. 자연 세계를 넘어선 존재자에 신경 쓰지 않는 실재론자들은 좀 더 엄밀하고 합리적인 방법으로, 자연과학의 방법으로 도덕 속성이나 사실의 존재를 구하려고 노력한다.

도덕 자연주의는 도덕 실재론의 한 방법이다. 일반적으로 도덕 자연주의는 환원적인 원칙이다. 즉 도덕 자연주의는 도덕적인 사실이나 속성을 도덕과 무관한 술어로 기술하고자 한다(Pigden 1993/2006: 64).

도덕 자연주의자는 자연주의의 오류가 자연주의를 위협하는 심각한 문제가 아니라고 주장한다. 먼저 흄의 법칙을 논해 보자. 전제에는 '~해야 한다'가 없으면서 결론에는 등장하는 타당한 논증이 있을까? 있다.

> P1 차를 마시는 일은 평범한 영국인들의 일상이다.
> P2 그러므로 차를 마시는 일은 평범한 영국인들의 일상이거나 모든 뉴질랜드인은 총살당해야 한다(Prior 1976: 90).

위의 논증은 비록 결론이 공허하긴 하지만 논리적으로는 타당하다. 그래서 존재로부터 당위를 도출할 수 없다는 도덕의 논리적 자율성을 위협한다. 물론 '공허하지 않게' 전제에 나오지 않은 용어가 결론에 등장하는 타당한 추론이 있느냐고 반론할 수 있다. 그렇더라도 도덕의 논리적 자율성이, '좋음'과 '쾌락'이 서로 다른 뜻이라는 의미론적 자율성, 자연 속성으로 환원될 수 없는 독자적인 도덕 속성이 있다는 존재론적 자율성을 함축하지는 않는다(Pigden 1993/2006: 70-74).

다음으로 도덕 자연주의자들은 무어에 대응해 어떤 개념의 의미가 다르더라도 두 개념이 가리키는 속성은 동일할 수 있다고 주장한다. 대표적인 예가 '물'과 'H_20'다. "물은 H_20인가?"라는 질문은 과거에는 불가능한 질문이었다. 그러나 현대 화학이 발전하면서 우리는 물이 H_20임을 알게 되었다. '물'과 'H_20'는 우리가 익히 아는 동일한 속성을 지칭한다. 'H_20'는

과학적인 개념이다. 즉 우리는 도덕 속성과 동일한 자연 속성을, 개념적인 의미 분석이 아니라 경험적인 탐구로 발견할 수 있다.

그러나 도덕 자연주의에는 근본적인 문제가 있다. 도덕 자연주의자는 인간 심리에 대한 흄의 모델을 받아들인다. 흄에 따르면 우리의 심성 상태는 두 가지로 이루어진다. 하나는 믿음, 다른 하나는 욕구이다. 앞서 말했듯이 믿음은 세계가 어떠한가를 기술한다. 따라서 참과 거짓을 판별할 수 있다. 욕구는 세계가 어떠해야 하느냐에 대한 각자의 태도, 소망, 기원이다. 이는 옳고 그름의 문제일 수 없다. 믿음과 욕구는 철저히 구분된다. 어떤 사람은 특정 욕구를 가지면서도 그와 관련된 믿음을 결여할 수 있고 그 반대도 마찬가지다. 더하여 믿음과 욕구 단독으로는 행동을 일으킬 수 없다. 누군가 '저기에 책이 있다'고 믿는다고 해보자. 그래도 책을 읽으려는 욕구가 없다면 책을 읽지 않는다. 반대로 책을 읽으려는 욕구가 있다고 해보자. 이때도 책이 어디 있는지에 대한 믿음이 없다면 책을 읽지 못한다. 오직 믿음과 욕구가 결합될 때만 정확히 책을 읽는 그 행동을 할 수 있다. 이게 왜 문제가 되는가?

우리는 도덕 판단이 자신의 평가적 태도와 독립적인 객관적 사실에 의거한다고 기대한다. 그런데 인간 심리에 대한 흄의 모델에 따르면 도덕 사실이 있다는 믿음만으로는 도덕 행위를 위한 동기를 일으키지 못한다. 다시 말해, 도덕 자연주의자가 도덕 사실과 동일한 자연 사실을 밝혀냈다 해도 그 자체로는 도덕의 규범성을 설명하기 어렵다. 한편 자연 사실이 우리 욕구와 관련된다면 도덕 판단은 주관적이다. 그렇다면 객관적인 도덕 사실은 불필요하다. 요컨대 도덕 판단의 객관성과 실천성은 서로 상치된다. 실재론은 도덕이 객관적이라고 주장한다. 도덕 판단은 객관적 사실에 대한 믿음을 표현한다. 더불어 도덕 판단은 행위를 위한 이유를 제공한다. 그러한 이유는 우리가 판단 내린 바로 그 행동을 하도록 동기화한다. 그러나 욕구 없이 동기도 없다. 우리는 어느 하나를 포기해야 할까? 그럴 수 없다. 도덕 판단의 객관성과 실천성은 하나도 포기할 수 없는 도덕의 중핵이다. 윤리학자 스미스(M. Smith)는 도덕 실재론이 맞닥뜨린 이 난처함을 '도덕 문제(The Moral Problem)'라고 부른다(Smith 1996). 뒤에서 논의할 조이스도 이 문제를 근거로 모든 도덕 자연주의를 거부한다.

도덕 사실이 마음 독립적이라는 직관은 실재론 논쟁의 핵심이다. 마음 독립적인 사실은 통상 '객관적인'과, 반응 의존적인 사실은 '주관적인'과 연결된다. 논의를 분명히 하기 위해 객관과 주관이 내포한 여러 의미를 정리해보자. '객관적인'은 적어도 세 가지 의미를 갖는다. 첫째, 어떤 대상은 인간의 믿음, 욕구와 독립적으로 존재할 때 객관적이다. 둘째, 어떤 대상은 적절함과 비적절함을 판단할 수 있는 여지가 있을 때 객관적이다. 셋째, 어떤 사람은 특정 문제에 대해 편향되지 않은 의견을 가질 때 객관적이다. 전통적인 도덕 실재론자가 의미하는 '객관적인'의 의미는 첫 번째다. 더하여 도덕 실재론자는 첫 번째 의미가 두 번째 의미와 결합된다고 생각한다. 도덕 문제에는 적절하거나 적절하지 않은 판단이 있다. 그 기준은 마음

독립적으로 주어진다(Kirchin 2002: 25-26). 도덕 사실은 마음 독립적이기에 우리가 모두 따르는 권위를 가진다. 도덕 반실재론자들은 마음 독립적인 도덕 사실의 존재를 중점적으로 공격한다. 그런 사실은 없다. 조이스는 진화가 이에 대한 강력한 무기를 제공한다고 본다.

그러나 '객관적인'이 여러 의미를 갖듯이 반드시 마음 독립적이어야 객관적일 필요는 없다. '객관적인'의 두 번째 의미만 만족해도 충분히 어떤 대상은 객관적이다. 이는 '주관적인'의 어떤 의미가 '객관적인'의 두 번째 의미와 결합될 수 있음을 시사한다. 그렇다면 마찬가지로 '주관적인'을 세 가지 의미로 나눠 보자. 첫째, 어떤 대상은 인간의 믿음, 욕구에 의존할 때 주관적이다. 둘째, 어떤 대상은 적절함과 비적절함을 가를 수 있는 여지가 없을 때 주관적이다. 셋째, 어떤 사람은 특정 문제에 대해 편향된 의견을 가질 때 주관적이다. '주관적인'의 두 번째 의미는 상대주의이다. 이런 상대주의는 취향이나 기호 판단에서 자주 볼 수 있다. 밥과 빵 중에 무엇이 더 나은지 가리는 일은 어리석고 무용하다. 그저 선호하는 맛이 다를 뿐이다. 그런데 정말 '주관적인'에는 옳고 그름을 정할 수 있는 어떤 기준도 원천적으로 불가능할까? 그렇지 않다. 우리는 설령 각자 취향이 다르더라도 어떤 취향이 더 바람직한지, 어떤 취향은 이상하고 나쁜지 토론하고 기준을 정한다. 누구도 썩은 밥을 좋아하는 사람을 정상이라 보지 않으리라. 누구도 소아성애를 개인적인 성적 취향이라 보지 않으리라. 하지만 두 번째 의미의 '주관적인'을 받아들이면 우리는 썩은 밥 애호가, 소아성애자를 괴짜, 범죄자라고 말할 수 없다. 이런 결론에 동의하기는 힘들다. 더 좋은 예는 색 반응이다. 우리는 색 지각에 올바른 경험과 잘못된 경험이 있다고 생각한다. 정상적인 관찰자와 달리 색맹은 색을 다르게 지각하는 게 아니라 그르게 지각한다. 따라서 두 번째 의미의 주관은 너무 극단적인 상대주의이다. 우리는 과도한 주관성은 버리고 첫 번째 의미의 '주관적인'과 두 번째 의미의 '객관적인'을 결합할 수 있다(Kirchin 2002: 26-30). 즉 도덕 속성이 우리 마음에 의존해도 무엇이 옳고 그른지 객관적 기준을 제시할 수 있다. 세계에 나타나는 객관적 속성인 색은 반응에 의존한다. 우리 눈에 장미는 빨갛고 소나무는 푸르다. 감수성 이론은 여기서 출발한다.

감수성 이론은 객관적인 도덕 속성의 존재가 우리 반응과 맞닿아 있다고 주장한다(McDowell 1985; Wiggins 1987). 우리가 옳다고 생각하는 도덕 가치는 특정한 반응을 일으키는 성향[8]적인 속성을 지닌다. 따라서 그런 가치의 토대인 도덕 속성은 반응 의존적이다. 근대 철학자 로크는 모양, 운동, 고체성 등 사물의 불변하는 본질적 성질을 제1성질로, 색, 소리, 맛 등 사물 자체의 성질이 아니라 주관과의 관계에서 성립하는 성질을 제2성질로 구분했다. 맥키 등의 반실재론자는 도덕 가치는 우리와 독립적으로 존재하는 제1성질이 아니라 제2성질

8) 성향이란 어떤 종류의 경험을 산출하는 대상의 속성이다. 예를 들어 유리는 힘을 가했을 때 깨지는 성향을 가진다. 이때 유리가 현재 깨지지 않았다 하더라도 유리는 깨지는 성향을 가진다. 즉 깨지는 성향은 유리의 고정 상태, 깨짐은 발생 상태이다. 유리의 고정 상태는 발생 상태로 정의되고 실현된다.

이라고 주장했다. 제2성질은 세계의 속성이 아니라 우리 마음이 소유한 특성이나 염원이 바깥으로 투사된 현상에 불과하다. 한데 우리는 제2성질에 대한 경험을 어떤 대상의 객관적인 제1성질로 착각한다. 그래서 도덕 판단은 거짓이며 오류다. 윤리학자 맥도웰(J. McDowell)은 맥키의 회의적 태도가 제2성질에 대한 심각한 오해라고 비판한다. 우리가 경험하는 제2성질은 해당 사물이 가진 진정한 속성의 지각이다(McDowell 1985: 112). 색 경험은 단순히 마음의 투사가 아니다. 색은 외부 대상이 간직한 속성이다. 색은 우리에게 특정 경험을 일으키는 성향을 보유한 채로 우리 바깥에 있다. 그리하여 색은 우리 반응으로 비로소 실현된다. 색 경험은 반응 의존적이면서 객관적이다.

> 어떤 물체는 정상적인/표준적인/최선의 조명 조건에서 그 물체를 관찰하는 정상적인/표준적인/최선의 관찰자에게 빨간색임의 감각을 불러올 수 있을 경우 오직 그러한 경우에만 빨간색이다(Kirchin 2002: 127).

이러한 정식은 첫째, 색 지각에 대한 성향적 설명을 준다. 빨간색은 내가 볼 때만 나타나는 속성이 아니다. 정상적인 조건에서 빨간색은 언제나 빨간색을 지각하게 한다. 따라서 빨간색은 내가 보지 않을 때에도 늘 빨간색을 일으키는 성향을 가진다. 빨간 사물이 있는 곳에 누군가 있다면 반드시 빨간색을 경험한다. 둘째, 색을 잘못 지각할 가능성이 있는 색맹 등의 사람을 배제한다. 색은 정상적인 사람이라면 누구나 동일하게 지각하는 객관적 속성이다(Kirchin 2002: 127-128).

감수성 이론가들은 색 경험에 대한 정식을 그대로 도덕에 적용한다. 도덕은 우리 감각과 연결된다.

> 어떤 행동/상황/사람은 정상적인/표준적인/최선의 조건에서 정상적인/표준적인/최선의 관찰자가 도덕적으로 옳다고 판단할 수 있을 경우 오직 그러한 경우에만 도덕적으로 옳다(Kirchin 2002: 129).

색 속성과 똑같은 의미에서 도덕 속성은 실재한다. 도덕 속성은 외부 세계에 있는 성향적 속성이면서 우리 반응으로 실현된다. 더불어 감수성 이론은 도덕 판단의 실천성도 잘 설명한다. 도덕 속성이 반응 의존적이기에 인간 심리에 대한 표준 모델이 일으키는 '도덕 문제'를 회피하고, 도덕 속성이 가진 성향이 행위 동기와 연결된다고 주장할 수 있기 때문이다. 남은 문제는 그러한 반응이 무엇이고, 반응이 도덕 속성과 어떤 식으로 관련 맺는지 자연주의적으로 설명할 수 있느냐다. 이 글이 다룰 프린츠는 감정적 성향[9]이라는 기준을 제시한다. 우리

9) 프린츠는 색을 느끼는 성향과 마찬가지로 '감정'과 '감정적 성향'을 구분한다. 감정은 행위자가 현재

는 일상적으로 도덕 문제에 감정적으로 반응한다. 도덕 판단은 우리를 동요시키고 고무한다. 이때 감정은 내적 상태일 뿐만 아니라 외부 사실에 대한 평가이다. 프린츠의 감수성 이론은 3장에서 자세하게 살펴보겠다.

전통적인 도덕 실재론 논쟁을 개관하면서 우리는 자연주의적 도덕 실재론, 즉 도덕 자연주의가 갖춰야 할 요건을 알 수 있다. 첫째, 도덕 판단에는 참과 거짓을 가릴 수 있는 인지적 요소가 있어야 한다. 진리치를 판단하는 사실은 보통 인식 주관과 독립적이라고 생각해 왔다. 그러나 감수성 이론이 보여주듯 우리 반응에 의존하는 사실로도 옳고 그름을 나눌 수 있다. 물론 반응 의존적 사실은 자연과학적 방법론으로 탐구할 수 있는 성질이어야 한다. 둘째, 도덕 판단은 행위 동기를 주어야 한다. 마음 독립적인 도덕 사실이 어떻게 행위 동기를 줄 수 있느냐는 까다로운 숙제이다. 이 또한 사실에 대한 관념을 반응 의존적으로 바꿈으로써 수월하게 풀 수 있다. 우리 반응이 세계의 사실을 반영하면서 동시에 나의 내적 상태에 일어난 어떤 변화라면, 인간 심리에 대한 흄의 모델 안에서 도덕 판단과 동기가 필연적으로 연결되어 있다는 동기 내재주의적 관점을 유지할 수 있다.

느끼는 상태이고 감정적 성향은 특정 감정을 느끼게 하는 속성이다. 프린츠는 감정적 성향을 따로 '감성(sentiment)'이라 명명한다. 프린츠는 도덕 속성이 성향적 속성인 감성으로 구성된다고 주장한다. 이에 대해서는 3.2절에서 자세히 다루겠다.

2.2. 오류 이론과 초기 진화적 도덕 반실재론

우리는 일상적으로 객관적인 도덕 사실이나 가치가 있는 마냥 사고하고 행동한다. 우리는 늘 어떤 판단이 참인지 거짓인지 따진다. 이런 이유로 맥키는 도덕적 진술이 진리치를 가질 수 있음을 인정했다. 그러나 그는 어떤 도덕 판단도 참이 아니라고 단언했다. "객관적인 가치란 없다(Mackie 1977/1990: 17)." 모든 도덕 믿음은 거짓이다. 우리는 도덕 문제에서 체계적인 오류에 빠져 있다. 이러한 맥키의 인지주의 도덕 반실재론을 '오류 이론'이라 부른다. 진화적 도덕 반실재론은 맥키의 오류 이론에서 발원한다. 맥키의 논의는 진화적 도덕 반실재론으로 가기 위해 거쳐야 할 필수 관문이다.

도덕 실재론을 논박하는 맥키의 첫 번째 전략은 '상대성 논증'이다. 상대성 논증은 모든 개인, 사회, 집단, 문화가 보편적으로 따르는 도덕 믿음이나 규범은 없다는 주장이다. 심지어 같은 공동체에서도 구성원들이 받아들이는 규범의 내용은 서로 다르다. 물론 역사나 과학 분야에서도 사람들이 가진 믿음은 여럿이며 종종 불일치가 발생한다. 한데 맥키가 보기에 이는 객관적인 가치나 사실이 없어서가 아니다. 다만 잘못된 증거나 사변적 추론 때문이다. 도덕 영역은 다르다. 도덕 문제의 불일치는 사람들이 객관적 도덕 사실을 파악하지 못했다기보다 그들의 삶의 방식이 각양각색이기 때문에 일어난다. 사람들은 일부일처제가 옳기 때문에 일부일처제를 고수하는 게 아니라 일부일처제를 고수하기 때문에 일부일처제를 옳다고 생각한다(Mackie 1977/1990: 43).

그러나 맥키 자신도 인정하듯이 상대성 논증은 약하다. 도덕 믿음이 문화적으로 다양하다는 주장은 객관적인 도덕 사실의 존재가 없음을 보증하지는 못한다(Mackie 1977/1990: 44). 도덕 객관주의자들은 도덕 사실이 구체적인 도덕 규범이 아니라 보다 일반적인 원리를 지시한다고 반박할 수 있다. 각 지역의 특수한 도덕 믿음은 보편 원리가 상이한 환경이나 문화와 결합하여 생겼을 뿐이다. 또한 도덕적 불일치가 반드시 생활 방식의 차이로만 발생하는지도 의문이다. 도덕적 주장도 여타 지식처럼 무지나 미신, 타성, 비합리적 권위 등에 의해 의견이 나뉠 수 있다. 이렇게 잘못된 근거에서 나온 도덕 믿음은 받아들일 수 없다. 전통에 따라 율법을 어긴 여성에 대한 명예살인은 옳다고 주장하는 사람을 용인할 수 있을까, 여성에 대한 억압은 단지 문화의 차이일 뿐일까? 그렇지 않다. 우리는 직관적으로 이런 전통, 율법, 문화는 불합리하다고 생각한다. 결국 불일치 자체는 효과적인 논박이 되지 못한다. 우리가 믿는 어떤 도덕 가치가 타인과 상충해도 그 가치는 객관적으로 실재할 수 있다.

맥키의 두 번째 전략은 '기이함 논증'이다. 기이함 논증은 매우 중요하다. 이로써 맥키는 도덕에 관한 핵심적인 직관을 공격했다. 바로 도덕 판단의 '객관적 규정성(또는 실천성)'이다. 도덕 사실은 우리 마음과 독립적인, 객관적 존재에 대한 믿음인 동시에 특정한 행동을 하게

만드는 규정적 요구이다. 맥키는 이를 '행위 이유'와 연결했다. 우리가 도덕 사실을 알면 사적인 욕구나 득실과 상관없이 그 행동을 해야 할 이유를 갖는다. 고로 그 행동을 해야 할 동기를 갖는다. 이 점에서 도덕의 규정성은 칸트의 정언 명령이다.

그런데 맥키가 보기에 이 세계 내에 객관적 규정성이라는 특성을 갖춘 존재가 있다는 주장은 기이하다. "객관적 가치가 있다면 그것은 우주의 그 어느 것과도 다른 매우 이상한 종류의 존재자, 성질, 관계일 것이다(Mackie 1977/1990: 45)." 기이함 논증은 앞서 스미스가 말한 '도덕 문제'와 같다. 맥키 역시 인간 심리에 대한 흄의 이론을 받아들였다. 행위 이유나 동기는 믿음과 욕구의 결합으로만 발생한다. 그렇다면 객관성과 규정성은 서로 모순된다. 어떻게 우리의 평가적 태도와 독립적인 어떤 사실이 그 자체로 행위를 위한 이유나 동기를 주는가, 객관적이면서 규정적인 사실이 정말 있다면 이를 어떻게 인식할 수 있는가? 우리는 도덕 사실을 알기 위한 매우 특별한 인식 기관을 가정해야 할지도 모른다. 그러나 그런 기관은 없다. 따라서 객관적 규정성은 자연 세계에 수용할 수 없는 기이한 속성이다. 객관적인 가치란 없다.

왜 주관적 속성에 불과한 도덕 가치가 객관적이라 착각할까? 맥키는 흄을 인용하여 이는 우리 마음에 있는 욕구나 감정 등을 세계에 투사하여 마치 객관적인 양 생각하는 데 불과하다고 주장한다. 도덕 가치는 특히 사회적 규제라는 임무를 맡고 있기 때문에 그 가치가 객관화되어야 한다는 강한 동기를 갖는다(Mackie 1977/1990: 50-51).

진화적 도덕 반실재론은 오류 이론을 계승한다. 진화적 도덕 반실재론자의 주장은 다음과 같다. 진화는 첫째, 오류 이론의 경험적 토대를 제공하며 둘째, 도덕이 자연선택의 산물이라면 우리는 필연적으로 오류 이론으로 귀착된다. 마이클 루즈의 메타윤리 이론으로 이를 살펴보자.

우리 조상들은 혼자서 살 수 없는 동물이었다. 그렇기에 동료와 협동하고 무임승차자를 처벌하는 등 사회적 조화를 이루는 일은 매우 중요했다. 협동 관계의 규제는 개체의 번식 성공도를 높이는데 크게 작용했다. 상황이 이렇다면 자연선택은 분명 협동을 잘 하는 개체를 선호하리라. 협동을 잘하려면 어떻게 해야 할까? 개체가 늘 협동 행동을 하도록 유전적으로 미리 결정하는 게 좋을까, 아니면 매순간 자신의 편익을 따지는 합리적 능력을 설계하는 게 나을까? 둘 다 아니다. 양 전략 모두 변화하는 상황에 적응하지 못하거나 너무 많은 자원을 소모하게 만든다. 선택은 그 중간의 길을 택한다. 인간에게는 주변 환경에 맞추어 특정 유형의 행동을 선호하는 어느 정도 생득적인 경향이 진화했다. 개체는 동료를 도우려는 욕구나 감정을 느낀다. 한데 일단 도덕 판단을 하는 심리 기제가 진화하면 적절한 감정이나 욕구가 없이도 그 행동을 해야만 한다고 느끼도록 발전한다(Ruse 1986: 221). 이제 개체는 이타적 욕구가 없더라도 동료를 돕는 게 의무라고 여긴다.

위와 같은 도덕의 진화에 대한 설명이 왜 오류 이론을 함축하는가? 루즈가 제시한 첫 번째 근거는 '잉여성 논증'이다. 루즈의 도덕 진화 모델이 맞다면 굳이 도덕적 행위를 정당화해주는 도덕 사실을 가정할 필요가 없다. 도덕 판단은 단지 생존과 번식 성공을 위해 진화한 적응일 뿐이다. "우리는 다른 사람을 돕고 그들과 협동해야 한다고 느낀다. 우리는 그렇게 진화했다. 이는 도덕의 기원과 지위에 대한 완전한 대답이다. 여기에 플라톤적인 가치의 세계는 필요하지 않다. 도덕에는 인간적 맥락을 벗어난 의미도 정당화도 없다(Ruse 1986: 252)."

루즈의 두 번째 근거는 '특이성(idiosyncrasy) 논증'[10]이다. 우리는 '정의', '공정성', '사랑', '연민' 등의 개념을 소중하게 생각한다. 왜 그런가? 우리가 가치 있다고 여기는 도덕 믿음과 개념들은 진화 역사를 반영하는 특이적인 선택의 산물이다. 즉 인간 종의 유전적 역사가 달랐다면 현재와 전혀 다른 개념과 믿음을 가졌을 수도 있다. 과거 조상들이 근친상간이 장려되고 강간이 용인되는 세계에 살았다면 지금 우리는 근친상간이나 강간을 옳은 행동이라 생각하지 않겠는가. 따라서 특정한 인간 본성을 넘어서는 필연적인 도덕 원리는 없다(Ruse & Wilson 1986: 186).

정리하면 도덕 판단이 오류인 까닭은 단순하다. 자연선택이 그렇게 만들었다. 객관적인 도덕 사실에 대한 믿음은 자연선택의 산물이다. 이는 왜 우리가 도덕을 객관적인 가치로 생각하는지 설명한다. 자연선택은 효율적으로 진화적 목적을 달성하기 위해 "객관성이라는 환상(Ruse 1986: 253)."을 심어주었다. 그리하여 도덕에 대한 진화적 설명은 오류 이론으로 귀결된다. 진화는 도덕 판단을 정당화하는 기획을 쓸모없게 만든다. 도덕 믿음이 생긴 인과적 역사를 알면 믿음의 정당화를 위한 특별한 사실을 들여올 필요가 없다. 우리는 객관적인 도덕 사실이 있는지 알 수 없고, 존재하지도 않는다.

진화가 도덕 판단의 정당성을 해친다는 주장은 설득력이 강하다. 진화는 분명 객관적인 도덕 사실의 인식과는 무관한 과정이다. 맥키의 오류 이론과 이를 바탕으로 한 루즈의 진화적 도덕 반실재론은 모두 마음 독립적인 사실만이 도덕 판단의 참을 보증한다고 전제한다. 마음 독립적인 도덕 사실이 없기에 도덕 판단은 거짓이다. 한데 루즈가 말하듯, 도덕은 인간 종에게만 특유하게 나타나는 행동 양식이다. 도덕은 조상들의 삶을 규제하고 번식 성공도를 높였던 적응이다. 그렇다면 도덕은 우리 반응에 의존하여 구성되는 인간 종의 역사적 산물이다. 인간이 없으면 도덕도 없다. 진화와 마음 독립적인 도덕 사실을 양립시키는 일은 애초에 불가능한 요구이다. 진화를 이용하여 도덕 판단의 참을 회복하려면 도덕 사실이 마음 독립적이라는 전제를 버려야 한다. 그때 진화는 특정한 반응을 형성했던 필연적인 과정으로 변모할 수 있다. 다음 절에서 루즈의 진화적 도덕 반실재론에 대응했던 선행 연구를 분석하여, 진화가 도덕 사실을 정당화할 수 있는 참된 과정이려면 어떤 조건이 필요한지 알아보자.

---

10) '특이성(idiosyncrasy) 논증'이라는 이름은 James(2011), p.170. 에서 따왔다.

2.3. 초기 진화적 도덕 반실재론에 대응한 선행 연구 비판

윤리학자 리처즈(R. J. Richards)는 진화적 사실로 인간이 추구해야 할 도덕 가치가 무엇인지 알 수 있다고 주장한다. 우리는 공동체를 이루는 구성원들의 복지와 아이들을 보호하는 데 힘쓰는 사회적 동물이다. 따라서 공동체의 좋음이야말로 우리가 성취해야 할 도덕적 좋음이다(Richards 1986: 286). 스펜서를 떠올리게 하는 이러한 주장은 당장 자연주의의 오류를 범한다. 사실 진술에서 가치 진술을 곧바로 이끌어 낼 수는 없다. 물론 앞서 보았듯이, 자연주의자들은 연역적인 논리법칙에 기대지 않고 경험적으로 도덕 원리를 찾을 수 있다. 그러나 리처즈는 연역적인 방식으로도 자신의 주장을 정당화할 수 있다고 단언한다.

리처즈는 모든 사람이 동의하는 경험적 사실과 도덕 가치가 있다고 생각한다. 그렇기에 전형적인 도덕적 인간의 믿음과 실천을 조사하여 필연적으로 '~이다'에서 '~해야 한다'를 도출할 수 있다. 도덕은 자연선택의 산물이다. 이는 경험적 사실이다. 공동체의 복지는 지고한 도덕적 좋음이다. 이 또한 경험적 사실로 뒷받침된다. 여기서 리처즈는 '~해야 한다'의 의미를 다르게 해석한다. 다음을 보자. "번개가 친다. 그러므로 천둥이 쳐야한다." 번개가 치면 천둥이 치는 현상은 물리 법칙에 따르는 필연적인 현상이다. 즉 경험 법칙이 '~이다'에서 '~해야 한다'로의 이행을 보증한다(Richards 1986: 286-287). 같은 방식으로, 인간은 도덕적이다. 따라서 인간은 공동체를 위해 행동해야 한다. 이는 진화가 산출한 피할 수 없는 조건이다.

리처즈의 논의는 즉각적인 의문을 부른다. 공동체의 좋음이라는 목적이 물리 법칙과 같은 의미의 도덕 법칙인가? 우리는 무조건적으로 공동체를 위해 행동하는가? 그렇지 않다. 가까운 친구, 자식, 부모에 대한 사랑은 사회적 가치를 최대화하기보다는 사적 관계로만 국한된다. 공동체의 관점에서 이는 이기적 행동이다. 그렇지만 누구도 친밀한 사람에게 쏟는 애정을 비도덕적 행동이라 비난할 수 없다. 여기에 리처즈는 사적인 도덕 행동도 사실은 공동체를 위한 배려라고 주장할 수 있다. 그러나 이는 너무 일반적이어서 인간의 모든 행위를 포괄한다. 그러면 공동체의 좋음이란 목적은 큰 의미가 없다.

또한 리처즈가 해석한 '~해야 한다'는 도덕 의무보다 약한 예측적 의무이다. 번개가 치면 천둥이 칠거라 예상할 뿐이다. 그렇다면 어떻게 예측적 의무가 도덕적 구속력을 가질 수 있을까? 우리는 직관적으로 도덕 의무는 어떤 욕구나 목적에 상관없이 해야 하는 행동이라고 여긴다. 한데 예측적 의무는 왜 약속을 지켜야 하는지, 거짓말을 하지 말아야 하는지에 대한 정언적 이유를 줄 수 없다. 더구나 예측적 의무는 전혀 도덕적이지 않은 사람에게 적용될 수 있다. 여기 은행 강도가 있다고 하자. 그는 위급 상황에 맞선 의무를 가진다. "경찰이 오면 인질을 잡아야 한다." 체포될 처지가 되면 인질을 잡으리라는 계획 역시 예측적 의무이다. 그러나 우리는 은행 강도가 도덕 행동을 하고 있다고 말할 수 없다. 리처즈의 '~해야 한다'가

가진 실천적 난점은 조이스도 지적한다(Joyce 2006: 159-160).

윤리학자 로체퍼(W. A. Rottschaefer)와 마틴슨(D. Martinsen)은 루즈의 윤리학이 너무 과장된 다윈주의라고 비판한다(Rottschaefer & Martinsen 1990). 루즈는 도덕이 진화의 산물일 때 객관성과 규정성을 함께 갖는 도덕 사실의 존재는 정당화되지 않는다고 주장했다. 이에 로체퍼와 마틴슨은 먼저 우리의 완전한 도덕 능력은 단순히 선천적인 발달 경로를 따르는 게 아니라 학습과 추론의 결과라고 반박한다. 예를 들어 어린 아이들이 일찍부터 나타내는 양심은 부모가 시행한 도덕적 훈련의 결과이다. 부모는 도덕적으로 올바른 행동과 자기 이익이 충돌하는 상황에서 아이들이 도덕 행동을 하도록 지도한다. 아이들은 칭찬의 유익함과 처벌의 두려움으로 도덕 원리를 내면화한다. 따라서 로체퍼와 마틴슨은 도덕의 객관성과 규정성은 본성이 아니라 양육으로 획득한 성격이라고 주장한다(Rottschaefer & Martinsen 1990: 155-156).

다음으로 로체퍼와 마틴슨은 도덕 속성이 비물리적인, 플라톤적 방식으로 존재할 필요가 없다고 주장한다. 도덕 속성은 자연 세계 내에 독립적으로 자리할 수 있다. 어떻게 가능한가? 도덕 속성은 생물학적으로 발생한 관계적 속성이다. 따라서 도덕 속성은 진화한 인간의 믿음, 행동 등이 나타내는 자연 속성에 수반[11]한다. 이는 마치 생물학적 적합도라는 속성이 유기체가 보유한 속성에 수반하는 모양과 같다(Rottschaefer & Martinsen 1990: 161-162). 로체퍼와 마틴슨은 자연 속성과 도덕 속성의 수반적 관계는 단지 생물학적 발생 과정에서 우연히 만들어지지 않았다고 말한다. 도덕은 도덕적 좋음에 관한 인간의 의식과 자유를 반영한다. 도덕적 좋음은 첫째, 인간의 생존과 번식을 증진하는 데 기여했던 역사와, 둘째, 인간의 사회문화적 학습, 이성과 관계 맺는다. 도덕 속성의 수반은 두 번째와 관련된 인간 활동으로 성취된다. 인간의 생물학적 적응을 높여주었던 좋음은 이를 추구하는 의식적 활동으로 도덕적 좋음이 되는 관계적 속성이다. 다시 말해 우리가 의식적으로 자연선택된 협동 행동을 한다면, 협동은 도덕적 좋음, 옳음을 예화한다. 더불어 그런 도덕적 좋음은 우리에게 생물학적으로 좋은 사물이나 사태가 다양한 만큼 다원적이다(Rottschaefer & Martinsen 1990: 163).

그러나 인간의 문화와 이성적 추론을 강조하는 로체퍼와 마틴슨의 주장은 진화적 도덕 반실재론에 대한 큰 반론이 되지 못한다. 루즈는 도덕 능력에 대한 문화나 양육의 역할을 수용할 수 있다. 루즈의 진화윤리학은 유전자 결정론이 아니다. 아이들은 자라면서 다양한 자극을 받고, 타인의 행동을 모방하며 도덕 믿음을 습득한다. 다만 그 과정에 선천적 제약이 있을 뿐이다. 결국에 부모의 양육과 학습은 선천적인 발달 경로에 포함된다. 따라서 로체퍼와 마틴슨

11) 수반이란 똑같은 자연 속성을 가진 사물들은 모두 동일한 도덕 속성을 공유함을 뜻한다. 이때 도덕 속성이 변했다면 자연 속성에도 변화가 있다. 즉 자연 속성이 변하지 않는다면 도덕 속성도 변하지 않는다. 그리하여 어떤 두 세계의 자연 속성이 같다면 도덕 속성도 같다. 수반과 도덕 실재론에 관한 논의는 주동률(1996)을 참고하라.

의 설명은 루즈와 큰 차이가 없다. 더구나 루즈는 아이들이 학습과 양육 덕에 도덕이 마치 객관적인 양 착각하는 거짓 믿음을 내재화한다고 반론할 수도 있다.

로체퍼와 마틴슨이 제시한 도덕 속성과 자연 속성의 수반적 관계는 어떠한가? 그들은 인간의 의식적 활동이 생존과 번식을 증진한 좋음을 선호하면 그 행동이 도덕적 좋음이라는 속성을 예화한다고 주장한다. 도덕적 좋음은 인간의 자유로운 행위에 수반한다. 이러한 도덕 속성은 우리에게 행위 이유를 주는가? 도덕 속성은 규범적 행동을 일으키는 인과적 역할을 해야 한다. 또한 그 규범성은 다른 타산적 이유와 구별되는 강력한 요구여야 한다. 반면 로체퍼와 마틴슨이 제시한 도덕 속성은 가언적인 요구이다. “X를 원한다면 Y를 해라.”, “사회적 조화가 좋다면 동료를 도와라.” 당연히 욕구가 소멸하면 가언 명령은 철회된다. 즉 가언 명령은 늘 취소될 가능성이 있어 어떻게 도덕적 욕구가 여타 이기적 욕구를 이길 수 있는지 설명하기 어렵다. 적합도를 증진한 좋음은 반드시 도덕적 좋음만은 아니기 때문이다. 다른 사람을 착취하는 행동이 나의 번식 성공에 도움이 된다면 기꺼이 그럴 수 있다. 게다가 인간이 의식적으로 적합도 증진을 위해 행동한다고도 볼 수 없다. 대개 우리는 도덕 행동의 유용성을 의식하지 않고 그냥 해야 하기 때문에 한다. 마지막으로 로체퍼와 마틴슨은 다원적인 도덕 속성이 무엇인지 예시하지 않았다. 단지 적합도 증진에 좋은 속성을 나열한다면 어떤 종류라도 포함될 우려가 있다.

윤리학자 캠벨(R. Campbell)은 특정한 도덕 믿음의 정당화를 시도하지 않는다. 그가 제기하는 질문은 단순하다. 왜 도덕이 생겼는가? 캠벨은 궁극적 물음에 대한 진화적 설명이 충분히 도덕 판단의 참을 보증한다고 주장한다(Campbell 1996). 답은 이렇다. 도덕을 갖춘 집단의 모든 구성원이 그렇지 않은 집단보다 훨씬 생존에 유리하다. 그러므로 도덕을 갖는 게 정당하다. 캠벨은 자신의 주장이 정당화를 위한 세 가지 의미를 만족시키기 때문에 객관적이라고 말한다. 첫째, 정당화는 진화가 사실이라는 의미와 같다. 둘째, 정당화는 특정인의 도덕 믿음에 의존하지 않는다. 셋째, 정당화는 이미 도덕이 정당하다고 믿는 사람의 선호나 욕구와 관련 없다(Campbell 1996: 24). 더불어 캠벨은 생존에 유용한 이익으로 규범성까지 설명한다. 즉 인간은 상호 이득을 위해 자연스럽게 다른 사람을 위한 행동을 한다(Campbell 1996: 26).

규범성에 대한 캠벨의 설명은 앞서의 논자들이 직면했던 비판을 그대로 받는다. 이익을 위한 도덕 행동은 도덕 요구의 특별한 성격을 설명하기 어렵다. 도덕 행동을 위한 이유는 다른 타산적 이유를 압도(override)할 수 있어야 한다.

조이스는 캠벨의 주장이 가진 난점을 잘 파악했다. 먼저 진화는 도덕이 현재 유용함을 함축하지 않는다. 도덕은 첫째, 과거에 유용했거나 둘째, 우리 조상들의 유전자에 유용했다. 캠벨은 개체의 적합도에 유용한 행동이 개체 자신에게도 유용하다고 착각하고 있다. 하지만 이

둘이 반드시 일치하지는 않는다. 또한 유용함이 정당화를 준다면 종교나 주술도 참이라고 생각할 여지가 있다. 즉 유용함은 도구적 정당화만 주지, 인식적 정당화는 주지 못한다. 인식적 정당화에는 객관적 사실이 필요하다. 마지막으로 도덕이 유용한 허구일 수 있다. 도덕이 객관화라는 믿음이 생존과 번식에 도움이 되기 때문에 실제로 그렇지 않은 데도 도덕을 참으로 믿을 수 있다(Joyce 2006: 161-163).

이상으로 루즈의 진화적 도덕 반실재론에 대응한 선행 연구들을 살펴보았다. 진화적 사실이 도덕 판단을 정당화하는 토대로 사용되려면 무엇이 설명되어야 하는가? 첫째, 우리가 자연선택 덕에 선천적으로 습득한, 관련된 영역이나 문제를 도덕적으로 사고하게 만드는 기초 가치는 일원적이지 않다. 사회적 조화나 공동체의 복지 같은 단일한 자연 속성은 인간 삶의 다면성을 반영하지 못한다. 그래서 쉽게 자연주의의 오류에 빠질 위험이 있다. 조상들이 마주쳤던 적응적 문제는 다양하기 때문에 거기에서 파생된 기초 가치도 다원적이라 생각한다. 그러한 기초 가치들을 명시할 수 있어야 한다. 둘째, 진화한 기초 가치의 여러 목록을 찾았다 해도 실제로 어떻게 그 가치들을 인식하고 따르는지 밝혀야 한다. 행위자는 문화나 환경의 영향으로 자신에게 중요한 가치를 모를 수도 있다. 그저 무지한 상태로 남으면 진화 역사에서 자연선택된 기초 가치가 있다는 주장은 큰 의미가 없게 된다. 셋째, 진화에 바탕을 둔 도덕 자연주의는 기본적으로 흄의 모델을 따른다. 그렇다면 어떻게 도덕적 욕구가 타산적 욕구를 이기고 도덕 행동을 일으키는지 해명해야 한다. 실천성은 도덕 판단의 핵심이다. 정언적 이유라는 믿음은 차치하고라도 우리가 일상에서 왜 도덕 명령을 강력한 요구라고 생각하고 순응하는지 답할 수 있어야 한다. 세 번째 질문은 '진화적 도덕 실재론'이 성공하기 위한 필수 요건이다. 다음 장에서 본격적으로 조이스와 프린츠의 메타윤리 이론을 비교한 뒤, 프린츠를 비판적으로 옹호하면서 제기된 난점을 해결할 수 있는 가능성을 모색해보자.

3. 조이스와 프린츠의 메타윤리 이론 비교

3.1. 도덕 판단의 본성

리처드 조이스의 책, 『도덕의 진화(*The Evolution of Morality*)』는 두 부분으로 이루어진다. 첫 번째는 도덕의 기원과 본성을 진화적 관점으로 탐구한다. 두 번째는 도덕이 진화의 산물이라는 사실이 어떤 메타윤리학적 함축을 갖느냐 논한다. 이 절에서는 먼저 첫 번째 주제, 진화와 도덕의 관계를 살펴보자.

조이스는 도덕의 진화를 다루기에 앞서 도덕적 사고의 가치란 무엇인가, 우리가 서로를 도덕적으로 바라봄으로써 얻는 이득이 무엇인가 묻는다. 누누이 언급했듯이, 그 답은 도움이나 협동 행동의 증진과 유지이다. 우리는 타인과 상호작용할 때 '덕', '의무', '정의' 같은 개념을 가진 구성원을 더 바람직하다고 생각한다(Joyce 2006: 13). 도덕적인 사람은 배신과 기만을 경계하고 다른 사람과 공동체의 안녕을 중요하게 생각하리라.

서로 도움을 주고받는 행동은 개체의 적합도를 높였다. 따라서 자연선택은 인간이 타인의 복지에 마음 쓰는 이타적 동기를 갖게 했다. 이때 이기주의자는 협동 행동이 사실은 개체의 번식적 이득을 위한 자기 본위적 동기에서 나온다고 반론할 수 있다. 하나 이기주의자는 지금 행동을 일으키는 '근접인(proximate cause)'과 그러한 행동이 왜 존재하는지를 설명하는 진화적 원인, '궁극인(ultimate cause)'을 혼동하고 있다. 진화생물학이 사용하는 근접인과 궁극인이라는 설명 층위의 분리는 인간 행동에 대한 흔한 오해를 부른다. "인간은 유전자를 전달하려는 본능 때문에 서로 사랑하고 동침한다."라고 말하는 경우를 보자. 이런 언술은 사람들이 각자의 유전자 사본을 복제하려는 의식 하에 배우자를 선정한다고 착각하게 한다. 정확히 말해, 인간은 유전적 적합도를 높이려고 행동하지 않는다. 되레 인간을 움직이는 요인은 사랑이 주는 기쁨과 행복이라는 근접 자극이다. 개체는 자신의 행동 근저에 있는 궁극인을 몰라도 된다. 그래야 더 쉽고 빠르게 행동을 이끌 수 있다. 즉 협동 행동이 궁극적으로 쌍방의 번식 성공률을 높이겠지만 진화적 이익을 자각할 필요는 없다. 더구나 협동 행동이 주는 이득이 반드시 유전자 복제를 의미하지도 않는다. 우리는 심리 수준에서 유전자의 눈으로 타인과 세계를 보지 않는다. 우리는 정말로 남을 위해 희생한다. 누군가에게 진정으로 마음 쓸 때 협동은 더 효과적으로 시행된다.[12] 인간이 다른 이를 소중히 여긴다는 사실은 사랑, 우정,

12) 이타적인 동기는 도움 행동을 규제하는데 이기적인 동기보다 더 신뢰할 만하고 효율적이다. 심리적 이기주의의 한 형태인 쾌락주의를 예로 들어 보자. 멀리서 자식의 울음소리가 들린다. 그 소리를 듣는 부모는 공포와 불안을 느낀다. 공포와 불안을 없애려면 아이를 도와야 한다. 따라서 아이를 도우러 간다. 여기서 부모의 도움 행동은 아이의 울음이 부모에게 공포와 불안을 일으키고 그 부정적 감정을 없애려면 아이를 도와야 한다는 믿음이 매개되어 이루어진다. 반면 심리적 이타주의는 아이

공감, 연민 같은 친사회적 감정의 존재로 입증된다. 상대방을 향한 감정은 협동 행동을 촉발하는 이타적 동기를 제공한다. 친사회적 감정은 가족과 친지, 낯모르는 사람을 돕도록 만드는 근접 기제다.

친사회적 감정만으로 도덕의 진화와 도덕 판단의 진면목을 설명하기에 충분할까? 조이스는 아니라고 말한다. 왜 그럴까? 조이스는 우리가 '원해서(want)' 하는 행동과 '해야 한다고(ought)' 판단해서 하는 행동이 다르다고 지적한다. 감정은 무엇인가를 원하는 욕구이다. 친사회적 감정은 사회가 평화롭고 조화롭기를 원하고 사람들이 서로 사랑하기를 원한다. 사람들은 살인이나 절도를 원하지 않기 때문에 그런 행동을 '억제(inhibition)'한다. 그러나 조이스가 보기에 도덕 판단은 원치 않음을 떠나, 살인이나 절도는 옳지 못하므로 해서는 안 된다는 '금지(prohibition)'의 개념을 포함한다. 도덕 판단은 금지의 개념을 이해하는 능력이 있어야 한다. 도덕에서 연민이나 혐오의 감정은 분명 중요한 역할을 담당한다. 감정은 동기를 일으킨다. 하지만 감정은 도덕 판단을 구성하는 필수 요건은 아니다. 우리는 타인을 사랑하면서도 그를 해치지 않으려는 어떤 의무도 갖지 않은 사람을 상상할 수 있다. 친사회적 감정의 발생만으로 인간이 어떻게 도덕적 옳고 그름을 판단하게 되었는지 설명할 수 없다(Joyce 2006: 50-51).

단지 누군가를 도우려 원하는 일은 도덕 판단이 아니다. 도덕 판단이란 정확히 무엇인가? 조이스는 먼저 도덕 판단이 어떤 심성 상태를 표현하는가 묻는다. 조이스는 도덕 판단에 대한 순수한 비인지주의도, 인지주의도 틀렸다고 답한다. 도덕 판단에는 인지적 요소와 비인지적 요소가 함께 있다. "한스는 독일놈(kraut)이다."라는 판단을 생각해보자. 이 판단은 '한스는 독일인이다.'라는 믿음과 함께 한스와 같은 국적을 가진 사람들에 대한 경멸의 감정을 표현한다. 즉 도덕 판단은 믿음과 능동적인(conative) 태도를 함께 표현한다(Joyce 2006: 54). 그런데 상술했듯이, 감정으로 대변되는 능동적인 태도는 무엇을 해야 한다는 도덕적 의무의 개념을 잘 포착하지 못한다. 도덕 행동을 원하는 욕구는 다른 욕구에 밀려날 수 있다. 이에 조이스는 도덕 판단에는 어떤 특별한 믿음이 결부되어 있다고 주장한다. 도덕 판단은 외부 세계에 대한 가치 중립적인 진술이 아니다. 도덕 판단에는 저항할 수 없는 모종의 힘이 있다. 바로 도덕 판단의 '실천적 힘(practical clout)'이다.

"살인은 도덕적으로 옳지 않아!"라는 판단은 쉽게 무시할 수 없는 숙고를 일으킨다. 이는 "살인 우우!"라는 감정 표현, "살인은 나를 화나게 한다."라는 감정 보고와 다르다. 느낌에 대한 진술은 그 자체로 청중에게 어떤 구속력도 주지 못한다. 상대방이 발화자의 내적 상태에 전혀 신경 쓰지 않는다면 "그래서 뭐?"라고 대꾸할 수 있다. 하나 도덕 판단의 실천적 힘

의 울음이 들릴 때 아이가 도움이 필요하다고 믿고 곧바로 아이를 돕는 행동을 산출한다. 심리적 이타주의의 진화에 관해서는 Sober & Wilson(1998)을 참조하라.

은 상대방이 어떤 욕구나 소망을 갖느냐에 상관없이 그 행위를 하지 말아야 할 이유를 준다. 도덕은 특정한 목적을 위한 수단이 아니다. "무고한 사람을 살인하지 마라."라는 언명은 자신의 별난 욕구와 소망을 들먹이면서 무시할 수 있는 성질이 아니다. 즉 도덕 명령은 정언 명령이다.

> 어떤 욕구를 만족시키려면 무엇을 해야 하는지 조언하는 가언 명령은 우리가 도덕 명령에 기대하는 실천적 힘(practical oomph)을 가질 수 없다. 도덕 명령은 우리가 바라는 특별한 목적을 들면서 회피할 수 있는 요구가 아니다. 사람들은 특별한 목적을 이유로 도덕 명령을 피할 수 없다. 도덕 명령은 불가피(inescapable)하다(Joyce 2006: 61).

조이스는 정언 명령이 도덕 체계가 가진 '협상 불가능한(non-negotiable)' 특성이라고 주장한다. 이는 정언 명령을 수용하지 않은 도덕 이론이나 체계는 도덕이라 볼 수 없다는 뜻이다. 도덕 판단이 가진 실천적 힘은 도덕의 핵심이다. 조이스는 도덕 판단의 실천적 힘을 두 가지로 나눈다. 하나는 '불가피함'이다. 도덕 명령은 각자의 욕구나 목적과 상관없는, 피할 수 없는 의무다. 다른 하나는 '권위'이다. 도덕 권위는 여타 관습이나 제도에 의존하지 않는, 그 자체로 따라야 할 무거운 이유를 주며 명령을 무시한 사람을 비합리적인 행위자로 만든다. 도덕의 규범성은 바로 이 불가피함과 권위로 이루어진다. 불가피한 권위를 가진 규범 체계에는 실천적 힘이 있다(Joyce 2006: 62). 결론적으로 도덕 판단은 불가피한 권위를 가진 실천적 요구에 대한 믿음을 표현한다.

자연선택은 우리 조상들이 도덕 판단에 불가피한 권위가 있다고 믿도록 설계했다. 타인에 대한 사랑과 관심만으로 협동 관계가 유지되기는 힘들다. 개체는 언제나 독단적 행동으로 이익을 독점할 유혹을 받는다. 이때 도덕은 사회적 관계의 붕괴를 막는 특별한 심리, 행동 기제로 작동한다. 행위자는 협동 행동이 도덕적으로 요구된다면 협동을 좋아하든 싫어하든 준수해야 한다고 생각한다. 또한 협동 행동이 회피할 수 없는 권위로 주어진다면 적절한 행동을 하지 않았을 때 비난받아 마땅하다고 인정한다. 다시 말해, 도덕적 사고는 마음속에서 타산적 계산을 멈추도록 만든다. 그리고 개체는 다른 사람도 자신과 같기를 요청한다. 이렇게 자기와 타인을 도덕적으로 보는 개체는 장기적으로 의지할 수 있는 좋은 동료다. 각자는 양심에 따라 옳은 일을 행하고 배신자는 처벌하려 한다. 따라서 도덕은 상호 헌신이며 우리를 묶어주는 공통의 접착제이다(Joyce 2006: 117-118).

자연선택은 어떻게 실천적 힘을 가진 도덕적 사고를 만들었을까? 조이스는 맥키, 루즈와 마찬가지로 도덕 판단은 우리 마음의 투사라고 주장한다. 도덕 판단이 나타내는 불가피한 권위는 마음의 창조물이다. 조이스는 자연선택이 감정을 이용하여 이런 착각을 만들었다고 생각한다. 프린츠의 이론을 논하면서 보겠지만 도덕 판단에서 감정은 중요한 기제로 작용한다.

동물 학대를 금지하는 판단의 근저에는 고통 받는 동물을 보며 느끼는 연민이 있다. 그러나 조이스는 감정 체험은 세계가 보유한 특징을 지각하는 게 아니라고 부정한다. 동물에게 느끼는 연민은 그 동물이 가진 속성이 아니라 우리 감정의 투사다. 그럼에도 그 장면을 보는 행위자는 마치 세계에 연민이라는 속성이 있는 양 오인한다. 왜 그럴까? 인간 삶의 어떤 영역에서는, 현실에 대한 정확한 지식을 버리는 게 적합도에 도움이 된다. 절벽 아래를 내려다보고 공포를 느껴 실제보다 그 높이를 과장하는 개체는 그렇지 않은 동료보다 생존할 가능성이 더 크다. 조이스는 도덕 영역이 바로 그런 경우라고 지목한다. 도덕 판단에는 참된 믿음이 필요하지 않다. 인간의 도덕 감수성이 마치 지각 기관처럼 세계의 도덕 사실을 탐지한다는 증거는 없다. 왜냐하면 자연선택의 목적은 "성공적인 사회 행동을 증진하는 데 있지, 세계의 특징을 탐구하는 데 있지 않다(Joyce 2006: 131)." 도덕 판단이 감정의 투사라고 해서 조이스가 도덕을 전적인 감정의 산물로 보는 건 절대 아니다. 다만 조이스는 우리 조상들이 도덕적 사고를 하기 전, 먼저 감정의 투사가 있었다고 한정한다. 일단 감정으로 이 세상에 친사회적인 바람이나 욕구를 흩뿌리고 나자, 감정 없이도 욕구와 무관한 도덕 이유가 있다는 믿음이 가능해졌다. 조이스는 이를 가리켜 투사적 경향이 지성적 힘으로 변모한 상황이라 말한다(Joyce 2006: 133).

프린츠는 조이스와 정반대로 감정[13]을 도덕 판단의 핵심으로 상정한다. 감정은 도덕 판단이 일어난 뒤에 따라오는 부차적 요소를 넘어 선다. 감정은 또한 도덕 판단을 추동한다. 우리는 도덕 가치를 느낀다. 도덕적 옳음과 그름은 우리가 느끼는 세계의 속성이다. 사악한 행위는 우리에게 부정적인 감정을 일으키며 부정적인 감정은 그 행위를 도덕적으로 나쁜 행위라 규정짓게 한다. 따라서 도덕 판단은 감정을 표현한다. 더 정확히, 도덕 판단은 어떤 사태에 대한 감정적 성향을 표현한다.

도덕 판단과 감정을 연결하는 프린츠의 입장은 두 가지 논제로 나뉜다. 첫째는 형이상학적 논제로, 도덕 속성은 본질적으로 감정과 관련된다고 주장한다(Prinz 2007: 14). 이는 도덕 속성의 존재를 색 지각에 유비하여 설명하는 감수성 이론과 같다. 외부 사물이 가진 빨강이라는 속성은 우리 시각으로 경험된다. 마찬가지로 옳고 그름의 도덕 속성은 감정을 느낌으로써 비로소 파악된다. 도덕 속성은 감정 없이 존재할 수 없다. 도덕 속성은 우리 반응에 의존한다. 이런 점에서 프린츠의 견해는 에이어의 정서주의와는 다르다. 도덕 판단은 감정이라는 비인지적 상태를 표현하지만, 감정이 도덕 속성과 관계되므로 판단의 참과 거짓을 가릴 수 있다. 도덕 판단에는 더 옳은 감정 반응이 있다. 따라서 프린츠는 조이스와 더불어 도덕 판단에는 인지적 요소와 비인지적 요소가 모두 있다고 본다. 프린츠의 감수성 이론에 대해서는

13) 여기서 '감정'으로 번역한 'emotion'은 '정서', '정감' 등으로도 번역될 수 있다. 이 글에서는 에이어의 '정서주의(emotivism)'와 차별하려고 'emotion'의 번역어로 우리가 일상에서 주로 사용하는 '감정'을 택하겠다.

다음 절에서 자세히 다루겠다.

둘째는 인식적 논제로, 도덕 개념은 본질적으로 감정과 관련된다고 주장한다(Prinz 2007: 16). 인식적 논제는 우리가 도덕 속성을 어떻게 아는가라는 물음에 답한다. 다시 색을 예로 들어 보자. '빨강'이라는 색 개념은 맹인도 가질 수 있다. 맹인은 문장 속에 포함된 색 단어의 의미를 나름대로 익힐 수 있다. 그러나 이는 보통 사람이 '빨강' 개념을 습득하는 방식과 판이하다. 맹인이 떠올리는 색 개념은 사물의 진정한 속성을 봄으로써 얻어지지 않았다. 따라서 맹인은 정말로 색을 알지 못한다. 도덕 속성도 매한가지다. 프린츠는 도덕을 아는 다른 방식이 있을 수 있다고 인정한다. 하나 진정한 도덕 속성은 오직 감정으로만 감지할 수 있다. 이때 도덕 속성이 감정에 의존한다면, 도덕 속성에서 발원하는 도덕 개념 역시 감정에 기반을 둔다. 도덕적 '옳음'이나 '그름'이라는 개념에 포함되는 개별적인 사례들은 감정적 상태로 이루어진다. 그렇다면 우리는 도덕 개념을 파악함으로써 상응하는 도덕 속성을 인식할 수 있다. "도덕 개념이 도덕 속성을 지시하고 도덕 속성이 감정과 구성적으로 관련 있다면, 도덕 개념을 이해하는 일 또한 감정과 결부된다는 생각이 합당하다(Prinz 2007: 15-16)." 요컨대 감정은 도덕 개념과 속성을 알아보는 핵심 요소다.

마지막으로 도덕과 감정이 본질적으로 관련된다는 말은 두 가지 의미를 지닌다. 하나는 감정이 도덕 속성과 도덕 개념을 구성한다는 뜻이다. 도덕 속성과 도덕 개념은 감정 반응의 결과이다. 이렇게 도덕 속성과 도덕 개념이 감정으로 형성되면 이들은 또한 감정 반응을 일으키는 원인으로 작용한다. 감정은 동기적 힘을 갖기 때문이다. 다른 하나는 도덕과 감정이 성향적 관계라는 뜻이다. 도덕은 당장 느끼는 감정이 아니라 특정 감정을 느끼게 하는 성향으로 정의된다. 예컨대 누군가 비도덕적인 행동을 보고도 피곤해서, 우울해서 등 여러 이유로 그에 맞는 감정을 느끼지 못할 수 있다. 그렇더라도 그가 이를 용인한다고 보면 잘못이다. 옳지 못한 행동에는 부정적 감정을 일으킬 수 있는 성향적 속성이 늘 존재한다(Prinz 2007: 19).

프린츠는 자신의 입장을 뒷받침하기 위해 현대 도덕 심리학의 경험 연구들을 제시한다. 뇌 신경이미지 연구는 도덕 판단과 뇌의 감정 영역이 긴밀히 연결됨을 보여준다. 도덕 판단에는 주로 감정 반응을 유발하는 복내측 전전두피질, 전대상피질, 편도체 등이 관여한다(Greene & Haidt 2002). 사람들은 일반적인 문장, "돌은 물로 이루어져 있다."보다 도덕과 관련된 문장, "그들은 무고한 사람의 목을 매달았다."를 접했을 때 뇌의 감정 영역이 더 활성화되었다(Moll et al. 2003). 도덕 판단과 감정 간의 인과적 관계는 도덕 판단의 인지적 성격을 중시하는 논자들도 동의한다. 그들은 흔히 도덕 판단이 먼저 일어나고 그 영향으로 감정이 발생한다고 생각한다.

그러나 프린츠는 감정이 도덕을 구성한다고 주장한다. 감정은 단지 도덕 판단에 곁따르는

현상이 아니다. 감정은 또한 도덕 판단을 일으킨다. 휘틀리(T. Wheatley)와 하이트(J. Haidt)는 피험자들에게 최면을 걸어 '받다(take)'와 '종종(often)'이라는 단어에 역겨움을 느끼도록 조건화했다. 그 뒤 연구자들은 피험자에게 두 단어 중 하나가 포함된 이야기를 들려주었다. 피험자들은 도덕적으로 중립적인 일화에서도("댄은 토론 활성화를 위해 교수와 학생 모두 흥미를 가질 수 있는 주제를 선택하려고[take] 노력한다.") 조건화된 단어에 반응하여 사례에 등장하는 주인공이 도덕적으로 문제가 있다고 판단했다. 게다가 피험자들은 자신의 직감을 정당화하려고 댄이 나쁜 이유를 지어냈다(Wheatley & Haidt 2005). 또한 사람들은 주변 환경이 더럽거나 악취가 나는 경우, 그렇지 않을 때보다 같은 사례에 대해 더 엄격한 도덕 판단을 한다(Schnall et al. 2008).

이상의 경험 연구들은 도덕 개념이 감정과 관련된다는 인식적 논제를 지지한다. 인간은 도덕과 무관한 상황에서도, 도덕과 무관한 요인으로도 감정이 촉발되면 도덕 판단을 한다. 감정은 도덕 판단을 야기한다. 이는 도덕 개념이 감정적 요소를 가져, 감정 유발이 도덕 개념을 적용하는데 영향을 끼치기 때문이다(Prinz 2007: 28). 도덕 판단에서 감정은 일종의 정보로 활용된다. 어떤 사안을 평가할 때 사람들은 자신의 내면으로 눈을 돌려 느낌이 어떤지 살핀다. 긍정적인 감정을 느끼면 그 행동은 옳고, 그렇지 않으면 그르다(Haidt 2012/2014: 127).

감정이 도덕을 구성한다는 주장은 급진적이다. 통념과 반대로, 도덕 판단이 합리적 이유나 이성의 활동과는 거리가 멀다고 말하기 때문이다. 도덕 판단은 단지 감정적 성향이 있느냐의 문제이다. 감정은 도덕적 태도를 갖는 충분 조건이다. 우리는 이치에 맞는 정당화가 부재해도 도덕적 태도를 가지며 도덕 판단을 한다. 사람들의 반성적 도덕 판단에는 도덕 이유보다 먼저 감정적 토대가 있다(Prinz 2007: 29). 이를 뒷받침하는 증거는 도덕 판단에 대한 하이트의 '사회적 직관주의 모델(social intuitionist model)'이다. 하이트는 사람들이 특정 도덕 영역에서 직관에 따른 매우 즉각적인 판단을 내리며 이성적 추론은 사후적으로만 일어난다고 주장한다(Haidt 2012/2014). 하이트는 피험자에게 근친상관에 대한 일화를 들려주었다. 남매가 별장으로 여행을 떠났다. 둘은 호기심에 합의하고 성관계를 맺었다. 남매는 피임했고, 성교를 즐겼으며, 다시는 하지 않기로 약속했고, 둘만의 비밀로 삼기로 했다. 사실상 이 사례에서 도덕적으로 문제가 될 만한 요소는 없다. 그러나 대부분의 사람들은 남매의 행동이 도덕적으로 옳지 못하다고 대답했다. 왜 나쁘냐고 묻는 연구자의 말에 사람들은 여러 가지 이유를 댔고 연구자는 이에 반론했다. 남매는 기형아를 낳을 수 있다. 그들은 피임을 했다. 사회에 안 좋은 영향을 준다. 남매는 이 일을 비밀에 묻었다. 서로 마음의 상처가 될 수 있다. 남매는 성관계를 즐겼으며 관계는 더 돈독해졌다. 결국 피험자들은 반론이 옳다고 인정했으나 오직 17%의 사람만이 처음의 판단을 바꾸었다. 남은 사람들은 여전히 왜인지는 모르지만 근친상간은 그냥 나쁘다, 역겹다고 진술했다(Haidt 2012/2014: 91-92). 이를 '도덕적 말막힘

(moral dumbfounding)'이라 부른다. 도덕적 말막힘 현상은 인간 이성의 무력함을 보여준다. 이성은 도덕 판단에서 별다른 역할을 하지 않는다. 그저 뒤늦게 와서 판단에 중요하지 않은 단편적 이유를 댈 뿐이다. 이미 옳고 그름은 감정에 따라 자동적으로 따라 나왔다. 하이트의 연구는 "이성은 정념의 노예."라는 흄의 선언을 지지한다.

그러나 프린츠는 도덕적 말막힘 현상이 아무런 이유 없이 생긴다고 해석하지 않는다. 다만 그 이유는 기초 가치에서 나온다. 기초 가치는 더 이상 설명할 수 없는 가치이다. 아동학대는 왜 나쁜가? 이런 질문은 이상하다. 아동학대는 그냥 나쁘다. "그건 그냥 나빠!"라고 말할 때 그 나쁨을 정당화하는 이유는 외부 사실에서 주어지는 게 아니라 인간 심리에 뿌리박힌 기초 가치에서 비롯한다. 따라서 "기초 가치는 이유를 제공하지만 이유에 의거하고 있지는 않다(Prinz 2007: 32)." 기초 가치는 감정에 바탕을 둔다. 인간은 어떤 도덕 문제에 대하여 직관적으로 옳고 그름의 감정을 느끼는 기초 가치를 갖고 있다. 기초 가치는 왜, 어떻게 생겼는가? 뒤에서 보겠지만 프린츠는 기초 가치가 자연선택으로 형성된 선천적인 적응이 아니라 감정이 문화와 결합하여 생긴 후천적 산물이라 생각한다.

도덕 판단이 감정적 성향을 표현한다는 입장은 동기 내재주의를 잘 설명한다. 도덕 판단은 어떤 행동에 '옳음', '그름'에 속하는 도덕 개념을 적용한다. 도덕 개념은 감정으로 구성된다. 감정에는 동기적 힘이 있다. 따라서 도덕 판단은 동기를 일으킨다. 그러나 윤리학자 브링크(D. O. Brink)는 '무도덕주의자(amoralist)'로 동기 내재주의의 단점을 지적한 바 있다(Brink 1989). 무도덕주의자는 자신이 옳다고 판단한 일을 실행할 아무런 동기도 느끼지 않는 사람이다. 무도덕주의자가 정말로 있다면 동기 내재주의는 논박된다. 현실에서 진짜 무도덕주의자로 자주 거론되는 사례가 있다. 바로 사이코패스다. 사이코패스는 보통 사람과 동일한 인지 능력을 보유하며 도덕 가치를 이해하는 듯 보인다. 적어도 우리 사회가 금지하는 행위가 무엇인지는 알고 있다. 그럼에도 사이코패스는 도덕적 실천에 무심하다. 쉽게 타인을 이용하고 버리며, 살인이나 고문, 시체 훼손도 서슴없이 저지른다. 반사회적 행동을 범할 때 어떤 거리낌도 없다. 무도덕주의자로서 사이코패스는 내재주의를 논박할까? 그렇지 않다. 사이코패스는 감정적 장애를 갖고 있다. 사이코패스의 감정 능력은 매우 비정상적이다. 사이코패스는 주요한 감정 반응(후회, 수치, 사랑 등)을 거의 나타내지 않고 특히 공포나 슬픔을 나타내는 얼굴 표정을 잘 인지하지 못한다. 사이코패스는 다른 사람이 겪는 고통에 무감각하다. 즉 사이코패스의 도덕적 무능력은 감정을 관장하는 뇌영역이 손상되어 발생한다. 사이코패스에 대한 경험 연구는 도덕적 사고와 행동을 지지하는 토대가 자신과 상대의 감정에 공감하는 능력임을 보여준다. 우리는 타인의 복지에 특별한 관심을 쏟는다. 그러한 동기와 행동의 근원에는 감정이 있다(McGeer 2008). 따라서 사이코패스는 내재주의를 논박하지 못한다. 사이코패스는 무도덕주의자가 아니다. 사이코패스는 진정으로 도덕을 이해하지 못한다. 오히려 사이코패스는

내재주의, 정확히 프린츠가 제시한 감정을 따르는 내재주의를 입증한다.

이러한 감정은 어디서 왔을까? 감정은 생존과 번식에 중요한 문제들을 해결하려고 자연선택이 설계한 적응 기제이다. 감정은 환경이 가하는 위협에 맞서 우리를 행동하게 만든다. 따라서 프린츠는 인간에게 행복, 슬픔, 공포, 역겨움, 분노, 놀람 등의 기초 감정들이 선천적으로 존재하며 이 감정들은 경험과 새로운 환경에 따라 때로 섞여 새로운 감정을 만들어 내거나, 다른 조건에서 전용되도록 배치되었다고 주장한다(Prinz 2007: 67). 도덕 감정이 그런 경우다. 도덕 감정은 도덕과 유관한 맥락에서 일어나는 감정이다. 도덕 감정은 기초 감정들에서 유래한다. 도덕 감정에는 타인을 향한 경멸, 분노, 역겨움 등이 있고, 자신을 향한 죄책감, 수치심 등이 있다. 우리가 어떤 도덕적 상황에 대해 승인과 불승인의 감정을 표현할 때 그 안에는 타인과 자신을 향한 도덕 감정이 내포해 있다.

구체적으로 감정은 어떤 방식으로 기능할까? 감정은 세계에 대한 어떤 평가나 판단일까? 프린츠는 이를 부정한다. 프린츠는 도덕 판단에 도덕 감정이 포함된다고 주장한다. 이때 감정이 외부 세상에 대한 어떤 생각이나 판단이라면 도덕 감정은 도덕 판단으로, 도덕 판단은 도덕 감정으로 정의되는 악순환이 발생한다. 프린츠는 다른 방법을 택한다. 감정은 신체 변화에 대한 지각이다(Prinz 2007: 56-60). 예를 들어 사람은 얼굴 표정을 바꾸기만 해도 심장 박동 같은 신체 변화가 일어나며 그에 걸맞는 감정이 야기된다(Levenson et al. 1990). 우리는 수치심을 느낄 때 얼굴이 빨개지고 입꼬리는 내려가며 몸은 움츠러든다. 그예 스스로를 다그치며 다시는 같은 행동을 반복하지 않으리라 다짐한다. 따라서 감정은 신체 변화가 보내는 신호이다. 이 신호에는 알맞은 행동을 야기하는 세계에 대한 정보가 들어 있다.

그런데 우리는 감정에 대해서 평가적인 용어를 사용한다. 특정 상황에서 어떤 감정은 적절하거나 비적절하고, 정당하거나 정당하지 않다. 뱀에 대한 공포는 정상적이고 귀여운 아기를 혐오하는 사람은 비정상적이다. 감정이 신체 변화에 대한 지각일 때도 감정의 합리성을 평가할 수 있을까? 프린츠는 그렇다고 말한다. 감정은 생존과 번식에 중요한 세계의 정보를 탐지하도록 진화한 장치이다. 근육 수축, 흐르는 땀 등 신체 변화로 일어난 감정은 개체와 환경 사이의 관계, 예를 들어 위험 등을 표상한다. 이는 마치 화재경보기의 알람 소리가 불을 표시하는 상황과 같다. 즉 감정은 나와 세계에 대한 '관심(concern)'을 표상한다. 프린츠는 감정이 관심을 나타낸다면 합당하거나 합당하지 않은 감정을 가릴 수 있다고 주장한다. 우리에 갇힌 뱀을 보고 두려움을 느낀다면 이는 적절하지 않다. 해를 가할 수 없는 뱀을 위험한 동물로 표상하는 상황은 오류이다. 그러나 자기 다리를 휘감고 있는 뱀을 보고 느끼는 두려움은 적절하다(Prinz 2007: 63-64).

프린츠는 감정의 본성과 기능에 대한 논의로 도덕 판단은 우리 마음의 투사가 아니라 지각이라고 결론 내린다. 우리는 감정으로 도덕 속성을 지각한다. 감정은 세계와 관계 맺는, 우리

몸에 일어나는 변화에 대한 지각이다. 따라서 내가 지각하는 승인과 불승인의 감정은 어떤 사태가 옳고 그른지 알려준다.

여기까지 조이스와 프린츠가 생각한 도덕 판단의 본성을 살펴보았다. 조이스는 무엇보다도 도덕 판단이 불가피한 권위를 가진 실천적 요구에 대한 믿음을 표현한다는 인지적 성격을, 프린츠는 도덕 판단이 감정적 성향을 표현한다는 비인지적 성격을 강조한다. 두 사람이 생각하는, 도덕 판단이 나타내는 핵심 요소의 차이는 도덕 사실에 대한 견해차와도 밀접한 관계를 맺는다.

3.2. 진화적 폭로 논증 대 감수성 이론

'진화적 폭로 논증'은 도덕이 자연선택의 산물이라면 도덕 판단의 참은 인식적으로 정당화되지 않는다고 주장한다. 우리는 객관적인 도덕 사실이 있는지 알 수 없다. 나아가 이러한 사실이 있다고 가정할 필요도 없다. 도덕은 허구이다. 도덕 판단의 인식적 정당성을 거부하는 조이스의 전략은 루즈와 유사하다. 도덕 믿음의 기원에 대한 지식은 그 믿음의 정당성을 훼손한다.

조이스는 사고 실험 하나를 제안한다. 여기 "나폴레옹은 워털루 전투에서 패배했다."라는 믿음을 생기게 하는 알약이 있다고 하자. 그리고 이러한 믿음을 없애는 해독제도 있다. 이제 "나폴레옹은 워털루 전투에서 패배했다."는 나의 믿음이 과거에 먹은 알약 때문에 생겼다고 하자. 내가 이를 알았다면 나폴레옹에 대한 믿음은 손상되는가? 조이스는 그렇다고 말한다. 물론 내가 목도한 진실이 곧바로 이 믿음을 거짓으로 만들지는 않는다. 다만 나폴레옹에 대한 믿음은 정당화되지 않는다. 나는 나폴레옹에 대한 참된 지식을 얻을 때까지 믿음을 철회해야 한다. 따라서 해독제를 먹어야 한다. 그러면 나는 나폴레옹이 지휘한 워털루 전투에 대해 불가지론자가 된다(Joyce 2006: 179-180).

"나폴레옹은 워털루 전투에서 패배했다."는 믿음이 직간접적으로 그에 맞는 사실로 생기지 않았다면 이 믿음은 정당화되지 않는다. 나폴레옹에 대한 믿음은 나폴레옹이 수행한 워털루 전투와 상관없는 사실 덕분에 생겼다. 즉 어떤 믿음이 왜 생겼는지를 알면 그 믿음의 정당성이 훼손되는 경우가 있다. 도덕 믿음이 그렇다.[14] 이때 나폴레옹에 대한 믿음은 도덕 믿음이고 믿음을 만드는 알약은 자연선택이다. 물론 자연선택이 반드시 개체를 기만하는 작용은 아니다. '1+1=2'라는 산술 믿음이나 나무에 과일이 몇 개 달렸는지 분간하는 시각 경험은 참이 아닐 경우 번식 적합도를 떨어뜨리기 때문에 세계와 대응하는 참이어야 한다. 그러나 도덕 믿음은 다르다. 도덕 사실에 대한 지각없이 옳음이나 그름과 관계된 믿음을 가지는 상태는 유용하다. 도덕 믿음은 도덕 사실에 근거하여 생기지 않았다. 우리가 옳다고 생각하는 행위는 그저 조상들의 유전자 사본을 잘 복제하는 데 도움이 되었을 뿐이다. 도덕은 진화의 산물이다. "도덕과 무관한 입증된 계보학(진화론을 말한다-인용자)을 사용할 수 있다는 사실은 우리가 도덕 믿음이 참이라고 생각할 어떠한 이유도 없음을 보여준다(Joyce 2006: 190)."

---

14) 한 가지 주의할 점이 있다. 조이스는 특정한 도덕 믿음이 가진 내용이 자연선택 덕에 선천적으로 주어졌다고 보지 않는다. 그는 세계를 도덕적인 용어로 범주화하는 도덕 개념이 선천적이라고 생각한다. 상황이 이렇더라도 회의적 결론은 바뀌지 않는다. 알약이 나폴레옹에 대한 일반적 개념을 생기게 한다고 해보자. 알약을 먹어서 생긴 나폴레옹과 관련한 개념은 이 개념과 결부된 믿음을 형성하게 만든다. 물론 이 믿음은 정당화되지 않는다. 나는 해독제를 먹어야 한다(Joyce 2006: 180-181).

조이스는 여기서 그치지 않는다. 그는 도덕 판단의 본성이 불가피한 권위를 가진 도덕 요구에 대한 믿음을 표현하는 데 있다고 주장했다. 도덕 판단은 욕구나 소망과 무관하게 따라야 할 정언적 이유를 나타낸다. 따라서 도덕 사실은 정언적 이유를 주어야 한다. 한데 도덕 판단의 실천적 힘은 진화의 산물이다. 실제로 정언적 이유는 없다. 남은 길은 오류 이론이다. 도덕 판단은 모두 거짓이다. 그렇다면 도덕 자연주의는 도덕 판단의 정언성을 수용할 수 있을까, 도덕 판단의 불가피한 권위를 자연 속성으로 설명할 수 있을까, 이로써 자연주의적인 도덕 실재론을 구할 수 있을까? 조이스는 그럴 수 없다고 말한다.

조이스는 도적 자연주의자들을 두 집단으로 나누어 비판한다. 첫 번째 집단은 도덕의 불가피한 권위를 자연주의적 틀 내에 수용할 수 있다고 주장하는 집단, 두 번째 집단은 불가피한 권위를 거부하는 집단이다. 첫 번째 집단부터 살펴보자.

어떤 도덕 자연주의자는 “침을 뱉지 마라.”와 같은 에티켓 규칙도 욕구와 무관한 행위 이유를 준다고 생각한다. 에티켓은 행위자가 침을 뱉지 못하도록 요구한다. 따라서 욕구와 상관없이 침을 뱉지 말아야 할 이유가 있다. 이렇게 에티켓이라는 관습이 정언성을 가진다면, 도덕 또한 우리가 정언적으로 따르도록 합의한 규범 체계일 수 있다. 그러나 조이스는 단지 제도적인 정당화를 제공하는 규범적 틀로는 도덕 판단의 불가피한 권위를 설명할 수 없다고 반박한다. 도덕의 실천적 힘은 특별하다. 도덕은 그 어떤 권위에도 의존하지 않는다. 도덕 명령은 제도를 초월하여 우리와 매우 밀접히 묶여 있다. 설사 우리가 절도를 용인하기로 결의했어도 절도는 나쁘다는 판단은 사라지지 않는다. 또한 나중에 다시 보겠지만, 에티켓이 늘 정언적 이유를 주지는 않는다. 에티켓을 지키려는 마음이 없거나 그보다 더 긴급한 사안이 있다면 에티켓은 무시될 수 있다. 누군가 위험에 빠졌다면 큰 소리로 고함을 쳐 남에게 불편을 끼치더라도 그를 도우리라. 결론적으로 도덕 자연주의자는 제도 내에서 만들어진 불가피한 권위가 아니라 실제로 그런 힘이 있음을 보여야 한다(Joyce 2006: 194).

도덕 자연주의자는 다른 방법을 제시한다. 도덕 요구는 우리가 정말로 해야 할 이유가 있는 행위와 관련된다. 따라서 우리는 어떤 사태를 잘 숙고할 때 무엇을 해야 하는지에 대한 충분한 이유를 얻을 수 있다. 윤리학자 하만(G. Harman)은 이 전략을 ‘실천적 추론 이론(practical reasoning theory)’이라 부른다. 실천적 추론 이론은 한 행위자가 적절하게 추론했을 때 하도록 원하게 될 행동은 그가 그 행동을 할 충분한 이유가 있는 상황과 같다고 주장한다(Harman 1986: 66). 실천적 추론 이론은 이유를 제공하는 도덕의 실천적 권위는 설명할 수 있을지도 모른다. 행위자가 심사숙고해서 나온 이유는 분명 실천적 권위를 가진다. 그러나 실천적 추론 이론은 도덕 판단의 불가피함, 즉 욕구나 바람과 상관없는 이유를 자연적으로 설명할 수 있는가? 조이스는 그럴 수 없다고 말한다. 적절하게 추론했을 때 원하게 될 무엇은 각자의 욕구에 따라 서로 다를 수 있다. 그래서 적절한 추론은 저마다의 득실과 독립적인 도

덕적 규정을 주지 못한다(Joyce 2006: 196).

적절한 추론이 개인의 욕구에 민감하다는 사실은 도덕적이지 않은 상황을 타당하게 만들 위험이 있다. 고양이에게 불을 붙이는 행동이 고양이를 괴롭히고 싶어 적절하게 사고하여 나왔다면, 적절한 추론이 어떻게 효과적으로 고양이를 괴롭힐지 찾게 한다면, 도덕 자연주의자는 이를 도덕적으로 옳다고 여길 셈인가? 결국 실천적 추론이 도덕 명령에서 기대하는 바와 같이 욕구와 상관없는, 정언적 요구로 나타나지 않는다면 도덕 자연주의는 실패한다(Joyce 2006: 196). 조이스가 보기에 지금까지 도덕의 불가피한 권위를 설명하는 도덕 자연주의의 어떠한 시도도 성공하지 못했다. 도덕 자연주의는 기본적으로 가언적인 체계이기 때문이다. "요점은 이렇다. 내가 따라야 할 이유는 어느 정도 나의 실제 욕구, 득실, 기획, 목적에 좌우된다. 그러한 다양한 사람들은 같은 상황에서 서로 아주 다른 이유를 가질 수 있다[하만이 말했듯이, 행위자의 이유는 그의 욕구에 "근원한다"](Joyce 2006: 198)."

다음으로 도덕의 불가피한 권위를 거부하는 두 번째 유형의 도덕 자연주의자를 보자. 그들은 특정 자연 속성이 사람들에게 항상 도덕적인 행위 이유를 준다고 보증할 수는 없지만 그래도 자연 속성과 행위 이유는 신뢰할만한 우연적 관계에 있다고 주장한다. 예를 들어 공리주의자들은 행복을 최대화하는 행동과 그러한 행동을 해야 할 이유 사이에 필연적인 연결을 포기하고 우리가 일반적으로 복지에 대해 신경 쓰며, 이 사실이 행복을 증진할 이유와 동기를 제공한다고 말할 수 있다(Joyce 2006: 200). 이런 상황에서 도덕 판단은 그 고유의 기능을 수행할 수 있는가?

조이스는 도덕에 실천적 힘이 없다면, 그래서 사람들이 가진 이유와 우연하게만 연결된다면, 이 또한 에티켓과 유사하다고 본다. 에티켓은 단지 관습에 불과하다. 행위자는 에티켓이 정한 규칙을 지켜야 하지만 그럴 필요가 없을 때는 언제든 원하는 대로 행동할 수 있다. 내가 혼자 집에서 마구 소리를 내며 음식을 게걸스럽게 먹는다고 하자. 누가 나를 본다면 이런 행동을 상스럽다고 생각하리라. 그러나 혼자 있는데 무슨 상관인가? 왜 에티켓을 따라야 하는가? 물론 에티켓의 관점에서 나는 비난받을 짓을 했다. 나는 이를 알고 있다. 그럼에도 나는 에티켓이라는 기준 틀을 무시할 수 있다. 조심히 밥을 먹어야 할 어떤 진정한 이유도 없다. 조이스는 이러한 논증을 도덕에도 적용한다. 나는 정말로 살인을 하고 싶은 욕구가 있다. 마침 기회가 좋아 처벌받지 않을 상황이 생겼다. 두 번 다시 같은 기회는 오지 않는다. 나는 아무런 두려움도 느끼지 않는다. 이때 나는 살인을 하지 말아야 하는가? 나의 살인 욕구는 도덕적 관점에서 비난받아 마땅하다. 나는 이를 알고 있다. 그러나 나는 이 특정한 준거 틀, 도덕이라는 규범 체계를 무시할 수 있다. 요컨대 도덕은 살인을 하지 말아야 할 제도 초월적인 진정한 이유를 주지 못한다(Joyce 2006: 202-204).

도덕 판단의 실천적 힘이 제거된 대가는 너무 크다. 나는 도덕 명령에 전혀 개의치 않거나

혹은 이익이 될 때는 따르고, 그렇지 않을 때는 묵살하는 기회주의자가 될 수 있다. 이에 대해 도덕 자연주의자는 도덕 요구에 실천적 힘이 없을지라도 우리에게 도덕적으로 그른 행동을 피하려는 강한 욕구가 있다고 반론할 수 있다. 그런 욕구는 도덕 판단의 내용을 준수하려는 한결같은 이유를 산출한다. 우리는 대체로 남의 물건을 훔치지 말아야 하고, 약속을 지켜야 하는 좋은 이유가 있다고 생각한다. 도덕 자연주의자는 이러한 일상적인 도덕 의견과 연결된 욕구에, 도덕적 그름과 동일시되는 자연 속성이 예화된다고 생각한다. 그렇다면 조이스는 과연 어떤 자연 속성이 도덕적 그름이라는 속성을 지시하는지 알아야 한다고 되묻는다(Joyce 2006: 206-207).

도덕적 그름과 동일시되는 자연 속성이 불필요한 고통을 피하려는 욕구라고 해보자. 어떠한 행동이 사람들에게 고통을 초래한다면 그 행동은 도덕적으로 그르다. 따라서 행위자는 고통을 삼가는 강한 욕구를 갖는다. 이로써 문제가 해결되는가? 조이스는 그렇지 않다고 말한다. 도둑질을 하지 말아야 할 이유가 불필요한 고통에서 온다면 내가 가진 행위 이유나 동기를 도덕 용어로 명명하는 일은 전적으로 잉여이다. 도덕적 그름은 나의 행동에 무엇도 추가하지 않는다. 나는 다만 고통을 피할 뿐이다. 그러나 어떤 행동을 "도덕적으로 그르다."라고 판단한다면, 도덕 자연주의자는 그 행동이 특정 자연 속성, 즉 불필요한 고통을 피하려는 욕구를 예화한다는 생각을 넘어서 무언가 실질적인 속성을 보여야 한다. 이러한 주장은 도덕 자연주의자에게 다음과 같은 질문을 제기한다. 당신의 이론에 따르면 왜 우리는 도덕적 담화를 다른 담화와 구별해야 하는가, 왜 우리는 도덕적 담화를 단순히 좋아함이나 싫어함, 사회적 조화에 유익한 행동이나 그렇지 않은 행동으로 생각해서는 안 되는가(Joyce 2006: 207)? 조이스는 이미 '~해야 한다'라는 도덕 판단과 무엇을 원한다는 욕구를 분명히 구분한 바 있다.

> 이 책의 앞장에서 나는 "왜 우리에게는 독특한 도덕적 담론이 필요한가?"라는 질문에 답했다. 도덕적 사고와 대화는 의지박약을 방지하는데 도덕과 무관한 실천적 고려보다 더 낫다. 또한 도덕은 상호헌신 장치이다. 도덕 자연주의를 비판하는 내 논증의 핵심은 도덕적 사고는 특별한 종류의 계산 중지[15]이며 욕구 초월적인 실천적 힘을 가진다는 사실이다. 이는 냉철한 타산적 계산보다 더 잘 유혹을 이겨낸다. 그러나 실천적 힘을 부정하는 도덕 자연주의자는 이를 수용할 수 없다. 도덕 자연주의자에게 도덕적 숙고는 단지 내가 욕구하는 게 무엇이며 이를 어떻게 성취할 수 있는가에 대한 고려이기 때문이다(Joyce 2006: 208).

15) 조이스는 도덕 판단과 개념이 이른바 '중지 기능(silencing function)'을 한다고 주장한다. 이는 어떤 행동이 도덕적으로 그르다는 사실을 이해하면 그 행동을 하려는 합리화나 내적 교섭 등의 계산적 사고를 중지한다는 뜻이다. 행위자는 타산적 고려를 즉시 멈추고 도덕 이유를 따르고자 한다(Joyce 2006: 111).

요컨대 불필요한 고통을 피하려는 욕구에는 도덕 이유를 다른 타산적 이유와 구별해 주는 독특한 성격이 없다. 따라서 어떤 도덕 자연주의도 도덕의 실천적 힘을 설명하지 못한다. 불가피한 권위라는 도덕 믿음은 진화의 산물이다. 실제로 그러한 사실은 없다. 종내 도덕 자연주의는 오류 이론으로 귀결된다. 도덕 판단은 모두 거짓이다. 그런데 도덕적 참이라는 허구가 폭로되면 우리는 어떻게 되는가, 도덕에 객관적으로 옳은 기준은 없으므로 무엇이나 다 가능하다는 회의주의에 빠지게 되는가? 조이스는 그렇게 보지 않는다. 도덕은 환상이지만 도덕적 담화는 여전히 실천적 역할을 한다(Joyce 2000: 727; 2006: 224). 도덕은 유용한 허구이다. 우리는 도덕의 실천적 권위를 믿지 않지만 도움이 되기에 받아들인다. 소설을 생각해보자. 독자는 소설의 내용을 믿지 않지만 진실한 감정을 느끼고 이야기에 동화된다. 똑같이, 행위자는 도덕을 믿지 않고도 진정한 도덕 감정을 경험할 수 있다. 그래서 조이스는 참과 연관된 도덕 믿음이 없이도 도덕 감정이 행위 이유를 준다고 생각한다(Joyce 2006: 226-227). 도덕 판단은 참으로 정당화되지 않지만 감정 덕분에 이 유용한 허구를 실천한다.

반대로 프린츠는 도덕 사실이나 속성이 우리 감정적 성향과 연결된다고 주장한다. 도덕 속성이나 사실은 감정 반응으로 구성되며, 감정 반응으로 판단의 참과 거짓을 가릴 수 있다. 도덕 속성은 반응 의존적 속성이다. 우스움이라는 속성을 생각해보자. 어떤 사람, 행위, 상황의 우스움은 우리를 즐겁게 하는 성향이 있을 때 참이다. 우스움이라는 속성의 존재는 분명 관찰자가 어떤 반응을 나타내느냐에 달려있다. 도덕도 동일하다. 어떤 사람, 행위, 상황의 그름은 관찰자에게 불승인의 감정을 불러올 경우 그르다(Prinz 2008a: 225). 그러므로 살인은 나쁘다는 판단은 단지 감정 표현을 넘어, 살인에 부정적 감정을 야기하는 속성이 있다는 주장이다. 부정적 반응은 도덕적 그름이라는 속성을 나타내며, 도덕적 그름이라는 속성은 감정 반응이 옳음을 정당화한다. 도덕 문제에서 특정한 반응 없이 무엇이 옳고 그른지 집어낼 수 없다. 이렇게 도덕 판단은 반응에 의존하여 진리치를 가진다. 프린츠는 자신의 감수성 이론을 다음과 같이 정식화한다.

> (S1) **형이상학적 논제** 어떤 행동은 특정 조건에서 보통의 관찰자에게 승인(불승인)의 느낌을 야기할 경우 도덕적으로 옳은(그른) 속성을 가진다.
> (S2) **인식적 논제** S1에 언급된 감정을 느끼는 성향은 일반적인 올바름(그름) 개념의 보유조건이다(Prinz 2007: 87).

앞에서 현대 도덕 심리학의 경험 연구가 감수성 이론의 인식적 논제를 지지함을 보았다. 어떤 행동을 올바르다고 판단할 때, 즉 '올바름'에 속하는 도덕 개념을 적용할 때 그 판단에는 감정이 매개된다. 아울러 인식적 논제는 형이상학적 논제를 뒷받침한다. 도덕 개념은 감정으로 구성된다. 그리고 그 개념이 형성될 수 있었던 근저에는 역시나, 감정으로 구성되고 또

한 감정을 일으키는 도덕 속성이 있다. 따라서 도덕 개념은 도덕 속성을 지시한다.

감정 반응에 바탕을 둔 감수성 이론이 왜 실재론인가? 사람들은 대개 '실재'와 '객관성'을 연결 짓는다. 객관성은 또한 '마음 독립적인'과 유사한 의미로 쓰인다. 그래서 흔히 어떤 대상이 '마음 독립적으로 객관적'일 때 실재한다고 생각한다. 감수성 이론은 이런 의미의 실재론을 거부한다. 감수성 이론은 반응 의존적 실재론이다. 여기서 프린츠는 실재에 대한 관념을 바꿀 필요가 있다고 주장한다. 실재론은 다음 두 조건만 만족해도 충분히 성립한다. 첫째, 어떤 대상은 참과 사실을 가릴 수 있을 때 실재한다. 실재는 분명 어떤 영역에 대한 진술이 참이며 사실과 대응된다는 뜻이다. 이 점에서 감수성 이론은 진리치를 가진다. 아동학대는 나쁘다는 판단은 감정 표현인 동시에 그런 행동에 부정적인 감정을 야기하는 속성이 있다는 주장이다. 아동학대는 정말로 우리를 화나게, 실망스럽게 만든다. 따라서 아동학대는 나쁘다는 진술은 참이다. 둘째, 어떤 대상은 인과적 효과를 낼 때 실재한다. 존재하지 않는 무엇이 인과적 힘을 행사할 수는 없다. 도깨비나 이무기는 세계를 변화시키지 못한다. 따라서 세계에 인과적 변화를 일으키는 그 무엇은 실재한다. 감정과 연결되어 있는 도덕 속성은 도덕 행동을 이끈다. 우리는 역겨움을 느끼기에 썩은 음식을 버린다. 마찬가지로 아동학대가 일으키는 노여움과 죄책감은 아동학대를 금지하고 범죄자를 처벌하게 만든다(Prinz 2007: 165-166). 이 두 가지 의미에서 감수성 이론은 도덕 실재론이다.

이에 대해 조이스는 앞서의 비판처럼 감정과 연결된 도덕 속성이 잉여가 아니냐고 물을 수 있다. 행동을 추동하는 원인은 감정 그 자체이지 감정을 유발하는 속성이 아니라고 말이다. 속성을 추가함으로써 우리 행동에 무언가 실질적인 차이가 생기는가? 프린츠는 그렇다고 말한다. 감정과 감정을 유발하는 속성은 서로 다른 설명적 역할을 한다. 먼저 감정을 유발하는 속성 없이 감정만 있는 상황을 보자. 내가 냉장고에 상한 음식이 있다는 거짓 믿음을 가진다고 하자. 이때 상한 음식을 버린다면 그 음식이 실제로 역겨워서가 아니라 역겹다고 믿기 때문에 그렇다. 이런 믿음은 어디서 왔을까? 아마 유통기한을 잘못 보았거나, 음식 고유의 색깔을 곰팡이로 착각했을 수 있다. 여기서는 역겨움이 행동을 결정짓는 진정한 요인이 아니다. 단지 그렇게 보일 뿐이다. 그렇다면 이제 정말로 음식이 상한 경우를 보자. 이 사례에서 역겨움은 썩은 음식이 역겹다는 사실에서 발생한다. 나는 역겨움을 느끼고 음식을 버린다. 즉 역겨움이라는 감정을 일으키는 음식의 속성은 믿음을 만드는 원인으로 작용하며, 그 뒤에 일어날 행동까지 규정한다. 그런데 다시 역겨움은 음식이 상했기에 발생하지, 상한 음식은 역겹다는 사실 때문이 아니라고 반론할 수 있다. 프린츠는 이를 부정한다. 상한 음식이 정말로 역겹지 않다면 그 음식은 감정을 야기할 수 없다. 도덕 속성도 똑같다. 도덕적 그름이라는 속성은 부정적인 감정을 일으켜 특정 행동을 하도록 고무한다. 아동학대가 어떤 감정과도 연결되지 않는다면 우리는 학대를 꺼려하지 않으리라(Prinz 2007: 167).

프린츠는 자신의 감수성 이론이 맥키가 기이하다고 표현했던 도덕 사실의 객관적 규정성을 설명할 수 있다고 주장한다. 여러 번 말했듯이, 감수성 이론가들은 도덕 속성을 색 지각에 유비하여 설명한다. 색을 경험케 하는 사물은 우리 밖에 있다. 하지만 색은 주관과의 관계에서 성립한다. 우리 시각을 거쳐 비로소 색은 실현된다. 도덕 속성도 매일반이다. 도덕 감정을 야기하는 도덕 속성은 우리 밖에 있다. 도덕 속성은 외부에서 우리가 특정 감정을 느끼도록 만들고, 다른 한편 우리는 감정 반응으로 도덕 속성을 예화한다(Prinz 2007: 89). 이렇게 도덕 속성은 세계 내에 실제로 존재하는 객관적 대상이면서 우리 감정 상태로 구성된다.

감정은 또한 규정적 힘이 있다. 감정은 동기를 일으킨다. 그리하여 감정을 유발하는 도덕 속성도 동기적 효과를 낸다. 도덕적으로 그른 행동은 불승인의 감정을 촉발한다. 불승인은 타인이 무슨 행동을 했는지 기술하는 데 그치지 않고 처벌, 사과, 미래에 더 나은 행동을 하도록 동기화한다(Prinz 2007: 89).

우리가 도덕 판단을 할 때 적용하는, 감정으로 구성된 도덕 개념은 이 객관성과 규정성을 매개한다.

> 감수성 이론은 형이상학과 동기 사이에 개념적 다리를 놓음으로써 맥키에 대응한다. S1은 도덕 속성에 대해 말하며 S2는 도덕 동기에 대해 말한다. 그 방법은 이렇다. 개념은 속성을 표상한다. 개념은 법칙적으로, 신뢰할 만하게 속성을 예화한다. 도덕 속성은 우리에게 감정을 일으키는 힘이다. 우리는 이러한 속성들을 어떻게 마음에 표상하는가? 우리는 색이 우리에게 야기하는 색 경험으로 색을 표상한다. 마찬가지로 우리는 감정으로 도덕 속성을 표상한다. 도덕 개념은 도덕 속성이 일으키는 감정을 통합하며 그럼으로써 이러한 속성들을 탐지하는데 사용된다. 감정은 동기적이다. 그러므로 개별적인 도덕 개념은 우리를 행동하도록 만든다. 우리는 개념을 이용하여 형이상학과 동기를 연결한다(Prinz 2007: 89).

이제 감수성 이론을 채택하면 도덕 실재론 문제가 해결되는가? 그렇지 않다. 즉각적으로 떠오르는 의문은 (S1)에 제시된 '보통의 관찰자'와 '특정 조건'이 과연 무엇인가라는 점이다. 프린츠는 특정 조건에 온전한 인식이 가능한 조건 등 이상적 상황을 제시하지 않고 '감성(sentiment)'[16]이라는 개념으로 대체한다. 감성은 자주 언급했듯이, 감정을 느끼게 하는 성향이다. 감정과 감성은 구분해야 한다. 감성은 현재 일어나는 감정 상태가 아니라 유관한 감정을 느끼게 하는 가능한 상태이다. 프린츠가 감성을 도입하는 이유는 명확하다. 감정은 변동이 있다. 우리는 무기력하거나 우울할 때 살인을 목격해도 그에 맞는 감정을 느끼지 않을 수 있다.

16) 프린츠가 사용하는 'sentiment'는 '감정을 일으키는 성향'을 일컫는다. 'sentiment'에 통용되는 번역어는 없다. 이 글에서는 '감수성(sensibility)'과의 연속성을 살리고 성향의 의미를 보존하려 '감정적 성향'을 줄인 '감성'이라는 번역어를 채택하겠다.

그러나 이를 살인을 용인한다고 해석하면 안 된다. 살인의 그름은 부정적인 감정을 느끼게 하는 성향에 달려있지 감정 그 자체가 아니다. 살인에는 여전히 도덕적 분노를 부르는 감정적 성향이 있다(Prinz: 2007: 129). 프린츠는 도덕 판단을 하는 외부 사태와 더불어 행위자 또한 감성이라는 성질을 보유한다고 생각한다. 이때 감성은 지금 일어나는 상태(감정)로 드러나는 유기체의 고정된 상태(감정적 성향)이다. 쉽게 말하면, 감성은 정보 처리라는 근접 과정에서 노출되는 장기 기억과 유사하다(Prinz 2007: 84).

프린츠는 감성을 이용하여 도덕 판단이 일어나는 조건을 명시하는 문제를 우회할 수 있다고 말한다. 엄밀히 말해, 도덕 속성은 특정 조건에서 즉시 감정을 야기해서가 아니라 어떤 관찰자가 가진 감성, 즉 감정을 느끼게 하는 성향으로 구성된다(Prinz 2007: 91-92). 그러면 우리는 조건이 없어도 대체로 관찰자가 어떤 행동에 대해 그에 맞는 감정을 경험하리라고 예상할 수 있다. 설사 관찰자가 감정을 체험하지 못하더라도 감성은 계속 존재하여 여전히 우리가 연관된 감정을 겪도록 한다.

다음으로 어떤 관찰자인가? 프린츠는 '보통'이라는 단어는 평가적인 의미로, 정의해야 할 용어를 이미 전제하는 문제를 일으킨다고 보아 관찰자를 한정하지 않는다. 도덕 판단을 하는 '나'는 대개 자신을 정상적이라 생각한다. 따라서 관찰자는 '우리', '너', '나' 모두 해당된다. 이를 최종적으로 정식화하면 다음과 같다.

> (S1') 어떤 행동은 관찰자에게 그 행동에 대한 불승인(승인)의 감성이 있을 경우 도덕적으로 그름(옳음)의 속성을 갖는다(Prinz: 2007: 92).

관찰자에게 특정 행동을 향한 감성이 있다면 감성은 관찰자가 감정을 느끼도록 만든다. 그때 그 행동은 역시 감정을 일으키는 힘을 가진다(Prinz 2007: 92).

프린츠의 감수성 이론이 가지는 함축은 반직관적이다. 이렇게 특정 조건을 감성으로 대치하고 관찰자를 한정하지 않으면 도덕 판단의 객관성은 약화된다. 다시, 감성이 발현되는 조건은 무엇인가? 우리는 모든 개인, 집단들이 동일한 상황에서 똑같은 감정을 느끼게 하는 한결같은 감성을 가지리라고 기대할 수 없다. 또한 도대체 어떤 관찰자의 반응이 옳은 판단인가? 관찰자마다 어떤 행동에 대해 겪는 감정이 다르다면 도덕 속성도 다르다는 이상한 결론에 이르게 된다. 우리는 지금까지 감수성 이론이 객관적이라고 주장해 왔다. 도덕 속성이 반응 의존적이어도, 어떤 판단의 적절함과 비적절함을 가리는 보편적인 기준을 제시한다면 충분히 객관적이라고 말이다. 그러나 프린츠의 정식은 널리 통용되는 일반적 규준을 주지 못한다.

프린츠는 자신의 이론이 강한 상대주의로 향한다는 사실을 안다. 상대주의는 도덕 실재론에 대한 통념과 어긋난다. 우리는 도덕 사실이 존재한다면 도덕 판단에 정답이 있다고 생각한다. 그래서 도덕 판단은 객관적이다. 살인은 그르다는 판단이 어떻게 개인이나 집단에 따라

다를 수 있겠는가? 하나 프린츠는 이에 개의치 않는다. 오히려 자신의 입장을 더 밀고 나가 도덕 사실의 객관성을 송두리째 부정한다. 도덕 사실은 주관적이다. 도덕 사실은 우리 반응으로만 진리치를 가진다. 도덕 판단이 객관적으로 참이어야 한다는 직관은 틀렸다. 도덕 판단은 주관적으로 참이다. 감수성 이론은 주관적 실재론이다(Prinz 2007: 138). 주관성에 대한 옹호는 당연히 도덕 상대주의를 낳는다. 도덕 판단은 개인, 집단, 문화 등에 따라 상대적이다. 프린츠는 특히 인간의 감정적 성향을 만드는 사회적 자극으로서 문화의 역할을 강조한다. 각 문화마다 옳고 그르다고 판단하는 행동, 사람, 상황은 다양하다. 모든 곳에서 모든 사람이 같은 행위에 대해 동일한 감정 반응을 나타내지는 않는다. 따라서 프린츠는 도덕 사실의 일원성을 부정하고 다원주의를 지지한다. 이는 다음 절에서 자세히 논하겠다.

각자의 주관적 반응에 의존하는 프린츠의 감수성 이론은 조이스가 말한 불가피한 권위라는 도덕 판단의 실천적 힘을 설명할 수 있을까? 감수성 이론은 근본적으로 흄 전통에 따른다. 적절한 감정 없이 믿음만으로 행위를 위한 이유나 동기는 생기지 않는다. 따라서 프린츠는 칸트의 정언 명령이라는 개념 자체를 의문시한다. 도덕을 정념으로부터 해방하려는 시도는 값비싼 대가를 치른다고 말이다. 순수 이성이 어떻게 도덕적으로 행위 해야 할 이유를 주는가? 왜 이성의 명령을 따라야 하는가? 오로지 합리성만 갖춘 인간은 사이코패스와 같다. 감정에 무심하다면 우리는 도덕적으로 행동할 수 없다. 도덕 명령의 규정적 힘은 감정에서 발생한다(Prinz 2007: 134-135).

그럼에도 프린츠는 자신의 감수성 이론이 도덕의 실천적 힘을 설명할 수 있다고 주장한다. 조이스는 도덕 판단이 단지 감정이나 욕구와 관련된다면 다른 사람의 그른 행동을 비판할 수 없고, 도덕적이고자 하는 동기를 약화시키며, 도덕 속성의 존재가 불필요하다고 말했다. 프린츠는 이 세 가지 문제에 대답한다. 첫째, 대립하는 두 사람이 도덕 가치를 공유한다면 나는 상대가 실수를 저질렀을 때 그의 기준에서 도덕 위반을 비판할 수 있다. 물론 서로가 완전히 다른 도덕 가치를 가진다면 상대방의 책임을 물을 수 있는 권위가 없을 수 있다. 그러나 프린츠는 이런 결론은 수용 가능하다고 말한다. 우리는 도덕적으로 매우 다른 사회를 발견할 때 그런 집단을 이국적이라 부르지, 도덕적으로 나쁘다고 하지 않는다. 둘째, 나에게 진심으로 책을 훔치고 싶은 욕구가 있고, 책 절도에 대한 주저는 내게 심어진 죄책감을 느끼는 성향에서 비롯됨도 안다고 하자. 그래도 죄책감의 감정은 그 행위를 막는 데 충분하다. 셋째, 도덕 권위가 감정적 성향과 밀접하다는 관점은 절대로 도덕을 훼손하지 않는다. 우리는 부정직함이 그르다고 말할 때 그 행동이 분노를 일으킨다고 주장하고 있다. 그리고 이런 사실은 행동을 인도하며, 처벌하려는 태도를 만든다. 따라서 감수성 이론의 반응 의존적인 설명은 도덕을 훼손하기는커녕 왜 도덕 사실이 실천적 함축을 가지는지 설명한다(Prinz 2008a: 226).

도덕 사실의 실재는 마음 독립적인 사실이 아니라 우리가 구성한 사실이 실재한다는 뜻이

다. 프린츠는 화폐를 예로 든다. 화폐와 화폐가 나타내는 가치는 우리가 어떤 목적으로 사용하려 만든 사실이다. 그럼에도 "이 화폐는 천원이다."라는 진술은 참이며, 천원이라는 화폐 가치는 우리 삶에 영향을 미친다. 즉 화폐는 사용과 반응 덕분에 실재하는 사회적 사실이다. 도덕은 화폐처럼 구성된 사실이다. 도덕 사실은 우리 도덕 감성에 달려있다. 프린츠는 이를 '구성적 감성주의(constructive sentimentalism)'라 칭한다(Prinz 2007: 167).

프린츠는 구성적 감성주의가 도덕이 감정의 투사인지, 아니면 감정의 지각인지를 놓고 벌어진 대립을 해소할 수 있다고 주장한다. 그에 따르면 지각과 투사 모두 일정 부분 옳다. 우리는 도덕 속성을 지각한다. 감정은 몸에 일어난 변화를 느끼는 상태다. 이로써 감정을 일으킨 외부 사태가 옳은지 그른지 판단한다. 하지만 도덕 속성은 우리 마음의 투사이기도 하다. 도덕 속성은 초원에 핀 꽃이나 날아다니는 꿀벌과 달리 마음에서 비롯된다. 즉 도덕 속성은 감성으로 구성되었다. 하나 일단 만들어지면 지각된다. 마치 건축된 빌딩을 지각하듯 말이다. 우리는 죄 없는 어린이가 학대당할 때 그 행동의 잘못됨을 감정으로 지각한다. 그리고 학대가 나쁘다는 감정 반응은 우리 감성이 구축했다(Prinz 2007: 167-168).

지금까지 도덕 사실의 존재를 두고 벌어진 조이스의 진화적 폭로 논증과 프린츠의 감수성 이론 사이의 대립을 살펴보았다. 조이스는 도덕 사실이 마음 독립적이라고 생각한다. 그렇다면 우리는 거추장스러운 도덕 사실을 가정할 필요가 없다. 도덕은 세계의 사태와 무관한 진화의 산물이다. 도덕은 단지 생존과 번식에 도움이 되었던 생물학적 적응일 뿐이다. 또한 조이스는 도덕 사실을 자연 사실로 설명하려는 도덕 자연주의의 모든 기획은 실패라고 단정한다. 도덕 판단은 불가피한 권위를 가진 정언 명령이다. 도덕 사실은 정언적 이유를 주어야 한다. 그러나 정언적 이유는 없다. 정언적 이유가 있다는 믿음도 진화의 소산이다. 고로 도덕 판단은 모두 거짓이다. 이에 프린츠는 도덕 사실이 마음 독립적임을 거부한다. 도덕 사실은 반응 의존적이다. 도덕 사실은 감정 반응으로 구성된다. 어떤 행동은 그 행동을 승인, 불승인하는 감정적 성향, 즉 감성이 있을 때 도덕적으로 옳은, 그른 속성을 가진다. 더불어 도덕 사실은 정언적 이유가 아니다. 감정 반응 없이 행위 이유는 생기지 않는다. 프린츠는 감수성 이론을 강하게 밀어붙여 도덕의 객관성을 거부하고 도덕 사실은 주관적 사실이며, 도덕 판단의 내용은 문화마다 다양하다는 도덕 상대주의를 주장한다. 두 사람의 이러한 차이는 도덕과 도덕 판단을 하는 능력이 생물학적으로 설계된 선천적 경향인지 아니면 학습과 문화 덕에 발달한 후천적 행동인지를 둘러싼 논쟁으로 연결된다. 조이스는 자신의 진화적 도덕 반실재론을 강화하려고 도덕 선천주의를, 프린츠는 주관적 실재론을 옹호하려고 도덕 비선천주의를 지지한다.

3.3. 도덕 선천주의 대 비선천주의

도덕 선천주의는 도덕이 '생득적(innate)'이라 주장한다. 생득적이란 의미는 적어도 세 가지로 나눌 수 있다. 첫째, 어떤 형질이 유전자로 결정되어 환경과 학습에 상관없이 날 때부터 발현된다는 뜻이다. 둘째, 어떤 형질이 자연선택의 산물이라는 뜻이다. 즉 그 형질은 우리 조상들의 적합도를 증진시켰던 적응이다. 셋째, 본질주의적인 개념으로 어떤 형질이 한 종의 구성원에게 전형적으로 나타난다는 뜻이다. 이 분류들은 서로 다르다. 다운 증후군은 유전적으로 결정되어 있고 환경 변이에 상관없이 태어날 때 나타난다. 그럼에도 다운증후군은 적응이 아니다. 반면 도구를 만드는 행동은 부모에게서 자식으로 전달되고 번식 성공도를 높인다는 점에서 적응이다. 그러나 이 행동을 하는 데 학습이 필요하지 않거나 발달상 고정된 형질은 아니다(Joyce 2013: 550). 조이스가 주장하는 도덕 선천주의는 도덕 판단이 다윈주의적 적응이라는 뜻이다. 이를 적응적 선천주의라 부르자.

조이스는 도덕 믿음이 아니라 도덕 판단을 하는 경향, 즉 세계를 도덕 개념으로 파악하고 도덕 판단을 불가피한 권위를 가진 명령으로 사고하는 특성이 선천적이라 주장한다. 도덕 판단은 문화를 막론하고 보편적으로 나타난다. 조이스가 제시하는 증거를 보자.

첫째, 도덕은 우리가 아는 모든 인간 사회에 존재하며 도덕적 계율은 이집트의 『사자의 서』, 메소포타미아의 『길가메쉬 서사시』 등에 나타나듯 고대부터 있었다. 도덕이 현대 문명 사회의 문화적 발명품이라는 증거는 없다(Joyce 2006: 135).

둘째, 도덕 판단의 핵심 내용은 문화마다 유사하다. 도덕은 기본적으로 사회적 행동을 효과적으로 관리하는 기제로 진화했다. 그래서 도덕 판단이 적용되는 범주는 한정되어 있다. 전형적인 영역은 다음과 같다. (1) 다른 사람을 해치는 행동. (2) 호혜성, 공정성과 관계된 가치. (3) 사회적 위계질서에 따르는 행동. (4) 신체적 문제에 대한 규제(월경, 성교 등). 마지막 네 번째 항목을 제외하고 앞선 세 항목은 사람들 간의 관계, 사회적 질서의 유지와 관련 있다. 즉 도덕은 무엇보다도 타인을 향해 있다(Joyce 2006: 65).

셋째, 도덕 판단을 하는 경향은 아주 어릴 때부터 명시적인 가르침이 없이도 발달한다. 여기서 조이스가 제시한 근거는 이른바 '자극의 빈곤 논증(poverty of stimulus arguments)'이다. 인간은 적절한 자극이 부족해도 언어를 학습하는 보편 문법을 생득적으로 보유한다. 도덕 선천주의자들은 이에 착안하여 어린 아이가 언어처럼 도덕을 학습하는 도덕 문법이 있다고 주장한다(Mikhail 2009). 성인 수준의 경험이 없는 어린 아이들도 단지 타인을 친절하게 대하는 수준을 넘어 옳고 그름의 개념을 금방 이해한다. 물론 조이스는 아이들이 무조건적으로 이런 능력을 나타낸다고 말하지 않는다. 다만 유아들의 마음은 도덕 판단을 더 쉬이 학습하도록 이미 준비되어 있다. 약간의 환경 자극으로도 충분히 이 마음을 일깨울 수 있다(Joyce

2006: 135).

넷째, 세 살짜리 아이들도 의무론적 규칙을 습득하고 사용하는데 뛰어나다(Joyce 2006: 136). 아이들은 직설법적 조건문과 의무론적 조건문을 구분한다. 한 실험을 보자. 한 무리의 장난감 쥐들이 집과 뒤뜰에 있다. 그 쥐들 중 몇몇은 누르면 찍 소리를 내고 몇몇은 아무 소리도 내지 않는다. 이때 아이들에게 이야기를 들려준다. 못된 고양이가 찍 소리를 내는 쥐를 찾는다. 이에 여왕쥐는 밖에 있으면 위험하니까 찍 소리를 내는 쥐는 집에 있으라고 명령한다. 한 명령은 직설법적 규칙으로 제시된다. "모든 찍 소리를 내는 쥐는 집에 있어라." 다른 명령은 의무론적 규칙으로 제시된다. "모든 찍 소리를 내는 쥐는 집에 있어야 한다." 어떤 쥐가 명령을 어기는지 알려면 어떻게 해야 할까? 의무론적 규칙을 들은 아이들이 두 배 이상 더 뒤뜰에 있는 쥐를 조사해야 한다고 옳은 답변을 했다(Cummins 1996).

다섯째, 아이들은 도덕적 규범과 관습적 규범을 잘 구분한다. 도덕은 누군가의 권위와 독립적이지만 관습은 그렇지 않다. 아이들은 잠옷을 입고 학교에 가는 행동이 잘못이라 생각한다. 그런데 선생님이 학교에 잠옷을 입고 와도 괜찮다고 말하면 더 이상 그 행동을 나쁘다고 생각하지 않는다. 반면 아이들은 친구를 때리는 행동은 선생님의 허락에 상관없이 그르다고 생각한다(Smetana 1981; Smetana & Braeges 1990). 일반적으로 도덕 규범의 위반은 관습 위반보다 더 심각하게 생각되고 처벌은 쉽게 정당화된다. 조이스는 이 예가 자신이 주장한 도덕의 특별한 실천적 힘을 잘 보여준다고 말한다.

> 발달 심리학의 이런 결과들은 도덕 판단을 하는 경향이 생득적임을 강하게 시사한다. 누구도 문화적 학습이 개인이 실천하는 도덕 판단의 내용을 결정하는데 중심적인 역할을 한다는 사실을 부정할 수는 없다. 다만 나는 도덕 학습을 위해 설계된 특별한 생득적 기제(또는 일련의 기제들)가 있다고 주장한다(Joyce 2006: 137).

문화를 막론하고 보편적으로 나타나는 도덕 판단의 특징을 가장 잘 설명하는 길은 무엇인가? 조이스는 도덕의 선천성을 가정하는 방법이 제일 그럴 듯하다고 주장한다.

프린츠는 도덕이 진화적 적응이라는 의미에서 생득적이라 보지 않는다. 물론 도덕은 우리 삶에 중요한 요소이고, 인간만이 가진 매우 독특한 형질이다. 그러나 이러한 사실이 곧 도덕이 적응임을 함축하지는 않는다. 인간에게는 예술, 옷 만들기, 불 피우기처럼 누구나 갖고 있지만 선천적이지 않은 능력들이 많다. 예술적 표현을 하도록 설계된 마음이 진화했다고 볼 근거는 없다. 다만 세계를 유형화하고 모방하려는 인지 능력이 있다면 그림을 그리고, 시를 읊는 행동은 충분히 나타날 수 있다. 도덕도 그렇다. 도덕은 그 자체가 아닌, 다른 목적을 위해 진화했던 인지 능력의 부산물이다.[17] 도덕은 다양한 생득적 인지 능력과 문화, 교육, 학습이 함께 작용하여 획득된다(Prinz 2007: 270). 따라서 프린츠는 도덕 비선천주의자다.

인간이 자연선택으로 갖게 된, 도덕과 무관한 선천적인 능력은 도덕 발달에 중요했다. 프린츠가 생각한 그 능력은 다음과 같다. 첫째, 무엇보다도 감정이다. 우리는 수치심, 죄책감 등 자신을 향한 감정과 분노, 혐오, 경멸 등 타인을 향한 감정을 타고 난다. 둘째, 규칙을 형성하는 능력이다. 우리는 잘못된 행위에 느끼는 부정적 감정을 규칙으로 전환한다. 이는 행위자가 직접 연루되지 않더라도, 제삼자에게 도덕적 태도를 가지는 행동 경향의 토대다. 셋째, 기억 능력이다. 우리는 잘못된 행동을 범한 사람을 기억한다. 그런 기억은 규칙 위반이 일어난 지 오랜 후에도 지속한다. 넷째, 모방 능력이다. 아이는 처벌을 받으면 자신이 범한 잘못된 행동에 대해 나쁜 감정을 느낄 뿐만 아니라 똑같은 행동을 한 다른 사람을 처벌하려는 성향도 얻는다. 다섯째, 마음을 읽는 능력이다. 인간은 타인의 심성 상태를 미루어 짐작하고 공감하는 능력이 있다. 타인의 고통은 나의 고통이다. 이는 친사회적 행동을 장려하고 감정에 대한 감정, 즉 메타감정을 배우도록 돕는다. 사람들은 상대의 감정을 알아보고 공명하기에, 누군가 나쁜 행동을 했을 때 그가 미안함과 죄책감을 느끼도록 요구한다. 그리하여 우리는 규범을 어기면 죄책감을 경험하기 때문에 올바르게 행동하려고 한다. 이런 행동이 일상화되면 메타감정이 생겨난다. 자신이 저지른 잘못된 행동에 전혀 죄책감을 느끼지 않거나, 누군가가 범한 그른 행동을 보고도 자신에게 혐오와 경멸의 감정이 없었다면, 그래서 그를 처벌하려 하지 않았다면, 나는 이런 과오에 대한 죄책감에 빠지게 된다. 즉 타인의 고통을 외면하는 일은 죄가 된다. 메타감정은 일종의 이차적 처벌로 기능하며 계속해서 다음 세대로 대물림되어 한 사회의 도덕 규범을 안정화시킨다(Prinz 2007: 270-272).

앞에서 프린츠는 인간에게 기초 가치가 있음을 주장했다. 우리는 다른 사람을 돕고, 자원을 공유하고, 협동하고, 성과 폭력을 통제하는 일 등을 중요하게 여긴다. 이런 가치들은 우리가 도덕이라 부르는 영역과 밀접하게 관련된다. 도덕 선천주의자들은 기본적인 도덕 가치와 이로써 생긴 '공평성', '호혜성' 등의 도덕 개념이 자연선택되어 진화한 적응이라 본다. 하나 프린츠는 선천주의를 거부한다. 기초 가치는 도덕과 무관한 생득적 능력과 문화가 결합하여 쥐게 된 구성물이다. 도덕은 문화적 구성물이다. 문화는 사회적 학습으로 어린 아이들이 기초 가치를 내재화하도록 훈련시킨다. 이때 도덕 교육은 기초 가치에 감정을 조건화하는데 달려 있다. "근친상간은 그냥 나빠."라고 말할 때, 그 근저에는 우리를 휘젓는 혐오의 감정이 있다. 프린츠는 이렇게 교육으로 특정 문화적 사례에 공포, 혐오 등의 감정을 결합시키면 새로운 가치를 기초 가치로 만들 수도 있다고 말한다(Prinz 2007: 195).

---

17) 여기서 프린츠는 '부산물'을 자연선택이 작용했던 형질이 아니며, 생득적 능력과 문화가 결합해 생긴 후천적 소산이라는 뜻으로 사용한다. 그러나 부산물이 반드시 후천적임을 의미하지는 않는다. 우리 몸을 이루는 뼈와 피는 개체의 생존에 필수적인, 선택이 작용한 진화적 적응이지만 그 하얗고 빨간 색깔은 뼈와 피를 이루고 있는 원소가 나타내는 부산물이다. 그래도 피와 뼈의 색은 선천적인 형질이다. 요컨대 우리는 선천적인 부산물과 비선천적인 부산물을 나눌 수 있다. 프린츠는 인간의 도덕 행동을 만드는 주요인으로 문화를 꼽으므로 도덕은 비선천적인 부산물이라고 강조한다.

문화는 다양하다. 따라서 기초 가치는 문화에 따라 서로 다른, 구체적인 도덕 규칙을 만든다. 프린츠는 인류학자 헨리히(J. Henrich)의 연구를 사례로 도덕 규칙의 문화적 다양성을 보여준다. '공평성'은 인간이 보편적으로 소유한 도덕 개념이다. 헨리히는 이 공평성이 문화에 따라 어떻게 나타나는지 탐구하려고 15개 소규모 수렵 채집 사회를 대상으로 최후통첩 게임을 진행했다(Henrich et al. 2005). 최후통첩 게임은 실험자가 제안자에게 일정 금액을 주고 응답자와 원하는 대로 나누라고 지시한다. 응답자는 제안자가 제시한 금액이 마음에 들지 않을 경우, 협상을 거부할 수 있다. 그러면 나누지 못한 돈은 실험자가 다시 가져간다. 경제학이 가정하는 합리적 인간 모델에 따르면, 제안자는 가장 최소의 금액을 제시하고 응답자는 어떤 금액이라도 일단 받아들이는 전략을 취하리라 예상한다. 한데 서구 사회에서 시행한 실험은 생각 밖이었다. 제안자는 대개 절반에 달하는 금액을 제시했고, 응답자는 제안이 조금이라도 불공평하다고 여기면 거절하여 둘 다 돈을 못 받는 쪽을 택했다.

수렵 채집 사회에서 실시된 최후통첩 게임은 집단 간 편차가 아주 컸다. 예를 들어 남아메리카의 마치겡가(Machiguenga) 사회에서는 제안 금액의 평균이 26%로 조금 낮았다. 그래도 많은 응답자들은 별다른 불만 없이 제안을 수락했다. 여기에는 경제적인 이유가 있었다. 마치겡가 부족은 화전민이어서 가족 밖의 사람과 협동할 일이 많지 않았다. 그래서 이 부족은 공평성과 관계된 규범을 장려하는 게 크게 중요하지 않았다. 파푸아뉴기니의 노(Gnau)와 오(Au) 부족은 정반대였다. 이 부족이 제안한 금액 평균은 서구 사회와 비슷했고, 때때로 50%를 넘는 매우 관대한 금액 제안도 있었다. 하지만 노와 오 부족 사람들은 높은 제안액을 거부하는 행동을 보였다. 여기에도 그럴만한 이유가 있었다. 노와 오 부족은 선물을 주고받는 관습이 있어 내가 선물을 받으면 반드시 보답을 해야 했다. 그렇기에 관대한 제안은 갚아야 할 무거운 부담이었다. 이로써 프린츠는 우리가 보편적이라 생각하는 개념들, 예컨대 '친절함', '공평성', '호혜성', '위계 서열' 등은 문화에 따라 그 구체적인 내용이 다르게 형성되고 유지된다고 주장한다(Prinz 2007: 275-276).

프린츠는 도덕 선천주의를 주장하는 조이스의 근거를 일일이 반박한다. 먼저 조이스는 도덕이 우리가 아는 모든 사회에서 나타나기 때문에 선천적이라고 주장했다. 그러나 프린츠가 보기에 도덕의 보편성은 비선천적인 설명과 양립 가능하다. 도덕은 근본적으로 감정 표현이다. 또한 모든 사회는 구성원들이 규칙을 순응하기를 요구한다. 이를 달성하는 효과적인 방법은 규칙에 감정을 짝지우기다. 위반자가 처벌받거나 추방된다면 감정적 성향의 조건화는 매우 자연스럽게 일어나리라. 그러므로 감정이 보편적인 만큼, 도덕 규칙에 대한 감정적 성향도 모든 사회에서 나타난다. 둘째, 과연 모든 사회의 도덕적 내용이 동일한가도 의문이다. 어떤 사회는 노예 소유를, 다른 사회는 식인 행위를 정당하게 생각한다. 또 어떤 사회는 굉장히 평등한 반면, 다른 사회는 엄격한 위계 질서를 갖고 있다. 이런 사실로 미루어 보아 도덕에는

사회적 학습과 문화적 환경이 본성보다 중요하다. 셋째, 아이들이 도덕 개념을 학습하고 깨우치는 과정이 빈곤하다고 보기 어렵다. 아이들과 늘 붙어 있는 양육자는 허용과 금지, 규칙, 규범 등을 자신도 모르게 가르친다. 아이들은 규칙을 위반하면 처벌을 받는다는 경험 덕에 의무적인 규칙과 기술적인 규칙의 차이를 배우고 모방으로 이를 강화한다. 넷째, 도덕과 관습의 구분도 아이들이 관습보다 도덕을 어겼을 때 더욱 강력한 처벌을 받기 때문에 가능하다. 가혹한 처벌은 도덕 규범을 어기는 행위에 더욱 부정적인 감정을 느끼도록 만든다. 그리하여 높은 수준의 부정적 감정은 도덕이 어떤 권위와도 독립된 체계로 보이게끔 한다. 대체로 타인에게 피해를 주는 행동은 강한 처벌을 받는다. 따라서 아이들은 선생님이 허락했어도, 이유 없이 친구를 때리는 행동은 도덕적으로 허용되지 않는다고 생각하게 된다(Prinz 2008a: 221-224).

이상으로 도덕이 적응이냐 부산물이냐에 대한 조이스와 프린츠의 견해차를 살펴보았다. 조이스는 도덕 선천주의를 주장함으로써 자신의 진화적 도덕 반실재론을 더 튼튼하게 한다. 반실재론에 기반을 둔 도덕 선천주의는 도덕 사실에 기대지 않더라도, 진화만으로 도덕 판단의 중요성과 영향력을 잘 설명한다. 그러나 이런 조이스의 전략에 떠오르는 의문이 있다. 진화론에 근거한 도덕 선천주의가 도덕 실재론을 지지하는 데 쓰일 수는 없을까? 도덕 선천주의와 도덕 실재론은 잘 어울리는 쌍이다. 우리는 무엇이 본유적으로 내재해 있다면 자연스레 그 대상이 실재한다고 생각하기 때문이다. 도덕 사실에 대한 관념을 뒤집는다면 진화론은 도덕 실재론을 옹호하는 근거가 되지 않을까, 진화는 반응 의존적인 도덕 사실을 만드는 과정이지 않을까, 반응 의존적인 도덕 사실은 선천적이지 않을까? 이 질문들이 다음 장에서 전개할 수정된 감수성 이론의 기초다.

한편 프린츠는 도덕 비선천주의를 주장함으로써 도덕 사실이 다원적이라는 주관적 실재론을 두둔한다. 이런 방법은 얼마나 견고한가, 정말로 프린츠의 도덕 상대주의와 도덕 실재론은 양립 가능한가? 그 답은 조금 회의적이다. 도덕 상대주의는 우리가 실재론에 기대하는 바를 충족시키지 못한다. 실재론은 사람들에게 보편적으로 적용되는 올바른 답이 있다는 직관을 설명해야 한다. 근친상간에 대해 서로 다른 견해를 가지면서 각자의 의견이 실재를 반영한다는 주장을 받아들일 수 있겠는가? 프린츠의 실재론은 실재의 개념이 너무 사소하다. 프린츠의 감수성 이론은 곧 비판에 직면한다. 이에 이 글은 다음 장에서 프린츠의 주장에 제기된 조이스의 비판을 알아보고, 프린츠의 감수성 이론을 보완한 수정된 감수성 이론을 제안하여 도덕 실재론을 구하고자 한다. 기본적인 생각은 이렇다. 도덕 사실이 마음 독립적이라는 전제를 버린다면, 진화는 반응 의존적인 사실을 만드는 필연적인 과정이다. 자연선택은 우리가 모두 동의하는 기초 가치를 형성했다. 기초 가치는 후일 도덕이라 부르는 모든 양식의 근원이다. 인간 종은 보편적인 도덕 감정을 일으키는 기초 가치를 가지며 이로써 특정 조건에서 도

덕 판단, 즉 진화한 기초 가치에 대한 우리 반응은 수렴된다. 따라서 도덕 사실은 반응 의존적으로 존재한다. 문제는 우리가 어떻게 도덕 판단의 합일을 기대할 수 있느냐다. 감정 반응이 일치할 수 있는 조건이 있을까? 이에 수정된 감수성 이론은 실제 관찰자가 어쩔 수 없이 영향 받는 편견과 무지에서 벗어난 이상적 관찰자를 도입한다. 이상적 관찰자의 상정은 전혀 새로운 방법이 아니며 어려운 문제를 한 번에 해결하는 신적 장치도 아니다. 우리는 뒤에 가서 이상적 관찰자의 존재가 완전히 합당함을 이해하게 되리라 자신한다. 조이스와 프린츠는 전혀 새로운 주장을 하는 게 아니다. 진화적 폭로 논증과 감수성 이론은 기존 메타윤리학의 전통 내에서 반복되는 논의를 충실히 따르고 있다. 다만 조이스는 도덕 반실재론에 진화를 끌어들였고, 프린츠는 객관성을 제거한 감수성 이론을 제안했다는 점만이 독특하다. 그렇다면 똑같이 메타윤리학의 전통 내에서 진화와 반응 의존성을 연결한 객관적인 도덕 실재론을 제안하지 못할 이유가 없다.

4. 프린츠의 감수성 이론에 대한 비판과 그 대응-수정된 감수성 이론의 제안

4.1. 도덕 판단의 객관성 확보

4.1.1. 도덕 상대주의 비판

프린츠의 감수성 이론은 도덕 속성이나 사실이 어떤 사태를 승인, 불승인하는 감정을 일으키는 감성으로 구성된다고 주장한다. 그런데 프린츠는 감성이 발현되는 보편적인 상황과 관찰자의 특성을 제시하지 않는다. 이는 감수성 이론이 상정하는 도덕 속성이 반응 의존적이며, 반응은 도덕 판단의 상대성을 함축하기 때문이다. 감정은 주관적이다. 감정은 개인적 역사와 문화적 배경에 영향 받는 내적인 심성 상태이다. 우리는 도덕 문제에 감정적으로 대응하지만 모두 같은 일에 같은 감정을 느끼지는 않는다. 결국 프린츠의 감수성 이론은 도덕 상대주의로 귀착된다. 프린츠는 도덕 상대주의가 실재론에 방해가 된다고 생각하지 않으며 주관적 실재론이라는 독특한 입장을 견지한다.

그러나 조이스는 도덕 판단이 일어나는 특정 상황과 관찰자를 명시하지 않은 정식이 큰 문제를 내포한다고 지적한다. 도덕 감성이 발현하는 때는 언제인가? 어떤 상황에서 관찰자는 도덕 위반을 보고 분노할 수 있지만 다른 상황에서는 냉철할 수 있다. 또한 특정 상황에는 관찰자의 양육배경, 그가 뇌를 다쳤는지, 홍적세에 사는지, 화성에 사는지와 같은 수많은 변화가 포함될 수 있다. 또한 관찰자는 누구인가? 나 혹은 그는 히틀러일 수도 있고 간디일 수도 있다. 히틀러가 유대인 학살에 승인의 감정을 느끼면 그건 좋음의 속성을 가리키는가? 결국 프린츠의 감수성 이론은 극단적인 상대주의로 간다. 도덕 판단이 일어나는 상황과 관찰자를 한정하지 않으면 감정을 느끼는 성향은 도덕 속성을 지시하는 데 실패한다. 이를 '불완전성 문제(incompleteness problem)'라고 부른다(Joyce 2008: 254-256; 2009: 511-513).

감정에 바탕을 둔 감수성 이론과 도덕 객관주의는 양립할 수 있을까? 그렇다. 앞에서 보았듯이, 어떤 사실이 우리 판단에 의존해도 무엇이 옳고 그른지 말할 수 있다. 감수성 이론과 도덕 상대주의는 동의어가 아니다. 감수성 이론이 적절한 판단의 기준을 제공할 수 있다면 '객관적'이라는 칭호를 받기에 충분하다. 따라서 이 글은 객관적인 감수성 이론, 곧 수정된 감수성 이론을 제안하고자 한다. 그 전략은 첫째, 인간 종이 공유하는 진화한 기초 가치의 존재. 둘째, 도덕 판단이 일어나는 조건을 명시하는 이상적 관찰자의 도입이다.

첫 번째부터 보자. 호모 사피엔스로서 인간 종은 공통의 가치를 만드는 전형적인 반응의 역사를 경험했다. 진화는 수십 만 년에 이르는 동안 이런 상황에는 저런 반응이 올바르다고 정립해 나간 과정이다. 예를 들어 사람들이 '우스움'이나 '부지런함'이라는 개념에 대해 떠올

리는 생각은 유사하다. 우스움이나 부지런함에는 범문화적으로 통용되는 반응이 있다. 도덕적 '옳음'이나 '그름'도 똑같다. 도덕과 관련된 문제에도 보편적인 반응이 있고, 이로써 기초 가치가 형성되었다. 인간에게 생물학적인 본유 가치가 있다면 도덕 사실은 반응 의존적으로 존재하며, 도덕 판단은 객관적으로 수렴될 수 있다. 프린츠의 도덕 상대주의를 비판하면서 더 자세히 논해보자.

프린츠는 도덕 판단을 진화적 적응으로 보는 도덕 선천주의자들이 도덕 내용의 문화적 다양성을 외면했다고 주장한다. 그러나 문화적 차이는 도덕 선천주의에 대한 비판이 될 수 없다. 도덕 선천주의는 절대로 사회적 학습과 환경의 영향을 부정하지 않는다. 고정된 도덕 믿음이 조상들의 번식 성공도를 증진시켰던 적응이라 보기는 힘들다. 적응적 행동을 만드는 데는 자연선택과 환경이 상호작용한다. 이때 선택은 환경 변화에 유연하게 대처할 수 있는 심리 기제를 선호한다. 주변이 어떠하든 변함없는 출력값을 내는 성향은 별로 도움이 되지 않기 때문이다. 경직된 행동 방침은 언젠가 큰 대가를 치른다. 더구나 무진한 일상의 사건들에 모두 대응하는 도덕 믿음을 타고 나는 일도 불가능하다. 그래서 조이스를 위시한 도덕 선천주의자들은 도덕 판단을 하는 능력과 이에 쓰이는 도덕 개념이 자연선택의 산물이라고 본다. 곧 도덕 선천주의는 도덕 믿음의 문화적 보편성이 아니라 개념의 심리적 보편성을 가리킨다. '옳음'이나 '그름'에 속하는 도덕 개념들의 목록은 범세계적이다. 물론 개념의 보편성이 도덕 판단의 일치를 대번에 보증하지는 않는다. 개념은 환경에 맞추어 특정한 믿음 내용을 산출한다. 같은 개념을 가진 두 개체는 어디에 살고 있느냐에 따라 다른 믿음을 습득한다. 하나 애초에 개념 생산은 인간 종의 기본적인 필요와 욕구를 반영한다. 개념 생산은 인간 종이 공통적으로 마주친 생태적 이유, 자연 조건, 선택 압력과 관련된다. 즉 인간 종이 보유한 도덕 개념을 형성했던 기초적인 가치는 근본적으로 동일하다. 우리는 모두 아프리카 사바나에서 활동했던 공통 조상의 후손들이다. 따라서 다른 모든 조건이 동등하다면, 서로 다른 환경 배경이 없다고 가정한다면, 도덕 판단은 기초 가치로 수렴된다.

앞에서 프린츠는 도덕 규칙이 다양하게 나타나는 예로 인류학자 헨리히가 실험한 15개 소규모 수렵 채집 부족의 최후통첩 게임을 들었다. 각 부족은 사는 환경에 따라 돈을 분배하는 비율이 서로 달랐다. 그러나 이 실험의 요점은 문화마다 '공평성'의 구체적 내용이 다르게 나타나는 데 있지 않다. 헨리히 외 연구자들은 15개 사회 어디서든지 '공평성'이란 도덕 개념을 공유한다는 사실을 발견했다. 경제 게임에서 어떤 부족도 자기만 이익을 독차지하는 행동을 보이지 않았다. 누구나 타인과 성과를 나눌 때는 공정해야 옳다고 여겼다. 경제학이 전제하는 이기적인 인간 모델에 반하는 증거는 서구 대학생들만이 아니라 소규모 부족에서도 마찬가지로 발견되었다(Henrich et al. 2005: 803-805). 요컨대 상이한 도덕 규범의 근저에는 반드시 공통 가치에서 발원한 도덕 개념이 있다. 선천적인 가치와 개념 덕에 문화적 차이로는 도

덕 판단이 하나로 모일 가능성을 배제하지 못한다.

다른 도덕 규칙도 매일반이다. 진화심리학에서는 근친상간 금기를 생득적인 심리 기제가 작동한 결과로 설명한다. 인간은 매우 어린 시기에 함께 살았던 사람과 교접하는 행위를 혐오하는 자동 반응을 타고 난다. 이를 '웨스터마크 효과(Westermarck effect)'라 부른다. 근친상간 혐오는 상대가 비혈연이어도 적용된다. 인종, 지역, 가문 등 다양한 배경을 가진 아이들을 공동 양육했던 이스라엘의 키부츠에서 어린 시절에 함께 자란 아이들의 혼인율은 매우 낮았다. 한데 프린츠는 웨스터마크 효과가 근친상간에 대한 문화적 차이를 설명하지 못한다고 주장한다. 어떤 사회는 사촌 간의 결혼은 허용하는 반면 현대 서구 사회는 사촌 간의 결혼도 금지한다(Prinz 2007: 283).

왜 문화마다 근친상간에 대한 견해가 다를까? 이 또한 근친상간에 결부된 도덕 개념이 달라서가 아니다. 앞서 말했듯이, 개념은 환경과 상호작용한다. 독특한 주변 여건에는 그에 맞는 규칙이나 관습이 생기기 마련이다. 자기 집단의 고유한 믿음을 받아들이는 개체가 생존과 번식에 더 유리하다. 즉 한 사회의 구성원들이 갖는 상이한 도덕 믿음에는 생태적 이유가 있다. 그렇다면 지역 특유의 자연 조건이 사라진다면 어떻게 될까, 정말로 아무 반대 없이 친지와의 혼인을 용인하는 문화가 있을까, 어머니와 혼인한 사실을 알게 되자 자신의 눈을 찌른 오이디푸스 이야기에 공감하지 않는 인류가 있을까? 그렇지 않다. 우리의 진화한 마음은 백지가 아니다. 인간의 믿음은 본성에 제약받는다. 개체가 수용해야 할 환경 압력이 없다면 누구도 근친상간을 허용하지 않으리라. 정확히 근친상간의 사례는 아니지만 티베트의 일부 촌락에서 이뤄졌던 일처다부제는 인간 본성의 힘을 입증한다.

티베트의 잔스카르(Zanskar)와 라다크(Ladakh) 지역에는 형제끼리 아내를 공유하는 통념에 반하는 문화가 있다. 일반적으로 진화는 수컷이 되도록 많은 암컷과 교미하려 한다고 예측한다. 수컷의 번식 성공률은 암컷의 수에 따라 증가하기 때문이다. 유성생식하는 종에서 수컷과 암컷은 양육 투자에 들이는 비용이 서로 다르다. 대개 정자만을 제공하는 수컷이 여러 달 동안 새끼를 임신하고 키워야 하는 암컷에 비해 훨씬 노력이 덜 든다. 수컷은 기회가 된다면 새로운 암컷을 만나는 게 이익이다. 따라서 한 명의 암컷을 여러 수컷이 공유하는 행동은 매우 드물고 이상한 현상이다. 그럼 일처다부제에 순응하는 티베트 사람들은 우리와 전혀 다른 관념과 습속을 가진 사람들일까? 아니다. 연구자들은 티베트의 일처다부제 관습에 생태적 이유가 있음을 발견했다. 잔스카르와 라다크 촌락은 농지를 경작하여 삶을 영위한다. 각 가정에서는 보통 장남이 부모에게 상속받은 농토에서 작물을 재배한다. 농토를 여러 명의 아들에게 나눠줄 수도 있지만 그렇게 하면 가족을 부양하지 못할 정도로 규모가 작아진다. 그래서 형제들은 한 아내와 공동으로 결혼하여 함께 농사를 짓는다. 즉 일처다부제는 티베트인들이 척박한 생활 여건에 적응한 결과이다. 여기서 또 물어보자. 이런 환경 요인이 사라진다

면 어떻게 될까? 그런 일이 실제로 일어났다. 잔스카르와 라다크 사람들에게 농사 외에 다른 수입원이 생겨 어려운 생활 여건이 나아지자 동생들 대다수는 자신의 아내를 찾아 떠났다. 일처다부제는 급속도로 무너졌다(Crook & Crook 1988). 티베트의 일처다부제 사례는 이 글의 핵심 전략을 지지해준다. 믿음을 개별화하는 환경 요소가 사라지자 종 보편적인 행동은 곧바로 나타났다. 그러므로 우리 도덕 개념은 심리적으로 보편적이며, 그 개념을 형성한 근본적인 이유는 동일하다. 도덕 개념에 근본적 불일치가 있다는 증거는 없다.

선천적인 도덕 개념을 만들었던 기초 가치는 무엇일까? 도덕 심리학자 하이트는 실험을 거쳐 도덕 판단이 주로 배려/피해, 공평성/부정, 충성심/배신, 권위/전복, 고귀함/추함, 자유/압제의 여섯 가지 가치 영역에서 범세계적으로 일어남을 발견했다.[18] 여섯 가지 기초 가치는 현재 우리가 도덕이라 부르는 심리와 행동의 근간이다. 이를 '도덕 기반 이론(moral foundations theory)'이라 부른다(Haidt & Bjorklund 2008; Haidt 2012/2014). 여섯 가지 영역은 조상들이 마주쳤던 적응 문제와 관련 있다. 남을 해하지 않고 아이들을 보살피기, 타인과 협력하기, 집단과 연합 만들기, 지위 확보하기, 공동체에 퍼지는 기생충과 전염병 방지하기, 지배에 저항하기 등이다. 기초 가치는 이런 진화적 원인들에서 생겨났지만 현대 사회에서 자주 마주치는 근접 원인들로도 자극된다. 약자가 겪는 고통에 반응하는 배려/피해 가치는 때로 만화에 등장하는 허구의 인물을 괴롭히는 행동도 도덕적으로 그르다고 판단하게 만든다. 그래서 근접 원인은 도덕 믿음의 개인적, 집단적 다양성을 만든다.

기초 가치에 바탕을 둔 도덕 판단은 특정 감정과 밀접하게 연결되어 있다. 예를 들어 배려/피해는 분노, 동정심, 공평성/부정은 경멸, 죄책감, 감사, 고귀함/추함은 역겨움을 일으킨다. 자연선택이 기초 가치와 감정을 짝지은 이유는 명백하다. 빠르고 자동적인 행동을 산출하는 심리 전략이 이성적 추론보다 훨씬 효과적이기 때문이다. 감정은 마음을 뒤흔들며 적극적인 행동을 산출한다. 더하여 기초 가치에 대한 감정 반응은 '친절함', '공평성', '충성심', '복종심', '경건함' 등과 같은 도덕 개념들을 구성한다. '공평성'이라는 개념에는 배신자를 욕하고 처벌하려는 감정이 들어 있다. 하이트의 도덕 기반 이론은 자연선택으로 생성된 기초 가치를 명시할 뿐만 아니라 인간에게 선천적인 도덕 규칙이 있음을 시사한다. 이 규칙은 하나의 도식으로 표현된다. "x가 아니라면 y를 해하는 행동은 하지 마라." 문화에 따라, 이 'y'에는 여러 행위자(가족, 이웃, 인류, 동물 등)가 포함되며 'x'는 예외 상황(복수, 운동 경기, 전쟁 억지)들을 정한다(Prinz 2008b: 380).

프린츠는 다른 글에서 선천적인 도덕 기반의 존재를 세 가지 점에서 비판한다. 그는 먼저 여섯 가지 도덕 영역은 보편적이지 않다고 지적한다. 도덕 기반은 문화마다 그 강조의 정도

18) 하이트는 Haidt & Bjorklund(2008)에서 배려/피해, 공평성/부정, 충성심/배신, 권위/전복, 고귀함/추함의 다섯 가지 가치가 도덕 기반을 이룬다고 주장했지만, Haidt(2012/2014)에서는 자유/압제를 추가했다.

가 다르다. 어느 지역에서는 피해 금지에 강력한 도덕적 금지를 만들지만, 다른 곳에서는 그런 문제를 덜 중요하게 생각한다. 도덕 기반은 모든 문화권에서 똑같이 중요하지 않다. 다음으로 학습을 이용하여 도덕 기반에 대한 대안적 설명이 가능하다. 도덕 기반은 특정 감정들과 연결되어 있다. 한 예로 사람들은 고귀함/추함의 영역에서 도덕적으로 불결한 대상이나 물건에 대해 역겨움을 느낀다. 이때 어떤 집단의 교육자는 구성원들이 올바로 행동하도록 특정 행동과 역겨움을 연합시킬 수 있다. 프린츠는 사회가 도덕 능력을 육성하려고 특정 행동에 역겨움을 느끼는 성향을 조건화할 수 있다면 도덕 기반은 학습으로 연장된, 도덕과 무관한 감정들의 소산이라고 주장한다. 마지막으로 프린츠는 하이트가 제시한 도덕 기반이 진정으로 '도덕'인가를 묻는다. 프린츠는 도덕 기반과 관련된 특정 감정들이 도덕 판단을 목적으로 진화한 도덕 감정이 아니라 도덕과 무관한 상황에서 사용된 기초 감정의 부산물이라고 주장한다. 따라서 선천적인 도덕 기반 또한 진화적 적응의 산물이 아니라 감정과 문화적 학습이 결속하여 생긴 부산물이다(Prinz 2008b: 381-383).

그러나 프린츠가 제기한 첫 번째 비판은 도덕 기반 이론이 주장하는 바와 같다. 단지 어떤 문화가 특정 영역을 더 중시하고 다른 영역은 덜 고려한다고 해서 도덕 기반이 보편적이지 않다고 볼 수는 없다. 도덕 기반은 도덕 판단 그 자체가 아니라 토대를 이루는 가치이다. 분명 도덕 기반을 자극하는 근접 원인은 문화마다 다양하고 학습과 경험 덕분에 가중치를 두는 영역도 다르다. 그렇더라도 문화는 우리 삶과 독립적인 별개의 실체가 아니다. 각 지역에 고유한 문화나 이념은 어디까지나 인간 본성과 환경이 상호작용하여 형성된다. 티베트의 사례에서 보듯 그런 독특한 생태적 조건이 없다면 본성은 동일하게 나타난다. 다시 말해, 정도차가 있을 뿐이지 도덕 기반은 종 보편적이다. 모든 문화가 여섯 가지 영역에 똑같은 중요성을 부여해야 한다는 요구는 비현실적이다. 두 번째 비판은 인간 본성의 힘을 간과했다. 도덕 기반 이론도 학습의 역할을 부정하지 않는다. 인간은 시체를 훼손하는 행동에 역겨움을 느끼도록 결정된 게 아니다. 다만 이런 행동에는 저런 감정을 더 쉽게 연합할 수 있도록 선천적으로 준비되어 있을 뿐이다. 적절한 자극이 없다면 알맞은 반응을 보이지 못하는 늑대인간이 될 수도 있다. 하나 학습 능력도 인간 본성의 일부다. 우리 반응은 무한하지 않다. 마음에는 한계가 있다. 인간은 뱀에 대한 공포는 쉽게 학습하지만 꽃에는 그렇지 못하다. 아무리 노력해도 사람들이 시체 훼손에 행복을 느끼게 만들기는 어렵다. 반면 사람들은 시체 훼손에 대한 역겨움 반응은 아주 수월하게 익힌다. 도덕 기반과 감정의 연결은 이런 의미에서 선천적이다. 세 번째 비판도 도덕 기반 이론을 논박하지 못한다. 프린츠는 도덕 감정이 원래는 도덕과 무관한 감정들에서 유래했기 때문에 부산물이라고 주장한다. 하지만 이는 어디까지나 추정이다. 도덕과 무관한 감정이 도덕 감정으로 전용되도록 선택되었을 가능성을 원천적으로 배제할 수 없다.

인간 종은 기초 가치를 공유한다. 기초 가치는 인류가 동의하는 상투적인 도덕적 진리를 만든다. “재미로 사람을 고문하는 짓은 나쁘다.”, “쾌락 때문에 무고한 사람을 살인하지 마라.”, “기분에 따라 약속을 어기는 일은 옳지 않다.” 등등. 다른 구실이 없을 때 상투적인 도덕적 진리에 동의하지 않는 사람을 상상하기는 힘들다. 앞서 프린츠도 더 이상의 이유가 필요 없는, “그건 그냥 나빠!”로 표현되는 기초 가치가 있다고 말했다. 프린츠는 기초 가치가 문화와 감정적 조건화 학습으로, 후천적으로 발생했다고 본다. 그러나 이는 너무 빈약한 설명이다. “그건 그냥 나빠.”는 자연선택의 논리이다. 자연선택은 개체가 궁극 원인을 의식하지 않고 감정에 따라 즉각 행동을 옮기는 비용-효율적 전략을 선호한다. 감정적 조건화는 특정 믿음을 학습하려는 준비가 되어 있을 때 경제적이다. 매우 어린 아이도 적은 경험으로 도덕과 관습을 구별하고 도덕 위반을 “그냥 나빠!”라고 생각한다. 더구나 하이트가 제시한 기초 가치의 여섯 가지 목록은 문화를 막론하고 나타난다. 따라서 기초 가치는 진화의 산물이라 보는 게 합당하다. 교차문화적으로 발견되는 상투적인 도덕적 진리에서의 일치는 진화적으로 제한된 기본적인 도덕 가치와 개념이 있음을 지지한다. 상황이 이렇다면 감정이 도덕 판단을 목적으로 선택된 심리 기제는 아닐지라도 이후에 감정과 기초 가치의 연결이 적응적일 가능성은 여전히 있다.

결론적으로 도덕 판단의 문화적 다양성은 너무 과장되었다. 문화는 인간 행동을 규제하는 독립된 실체로 저절로 생겨나지 않았다. 문화는 인간 본성과 생태적 환경이 상호작용함으로써 발생했다. 이는 인간의 문화가 많은 부분 유사한 특징을 보이며 또 완전히 극단적이고 전혀 새로운 문화가 생겨나지 않는 이유이다. 문화는 인간의 생물학적인 토대 안에 있다.

4.1.2. 이상적 관찰자의 도입

이상의 논의를 참고할 때 프린츠의 감수성 이론을 보완할 수 있는 특정 조건은 도덕과 무관한 모든 사실이 일치하는 상황이다. 도덕 판단을 둘러싼, 도덕 이외의 모든 사실이 일치하면 특정 사람, 행동, 사태에 대한 불승인, 승인의 감정 반응은 진화한 기초 가치로 수렴된다. 도덕적 논쟁이나 불일치는 애초에 공유하는 가치와 거기서 생겨난 도덕 개념 없이는 일어날 수 없다. ‘공평함’이라는 개념이 없는 사람과 어떤 행동이 공평한지 논하는 일은 무용하다. 따라서 도덕 가치가 선천적이라면 도덕 실재론자는 원리상 도덕적 불일치를 해결할 수 있다. 도덕적 불일치는 도덕과 무관한 사실과 전제들의 불일치에서 발생하기 때문이다(Brink 1989: 197-210). 히틀러가 유대인 학살에 승인의 감정을 느꼈다고 해서 결코 그 감정이 도덕적 좋음을 지시할 수는 없다. 히틀러의 감정은 잘못되었다. 히틀러는 유대인에 대한 잘못된 믿음과 사실에 바탕을 두고 판단했다. 히틀러는 인종 간의 우열이 있으며 아리안 족이 유대인보다

신체적, 정신적으로 더 뛰어나다고 생각했다. 그러나 이러한 믿음은 명백히 잘못이다. 히틀러의 의견은 도덕적으로 그르다. 감수성 이론은 도덕과 무관한 사실에 관한 믿음이 잘못일 경우 그에 따르는 감정이 잘못되었다고 말할 수 있다. 도덕 판단은 감정적 성향의 표현이지만 표출된 감정이 옳거나 그르다는 주장이 포함되어 있다. 행위자의 도덕 감정은 잘못된 사실에 근거하지 않을 때 정당하다. 요컨대 어떤 행동에 대해 느끼는 더 나은 감정, 더 나쁜 감정이 있다. 이는 감수성 이론이 적절한 감정이라는 기준으로 도덕 판단의 객관성을 수용할 수 있음을 보여준다.

어떤 도덕 반실재론자들은 도덕적 논쟁에 끊임없는 불일치가 있다는 점을 들어 객관적인 도덕 사실이 있다는 주장을 부정한다. 그러나 정말 도덕적 불일치가 합의 불가능한 근본적 불일치인가? 그렇지 않다. 어떤 사람은 낙태는 살인이라며 반대하고 다른 사람은 찬성한다고 해보자. 두 사람이 대립하는 원인은 무엇일까? 두 사람은 낙태라는 행위를 평가할 때 서로가 중요시하는 도덕 외적 사실이 다르다. 한 사람은 태아도 사람이라 주장하고 다른 사람은 태아는 사람이 아니라 주장한다. 즉 두 사람은 태아도 사람인가, 사람이라면 언제 태아를 사람이라 볼 수 있느냐를 놓고 대립한다. 한데 둘은 해를 금지하는 가치에서 발생한, 무고한 생명을 죽여서는 안 된다는 규칙은 공통적으로 갖고 있다. 따라서 언제 태아를 사람으로 보느냐에 합의하면 논쟁은 해결된다. 낙태에 대한 우리 감정 반응은 하나로 일치된다.

도덕과 무관한 모든 사실이 일치하는 이상적 상황을 가정하면 관찰자도 이상적 관찰자여야 한다. 현실의 관찰자는 때로 도덕 외적인 사실에 무지하거나 정보가 부족하기 때문이다. 잘 모르는 상태에서 나온 판단이 도덕 사실을 반영한다고 보기는 어렵다. 그렇다면 이상적 관찰자는 어떤 관찰자인가? 이상적 관찰자가 구비한 성격의 현대적 판본은 퍼스(R. Firth)가 제시했다. 퍼스는 이상적 관찰자가 도덕과 무관한 모든 사실에 대해 박식하고, 완전한 지각 능력이 있으며, 사심 없고, 공정하며, 일관적이고, 그 외 다른 모든 점에서 정상인 인간이라 가정했다. 도덕 판단을 위한 완전한 능력을 갖춘 이상적 관찰자의 판단은 절대적 기준이 된다. 어떤 행동은 이상적 관찰자가 옳다고 반응할 경우 오직 그러한 경우에만 옳다(Firth 1952).

퍼스는 이상적 관찰자의 반응이 무엇인지는 명시하지 않았다. 물론 이 글은 해당 반응이 특정 행동에 대해 승인과 불승인의 감정을 느끼는 감성이라고 생각한다. 더하여 이 글은 상술했듯, 이상적 관찰자의 다른 어떤 성격보다도 도덕과 무관한 모든 사실을 알고 이해하며 거짓 믿음이 없다는 특질을 최선으로 꼽는다. 이상적 관찰자는 갈등하는 사안, 인간 본성, 환경 여건 등등에 대한 완전한 지식을 갖고 있으며 이로써 자신이 승인하는 가치를 추구하고, 불승인하는 가치를 금지하는 실천적으로 합리적인 인간이다. 이상적 관찰자는 인종 간의 우열이 없으며 아리안 족과 유대인이 모두 동등한 인간이라는 사실을 안다. 따라서 이상적 관찰자는 히틀러의 학살 행위를 불승인한다. 도덕 판단과 관련된 충분한 정보를 아는 이상적

관찰자는 자신의 개인적인 배경과 문화적 관습에서 벗어나, 공평무사(impartial)하게 인간 종으로서 보유한 생물학적인 보편 가치를 느끼도록 돕는다.

그렇다면 어떤 행위는 현실의 관찰자가, 이상적 관찰자가 불승인하는 행위를 불승인할 때 도덕적으로 그른가? 그렇지 않다. 이상적 관찰자는 말 그대로 이상화된 완전한 존재이다. 전지전능한 자가 주는 규준이 있다면 사실상 현실의 관찰자가 나타내는 반응은 아무런 의미도, 효력도 없다. 우리는 이상적 관찰자와 현실의 관찰자가 연결되기를 바란다. 현실의 관찰자가 자신의 제한된 조건에서 가능한 이상적인 판단을 내리길 바란다. 그리하여 현실의 관찰자가 도덕 사실에 다가가기를 바란다. 따라서 이상적 관찰자의 반응은 현실의 관찰자가 발 딛고 있는 환경과 그의 한계 내에서 고려되어야 한다(Railton 1986: 173-175). 그렇다면 어떤 행동은 현실의 관찰자가 느끼는 승인, 불승인의 감정을 이상적 관찰자도 느낄 수 있을 때 옳거나 그르다. 이를 반영하여 다음과 같은 수정된 감수성 이론을 제안한다.

> (수정된 감수성 이론) 어떤 행동은 현실의 관찰자가 승인(불승인)의 감성을 가질 때, 도덕과 무관한 모든 사실을 알고 이해하며, 실천적으로 합리적인 이상적 관찰자 또한 승인(불승인)의 감성을 가질 수 있을 경우 오직 그러한 경우에만 도덕적으로 옳은(그른) 속성을 갖는다.

도덕 판단은 음식이나 음악 취향에 대한 주장과는 다르다. 자신의 사적인 선호를 설득하는 데는 이상적인 관점이 필요하지 않다. 나는 짜장면을 좋아하고 너는 짬뽕을 좋아한다. 그뿐이다. 그런데 도덕 영역에서 사람들은 나의 판단은 옳고 너의 판단은 틀렸다고 말한다. 자신은 진리를 말하고 있다고 생각한다. 왜 나의 판단이 옳은가? 이때 사람들이 자신의 도덕적 의견을 정당화하는 방식은 이상적 절차와 닮아 있다. 사람들은 종종 자신의 판단이 이상적 관점에서 나온 판단과 일치하거나 근접하다고 느낀다. 사람들은 도덕과 무관한 모든 정보를 알고 실천적으로 합리적인 어떤 완전한 존재가 있다면 자신의 판단을 승인하리라 자신한다. 때로 사람들은 무엇이 옳고 그른지 불명확할 때 자신이 전지전능한 존재라면 어떻게 판단할까 상상해 본다. 이는 우리가 일상에서 자주 경험하는 바다. 요컨대 도덕 판단이 가진 권위는 주관적 선호를 넘어선 곳에서 온다. 이상적 관찰자의 감성은 도덕 판단에 객관적 힘을 부여한다.

이상적 관찰자는 현실의 관찰자와 상호작용하며 실제 사람들의 사고, 감정, 행동에 영향을 미친다. 이로써 도덕 판단은 더욱 개선되고 향상된다. 내가 동성애를 불승인한다고 해보자. 나는 모든 동물이 자손 생산을 위해 이성을 사랑하게끔 태어난다고 주장한다. 동성애는 자연을 거스른다. 따라서 그르다. 하나 동성애가 본성에 맞지 않아서 나쁘다는 생각은 명백히 잘못이다. 자연 상태에서 많은 동물 종은 동성애 행동을 보인다. 동성애는 비정상적인 질병이 아니다. 게다가 자연적인 특질이 곧 도덕적 좋음을 함의하지도 않는다. 나는 동성애에 대한 정확한 지식을 갖지 못했다. 그래서 나는 잘못된 판단을 했다. 사람들은 나의 동성애 혐오를

혐오한다. 결국 나는 사회적으로 지탄받으며 자신의 무지에 대항하는 다른 사실에 부딪힌다. 나는 처음의 의견을 철회하도록 요구된다. 내가 동성애와 관련된 다른 사실을 받아들이고 자신의 편견에서 한걸음 물러나게 되면, 동성애에 대한 감정이 잘못임을 깨닫게 된다. 남에게 해를 끼치지 않는, 자유로운 개인의 욕구를 억압하는 행위는 배려/피해 가치에 어긋난다. 이상적 관찰자는 나의 동성애 혐오와 탄압을 불승인하리라. 오직 바른 지식에서 나온 감정만이 이상적 관찰자가 승인하는 감정이다. 이렇게 우리는 도덕적 논쟁과 그에 따른 교정을 거쳐 이상적 관점을 획득한다. 논쟁은 사실을 둘러싼 싸움이다. 도덕에서 이성이나 추론이 하는 역할은 올바른 도덕 판단을 위한 자료를 모으는 데 있다. 일단 명확한 지식을 갖고, 도덕과 무관한 사실에 대해 합의를 보면 특정 판단은 진화한 기초 가치 영역 중에 하나로 포섭된다. 그러면 도덕 판단에서 다시 즉각적인 감정 반응이 일어난다. 이제야 나의 감정은 이상적 관찰자 또한 승인하는 반응이 된다. 이상적 관찰자는 동성애에 대한 충분한 정보에서 나온 나의 감정을 승인한다.

프린츠가 인간이 보유한 선천적인 능력이라고 말한 메타감정은 이상적 관찰자와 현실의 관찰자가 연결된다는 주장을 뒷받침하는 좋은 근거다. 메타감정은 어떤 사안에 대한 행위자의 감정 반응이 적절한지를 점검하는 감정이다. 사람들은 자신이 벌인 못된 행동에 죄책감을 느끼지 않았을 때 죄책감을 느끼거나, 타인이 저지른 나쁜 행동에 분노하지 않았을 때 자신의 무심함에 수치심을 느낀다. 메타감정을 느끼는 사람은 가상적인 타인의 눈으로 자신의 감정을 평가한다. 메타감정은 인간 종이 사회적 동물로서 평판에 극도로 민감했기에 발생했다. 내 행동이 타인이나 집단에 수용될 수 있는지는 생존과 번식이 걸린 중요한 문제이다. 그 가상적인 타인의 눈이 바로 자신을 둘러싼 동료와 공동체, 나아가 다른 공동체 모두 동의할 수 있는 감성을 가진 이상적 관찰자이다. 사회 속의 인간은 가상적인 눈, 즉 이상적 관찰자가 자신의 행동을 승인하는지, 불승인하는지 신경을 곤두세우며 스스로 부정적인 감정을 느낄 경우 문제가 된 행동을 고치려 한다. 특히 주목할 만한 감정은 죄책감이다. 사람들은 죄책감을 표현함으로써 사회적 유대를 강화한다. 죄책감은 뉘우치고, 사과하고, 다시는 같은 행동을 반복하지 않겠다는 참회를 동반한다. 이는 다른 사람들에게 자신이 다시 협동을 재개해도 되는 동료임을 신호한다. 모든 연령 집단에서 죄책감을 느끼는 성향은 이후에 반사회적 행동을 저지를 확률을 낮추었다(Tangney et al. 2013). 죄책감은 자신이 올바른 감정과 행위를 보였는지 평가하는 자기반성적인 감정이다. 종교를 믿는 사람들은 신이 명한 행동을 하지 않았을 때 신의 이름 앞에 죄책감을 느낀다. 우주의 순리와 인간의 운명을 안다고 생각되는 신은 나를 평가하고 벌을 내릴 수 있으며 행동의 근간으로 삼기에 충분한 존재이다. 이상적 관찰자도 이와 비슷하다. 내가 죄책감을 느끼는 이유는 자신의 감정과 판단을 정당화하는 초월적 관점에 따르지 않았기 때문이다.

그러나 메타감정이 있기에 이상적 관찰자를 신처럼 현실의 관찰자와 별개인 존재로 생각할 필요가 없다. 이상적 관찰자는 현실의 관찰자이다. 인간은 자기객관화와 자기반성이 가능한 독특한 동물이다. 인간은 문화적 압력과 편견에서 잠시 벗어나 자신을 타인의 눈으로 보며 감정을 객관화할 때 이상적 관찰자의 관점에 서게 된다. 나는 묻는다. 내 감정은 올바른가? 예를 들어 보자. 무슬림 남성인 나는 여성에게 부르카를 씌워 통제하는 행동이 잘못임을 모를 수 있다. 나는 여성의 신체를 가리고 가정에만 억압해 두는 문화에 아무런 불승인의 감정도 느끼지 않는다. 그런데 내가 여태까지의 믿음을 무너뜨리는 사건을 경험한다고 하자. 어머니와 여동생이 부르카를 입지 않았다는 빌미로 낯선 남성들에게 구타당하는 변고를 당했다. 나는 절망하고 분노하며 부르카 관습을 용인했던 감정과 행동을 자책한다. 이제 나는 습관과 타성에서 벗어나 새로운 사실을 접하고 배우며 가상적인 타인의 눈으로 자신의 감정을 객관화한다. 마침내 나는 여성에 대한 강제와 구속이 인간의 보편적인 가치에 따라 그른 행동임을 알게 된다. 나는 이전의 판단을 수정한다. 그리하여 나는 더 올바른 감성을 갖게 된다. 이렇게 메타감정은 현실의 관찰자가 도덕 외적인 사실에 대한 정보를 습득케 하고 이후 더 나은 도덕 판단을 하도록 추동한다. 현실의 관찰자는 점진적으로 이상적 관찰자가 된다.

프린츠는 도덕의 객관성을 부정하는 장에서 한 절을 할애해 이상적 관찰자 이론을 비판한다. 요점은 이렇다. 첫째, 이상적 관찰자는 공평무사하지 않다. 도덕 판단은 가치에 대한 반응으로서 감정으로 표현된다. 그런데 가치와 감정은 문화의 산물이다. 따라서 이상적 관찰자의 반응은 문화마다 다를 수 있다. 둘째, 이상적 관찰자는 우리가 반대하는 행동을 승인할 수 있다. 문화적 영향을 거부한다면 우리는 이상적 관찰자의 마음이 자연 상태라고 봐야 한다. 그러나 자연은 온화하지 않다. 문명화되지 않은 존재는 무섭고 잔인하다. 야만적인 이상적 관찰자는 살인을 허용할 수도 있다. 셋째, 이상적 관찰자의 도덕 판단이 권위를 갖는지 의문이다. 많은 사람들은 도덕 판단을 할 때 초월적 관점을 참조하지 않는다. 그저 자신이 속한 공동체의 의견을 따를 뿐이다. 왜 일부일처제를 옳다고 하는가? 단지 우리 집단이 일부일처제를 가치 있게 여기기 때문이다(Prinz 2007: 142-144).

세 가지 지적은 모두 프린츠가 도덕 비선천주의자이며 도덕 상대주의자의 입장을 견지하는 데서 비롯한다. 하지만 앞서 논의했듯이, 기초 가치와 감정의 연결은 선천적이다. 정상적인 발달 과정을 거치면 인류는 매우 쉽게 보편적인 가치 목록을 습득한다. 여섯 가지 도덕 기반은 범세계적으로 나타난다. 올바른 판단을 가로막는 주변 사실을 제한다면 논쟁 중인 사안은 기본적인 도덕 가치로 포섭되어 감정 반응은 수렴된다. 그러므로 완전한 정보를 가진 이상적 관찰자의 반응은 문화를 막론하고 동일하다. 또한 자연 상태가 미개를 허락할 수 있다는 프린츠의 주장은 인간 본성에 대한 오해다. 인간의 성격은 단일하지 않다. 우리는 때로 타인을 이용하는 이기적 면모를 보이지만, 낯모르는 사람을 위해 자신을 희생하는 이타적 존재이기

도 하다. 도덕 행동을 만드는 문화적 압력이 없어도 인간은 도덕적이고자 한다. 즉 도덕은 인간 본성의 일부이다. 가치에 대한 존중은 조상들의 삶에 중요했다. 이상적 관찰자는 그 본성을 실현하는 요소이다. 마지막으로 이상적 관찰자의 반응이 아무런 구속력이 없다는 평은 동의하기 힘들다. 우리 집단이 일부일처제를 옳다고 여긴다는 대답은 도덕 판단에 대한 정당화가 될 수 없다. 그러면 상대방은 다시 "왜 당신네는 그렇게 생각하느냐?"고 되물을 수 있다. 사람들은 도덕 문제에서 주관적 의견을 정답이라 생각지 않는다. 상대방을 설득시키고 행동의 변화를 일으키려 한다면 객관적인 토대를 제공해야 한다. 그렇기에 도덕 판단에서 충분한 정보와 공평무사함은 실천을 야기하는 강력한 척도다. 프린츠의 직관은 틀렸다. 많은 사람들은 자신의 판단을 공고히 하고자 초월적 관점에 서려 한다.

계속된 논의에서 알게 되겠지만, 프린츠가 이상적 관찰자 이론에 가한 세 가지 비판은 조이스가 프린츠의 감수성 이론에 제기한 문제와 유사하다. 프린츠의 의도와 달리 세 물음은 자신에 대한 공격이며 되레 문제를 해결하는 대안은 수정된 감수성 이론이다.

프린츠가 제시한 감성이라는 성향을 이상적 관찰자의 성향으로 대체함으로써 극단적 상대주의의 문제는 해결된다. 도덕 판단은 주관적 감정에만 근거하지 않는다. 행위자가 느끼리라 기대되는 감정은 이상적 관찰자가 느낄 수 있는 감정으로 제한된다. 이상적 관찰자는 도덕과 무관한 모든 사실을 알고, 실천적으로 합리적인 관찰자이다. 이상적 관찰자는 잘못된 지식에서 나온 행위자의 감정을 불승인한다. 올바른 도덕 판단은 정확한 사실에서 나온 감정이어야 한다. 그때 도덕 판단은 진화한 기초 가치로 포섭된다. 설사 당신의 감정이 아동학대가 옳다고 승인해도 이상적 관찰자의 관점에서 아동학대는 도덕적으로 그르다. 당신의 감정 반응은 잘못된 사실에 근거했다. 당신의 판단은 틀렸다. 우리는 모두 아동학대를 용인하는 가치를 거부한다. 아동학대는 기초 가치에 어긋난다. 따라서 수정된 감수성 이론은 도덕 판단의 객관성을 회복한다. 인간에게는 자신의 마음에 가상적인 타인의 눈을 들이는 능력이 있다. 이런 시점은 감정과 행동 전반에 큰 영향력을 행사한다. 우리는 누가 보지 않아도 자신을 객관적으로 평가한다. 사회적 동물인 인간에게 이상적 관찰자를 소환하는 역량은, 역시 사회적 관계의 규제인 도덕 판단을 위한 적응이다.

수정된 감수성 이론이 도덕의 객관성을 잘 설명한다 해도 여전히 남는 문제가 있다. 바로 도덕의 실천성이다. 이상적 관찰자는 현실의 관찰자에게 어떻게 도덕적 동기를 줄 수 있는가, 현실의 관찰자는 이상적 관찰자가 승인할 수 있는 행동을 왜 해야 하는가? 조이스가 주장하듯 도덕 명령에는 불가피한 권위가 있다. 도덕적 이유는 욕구나 득실과 상관없는 정언적 이유이다. 이상적 관찰자는 나에게 그런 이유를 줄 수 있는가?

4.2. 도덕 판단의 정언성 설명

조이스는 도덕 판단이 품은 실천적 힘을 강조한다. 도덕 판단은 불가피한 권위에 대한 믿음을 표현한다. 그리하여 도덕 판단은 개인적인 욕구나 목적에 상관없이 따라야 할 정언적 이유를 준다. 일상에서도 우리는 도덕 이유가 다른 이유보다 우선한다고 생각한다. 이러한 도덕의 정언성은 협상 불가능한 핵심이다. 정언 명령 없는 도덕은 도덕이 아니다. 감수성 이론은 도덕의 정언성을 수용할 수 있는가?

프린츠는 감정이란 곧 동기적 상태이므로 감정을 일으키는 반응 의존적인 도덕 속성이 우리 행위를 인도한다고 주장한다. 그러나 조이스는 감수성 이론이 기껏해야 어떤 행위를 하도록 유도하거나 자극할 뿐 요구하지는 못한다고 반박한다. 어떤 사람이 특정 상황에서 낙태에 대해 부정적 감정을 야기하는 성향, 즉 감성 덕에 분노라는 감정을 느꼈다고 하자. 이때 분노라는 감정 반응은 단지 그가 낙태를 하지 않으려는 마음을 고무할 뿐 '하지 말아야 함'이라는 도덕 요구는 아니다. 도덕 요구는 감정이나 욕구와 상관없이 주어진다. 한데 반응 의존 속성은 반드시 개인의 감정, 욕구와 연결된다. 어떤 행동이 도덕적으로 그르다는 믿음 자체는 행위 이유를 주지 못한다. 도덕적 그름은 그 행위를 금하는 감정과 연결되어야 비로소 이유를 주는 힘을 가진다. 감수성 이론은 흄적인 도덕 자연주의이다. 행위 이유는 가언적이다. 가언 명령은 다음과 같이 표현된다. "부정적인 감정을 느끼지 않으려면 도둑질을 하지 마라!" 가언 명령은 타산적 계산이다. 도덕 이유보다 더 이로운 이유가 있다면 도덕 행동을 하지 않을 수도 있다. 그렇다면 적절한 감정이 없을 때도 왜 우리는 도덕에 신경 써야 하는가, 사악한 감정에 따라 행동한 자를 비난하는 도덕적 확신은 어디서 오는가? 여기에 대답할 수 없다면 감수성 이론은 도덕이 부과하는 특별한 의무를 설명하지 못한다. 이를 '실천적 유관성 문제(practical relevance problem)'라 한다(Joyce 2008: 513-514; 2009: 256-258).

조이스가 제기하는 비판은 매우 근본적이다. 그는 도덕 판단의 정언적 규범성은 도덕을 다른 영역과 구별하는 제일 요소라고 생각한다. 정언성에 근거하지 않는 감수성 이론과 같은 도덕 자연주의는 애초에 뿌리부터 잘못되었다. 정말 그런가? 조이스에 맞서 우리도 근원적인 질문을 던져 보자. 조이스는 도덕의 정언성을 의심할 수 없는 진리라고 전제한다. 그러나 정언 명령이라는 개념은 얼마나 확고한가, 과연 정언적이어야만 도덕적인가, 왜 도덕 사실은 정언적 이유를 주어야 하는가, 어떤 감정도, 욕구도 없이 행위를 위한 동기가 어떻게 생길 수 있을까, 좋음을 실천하는 일이 내게 어떤 이익이 되지 않거나 바라는 바와 합치하지 않을 때도 그 행위를 해야만 할까? 도덕 명령이 다른 이유를 압도하는 요구라는 직관 못지않게 도덕적 실천이 감정과 밀접한 관련을 맺는다는 직관도 상식에 잘 부합한다. 이에 이 글은 정언성이 도덕의 협상 불가능한 핵심이 아니라고 주장하고자 한다. 도덕이 정언적이지 않다고 해서

도덕의 실천적 힘이 사라진다고 볼 수 없다. 도덕 사실은 정언적 이유를 줄 필요가 없다. 먼저 프린츠의 대응부터 살펴보자. 프린츠 또한 도덕 명령이 정언적이라는 칸트식 전통을 의심한다.

프린츠는 자신의 감수성 이론이 도덕 요구가 정언적이라는 직관을 설명할 수 있다고 주장한다. 그 요지는 첫째, 감수성 이론은 감정과 독립적이라는 정언적 특징을 포함한다. 도덕 가치는 감정을 느낄 수 있는 감정적 성향인 감성으로 구성된다. 즉 어떤 행위의 그름은 현재 느끼는 감정이 아니라 해당 감정을 일으킬 수 있는 성향으로 규정된다. 따라서 지금 살인에 대해 분노하지 않더라도 살인은 도덕적으로 그르다. 살인에는 여전히 분노를 일으키는 성향이 있다. 둘째, 순수 이성만으로는 행위 이유를 줄 수 없다. 흄적인 행위 이유가 오히려 도덕의 실천성을 더 잘 설명한다. 다시 도덕에 관한 근본적인 질문을 생각해보자. 왜 나는 도덕 규칙을 따라야 하는가? 칸트의 정언 명령은 여기에 답할 수 없다. 규칙에 따르지 않으면 비합리적이어서? 그러나 왜 합리적이어야 하는가? 어떤 사람은 타인을 착취하는 행동이 비합리적임을 알아도 눈앞에 이익 때문에 전혀 개의치 않기도 한다. 요컨대 순전한 합리적 규칙은 그 자체로 동기를 일으키기 어렵다. 행위자는 도덕 규칙을 잘 이해하더라도, 규칙에 따르고 싶은 욕구가 없다면 어떤 요구에도 움직이지 않는다. 규칙은 무엇보다도 감정과 연관되어야 행위를 추동한다. 나는 이성이 아니라 감정적으로 살인에 무심하기 어렵다. 살인은 왠지 모르지만 나쁘고 불쾌하다. 살인을 용인하는 일은 심각한 감정적 비용을 초래한다. 따라서 나는 살인을 금한다. 결론적으로 프린츠는 감수성 이론이 약한 정언 명령이라고 주장한다. 도덕 명령은 내가 피곤하거나 근심에 빠져 있다고 해서 사라지지 않는다. 명령은 언제나 나와 묶여 있다. 그러나 도덕 명령이 나의 감정적 성향에 바탕을 두지 않는다면 그에 따라야 할 의무는 결코 생기지 않는다(Prinz 2007: 128-136). 이 글은 프린츠가 제시한 의견에 전반적으로 동의한다. 따라서 도덕 판단의 실천적 힘이 약한 정언 명령이라는 그의 주장을 보강하여 조이스의 비판에 대응하고 수정된 감수성 이론이 어떻게 행위 동기를 줄 수 있는지 논의하겠다. 정언 명령부터 고찰해보자.

정언 명령은 단일한 개념이 아니다. 필리파 풋(P. Foot)은 정언 명령을 두 가지로 구분한다. 첫째, 도덕 의무는 사람들의 욕구와 별개로 적용 가능하다는 의미에서 정언적이다. 둘째, 도덕 의무는 사람들의 욕구와 별개로 해야 할 이유를 준다는 점에서 정언적이다(Foot 1972). 하나는 '적용 가능성'이고 다른 하나는 '이유를 주는 힘'이다. 풋은 적용 가능성과 이유를 주는 힘이 서로 다르다고 주장한다. 정언적으로 적용 가능하지만 복종을 위한 이유를 부과하지 못하는 의무가 있다. 에티켓이 그렇다. 식사 중에 코를 푸는 행위는 에티켓에 어긋난다. 에티켓은 욕구와 상관없이 적용된다. 그러나 행위자가 에티켓을 따르는 데 무관심하고, 따르지 않았을 때 어떤 일이 생길지 전혀 신경 쓰지 않는다면 에티켓을 지켜야 할 이유는 없다. 그렇

더라도 식사 중에 코를 풀지 말라는 에티켓이 거짓이 되거나 철회되지는 않는다. 에티켓은 여전히 정언적으로 적용된다. 풋은 도덕도 매한가지라 본다. 도덕 판단은 정언적으로 적용 가능하지만, 도덕적이고자 하는 욕구나 바람이 없다면 행위를 위한 이유는 생기지 않는다. 도덕은 실상 가언적인 체계이다. "이유를 주는 힘이 도덕 판단이 나타내는 규범적 성격을 보장하지 못한다는 점은 명백하다. 도덕 판단은 예의범절에 대한 판단, 사교 규칙에 대한 판단 등등과 같은 의미에서 규범적이다(Foot 1972: 310)."

흄적인 도덕 자연주의자는 도덕 판단이 주는 정언적 이유는 거부하고 적용 가능성만을 받아들인다. 도덕 요구는 욕구와 상관없이 불가피하게 주어진다. 그러나 적절한 욕구 없이 행위 이유도 없다. 도덕의 정언성은 바로 적용 가능성이다. 프린츠가 택한 약한 정언 명령이라는 전략도 이와 같다. 그렇다면 다시 묻자. 적용 가능성만으로 충분한가, 도덕의 정언성은 해결되었는가?

조이스는 정언 개념에 대한 풋의 구분을 인정한다. 그러나 그는 어떤 체계가 도덕이라 불리려면 도덕 명령의 정언적 적용 가능성과 이유를 주는 힘을 통합해야 한다고 반박한다. 앞에서 조이스가 강조했듯이, 도덕과 에티켓은 근본적으로 다르다. 도덕은 단순히 사회적 관습이나 제도적 구성물이 아니다. 왜 그런가? 우리는 누군가 악한 행동을 했을 때 그에 대한 도덕적 책임을 묻는다. 도덕 위반자에게 책임을 묻는 행위는 그에게 해당 행위를 하지 말아야 할 이유가 있었음을 전제한다. 그런 이유는 정언적이다. 나치에 대한 비난은 그들이 도덕 의무를 따르려는 욕구가 없었다는 사실 때문에 면제되지 않는다. 나치는 도덕 요구에 신경 쓰려는 마음이 전혀 없었다는 핑계로 처벌을 피할 수 없다. 이에 조이스는 묻는다. 나치를 책망할 수 없는 도덕 체계를 도덕이라 부를 수 있는가(Joyce 2001: 43)? 그럴 수 없다. 정언적 도덕 이유가 있다는 생각은 책임을 묻는 도덕적 실천에 뿌리 깊게 박혀있다. 따라서 조이스는 정언적 이유를 부정하는 도덕 자연주의를 부정한다.

그러나 테렌스 쿠네오(T. Cuneo)는 정언적 이유에 대한 직관이 틀렸다고 주장한다. 도덕 행위가 정언성과 매우 밀접하다는 관찰만으로 실제로 정언적 이유가 있다고 말할 수는 없다. 우리가 굳게 믿던 어떤 습속들이 나중에 거짓으로 판명되는 사례는 흔하다. 또한 도덕적 실천이 정언적 이유와 무관하다해서 도덕을 버릴 필요도 없다. 그저 우리는 정언 명령이라는 과거의 개념을 수정하면 된다. 상대성 이론은 참이지만, 우리는 사건이 동시적으로 일어난다고 생각하는 습관을 버리지 않는다. 다만 동시성에 대한 이해를 조금 변경할 뿐이다(Cuneo 2011: 114). 따라서 쿠네오는 흄적인 도덕 자연주의가 정언적 이유를 따르는 양 보이는 도덕 행위를 설명할 수 있다고 주장한다. 그의 논의를 따라가 보자.

쿠네오는 도덕 자연주의에 반대하는 조이스의 '정언성 논증(categoricity argument)'을 다음과 같이 정리한다.

(1) 도덕 사실이 있다면 필연적으로 정언적 이유가 있다.
(2) 정언적 이유는 없다.
(3) 따라서 도덕 사실은 없다.
(4) 도덕 사실이 없다면 도덕 자연주의는 거짓이다.
(5) 그러므로 도덕 자연주의는 거짓이다(Cuneo 2011: 118).

조이스는 맥키의 오류 이론에 기반을 둔 진화적 도덕 반실재론자이다. 도덕 판단은 정언적 이유에 대한 믿음을 표현한다. 하나 정언적 이유를 주는 도덕 사실은 없다. 도덕의 정언성은 진화의 산물이다. 자연선택은 도덕 판단을 정당화하는 도덕 사실과는 상관없는 과정이다. 도덕 판단은 모두 거짓이다. 반대로 도덕 자연주의자는 도덕 사실이 있다고 생각한다. 그러한 사실은 흡적인 이유를 준다. 행위 이유는 적절한 욕구가 있을 때 발생한다. 정언적 이유는 없지만 도덕 사실은 있다. 도덕 자연주의자는 조이스가 제기한 논증의 결론을 거부한다.

쿠네오는 정언성 논증의 결론을 물리치고자 먼저 우리에게 상투적인 도덕적 진리가 있음을 상기시킨다. “재미로 살인하는 짓은 나쁘다.”, “체면을 지키려고 배우자에게 거짓말 하는 행동은 나쁘다.”, “단지 기분이 상했다고 약속을 어기는 일은 나쁘다.” 등등. 이런 언명들은 널리 수용되고 누구나 참으로 생각한다. 따라서 좋은 도덕 체계라면 상투적인 도덕적 진리를 포함해야만 한다. 사람을 죽이는 일을 즐거워하는 사람이 있다면 우리는 그가 미쳤거나, 제대로 된 교육을 받지 못했다고 생각하리라. 그런 사람은 뭔가 이상하다(Cuneo 2011: 119).

이러한 사실을 마음에 두고 조이스의 비판을 다시 보자. 조이스는 일상의 도덕적 직관을 중히 여긴다. 우리가 논증의 첫 번째 전제를 받아들여야 하는 까닭은 도덕적 책임을 묻는 매일의 사고방식과 실천 덕분이다. 이는 도덕 체계가 정언적일 때 가능하다. 나치는 인종 학살을 중단해야 할 욕구와 무관한 이유가 있었다. 쿠네오가 공격하는 지점은 여기다. 쿠네오는 조이스를 좇으면 상투적인 도덕적 진리가 파기된다고 주장한다(Cuneo 2011: 119). 상투적인 도덕적 진리는 정언적이지 않기 때문이다. 정언적 도덕 판단은 조건문이 없어야 한다(“사람을 살인하지 마라!”). 상투적인 도덕적 진리처럼 ‘쾌락 때문에’, ‘체면 때문에’라는 요건이 들어가는 순간, 누군가는 자신이 진실로 쾌락과 체면을 원한다고 말하면서 도덕 명령을 어길 수 있다. 조이스는 그런 사람에게 도덕적 책임을 물을 수 없다고 말했다. 그러나 우리 직관은 정반대다. 우리는 쾌락 때문에 살인을 저지른 자를 악인이라고 비난한다. 욕구는 악인의 명령 불복종을 정당화하지 못한다. 즉 욕구에 토대를 둔 상투적인 도덕적 진리 또한 도덕적 사고와 실천의 핵심이다. 상투적인 도덕적 진리가 없는 도덕 체계는 좋은 도덕 체계일 수 없다.

여기서 쿠네오는 묻는다. 왜 상투적인 도덕적 진리를 버려야 하는가? 우리는 논증의 첫 번째 전제, 정언적 이유가 있다는 전제를 거부할 수 없는가? 그럴 수 있고 그래야 한다. 반드시

정언적 이유를 받아들일 필요는 없다. 이를 위해 쿠네오는 두 가지 논증을 제시한다.

> A: 도덕 사실이 있으면 정언적 이유가 있다는 주장은 개념상 필연적이다. 그러나 정언적 이유는 없다. 따라서 도덕 사실은 없다. 하지만 도덕 사실이 없으면 행위자가 그릇되게 행동했다라는 도덕적 결함을 나타낸다고 할 수 없다. 일상 조건에서 행위자가 그냥 그렇게 하고 싶어서 약속을 어기고 사람을 죽였다고 해보자. 이는 상투적인 도덕적 진리가 거짓임을 보여준다. 왜냐하면 일상 조건에서 행위자가 단지 그렇게 하고 싶어서 약속을 어기고 사람을 죽였지만 그릇되게 행동했다라는 도덕적 결함을 내보인다고 할 수 없기 때문이다.

> B: 일상 조건에서 행위자가 단지 그렇게 하고 싶어서 약속을 어기거나 사람을 죽인다면, 그가 그릇되게 행동했다라는 도덕적 결함을 나타낸다는 주장은 개념상 필연적이다. 일상 조건에서 행위자가 단지 그렇게 하고 싶어서 약속을 어기거나 사람을 죽인다고 해보자. 여기서 그가 그릇되게 행동했다라는 도덕적 결함을 내보인다는 사실이 따라 나온다. **그가 그릇되게 행동했다는 도덕적 결함을 드러낸다면 도덕 사실은 있다. 그는 도덕적 결함을 나타냈고 도덕 사실은 있다. 그러나 정언적 이유는 없다. 따라서 정언성 논증의 첫째 전제는 거짓이다. 도덕 사실은 있지만 정언적 이유는 없다**(Cuneo 2011: 120, 강조는 인용자).

쿠네오는 묻고 답한다. 우리에게 B가 아니라 A를 택할 이유가 있는가? 없다. 도덕 사실의 정언성을 붙잡는 대가는 크다. 그러려면 정언적 이유만큼 중요한 요소를 포기해야 한다. 하지만 사람들은 보통 즐거우려고 살인을 범하는 사이코패스의 행동이 도덕적으로 그른 속성을 가진다고 생각한다. 이런 직관을 내던지라는 요구는 너무나 독단적이다. A는 비현실적인 회의주의다. A를 택해야 할 좋은 이유는 없다(Cuneo 2011: 120-121).

그렇다면 조이스는 다시 되물을 수 있다. 그럼 정언적 이유는 도대체 무엇이고, 어디에 있는가, 흄적인 도덕 이유는 타산적 고려를 멈추게 하는 계산 중지의 기능을 할 수 있는가? 더불어 도덕 자연주의가 옳다 해도 도덕 판단은 근본적으로 오류다. 정언적 이유가 없는데도 있는 양 말하고 행동하니까.

이에 대해 쿠네오는 흄적인 도덕 이유도 정언적이라고 변론한다. 어떻게 가능한가? 일상의 도덕적 사고와 실천은 오류를 포함한다. 다시 말해, 우리는 도덕 이유의 본성을 정언적이라 잘못 생각한다(Cuneo 2011: 123). 쿠네오는 윤리학자 마크 존스톤(M. Johnston)을 인용하여 도덕 이유의 정언성을 포켓볼을 배우는 방식에 비유한다. 때로 구멍에서 먼 공을 넣으려 예상한 경로를 따라가는 유령 공이 있다고 상상하는 방법이 도움이 된다. 나는 흰 공을 유령 공을 향해 쳐서, 흰 공이 목적했던 색 공을 건드리게 만든다. 실제로 유령 공은 존재하지 않지만 '유령 공 상상하기'는 당구를 익히는 좋은 길잡이다(Johnston 2010: 17).

쿠네오는 유령 공처럼 도덕 이유를 정언적으로 생각하는 행위는 매우 자연스러우며 도덕

요구에 따르는 최선의 방법이라고 주장한다. 정언적인 사고 덕에 우리는 나와 타인의 복지에 마음 쓰고, 도덕이라는 이름으로 묶인다. 이렇게 도덕적으로 행동하게끔 진정한 이유를 주는 욕구를 형성하고 유지하는 일은 유용하다. 당구의 경우는 의식적으로 거짓 표상을 만들지만 도덕은 그렇지 않다. 우리는 교육 때문이든 무엇이든, 무의식적으로 도덕 이유가 정언적이라고 생각한다(Cuneo 2011: 123-124).

위의 논의를 참고하면, 정언 명령은 도덕의 협상 불가능한 핵심이 아니다. 도덕 사실은 정언적 이유를 주는 마음 독립적인 사실이 아니다. 도덕 사실은 욕구에 대한 반응으로 구성된다. 물론 그 반응은 어떤 행위를 승인, 불승인하는 감정적 성향이다. 따라서 도덕적으로 행동하게 만드는 이유는 오직 감정과 연결되었을 때만 발생한다. 그런데 감정으로 구성된 도덕 사실은 성향적 속성이라, 어떤 조건에서 행위자는 도덕 요구에 걸맞는 감정을 느끼지 못할 수도 있다. 이는 늘 경험하는 사실이다. 타인의 복지를 증진하라는 도덕 명령은 종종 개인적인 계획, 목적, 소망과 충돌한다. 때로 우리는 다른 욕구를 내세우며 도덕 명령을 거부한다. 예를 들어 나는 부자면서도 더 많은 돈을 축적하려고 가난한 자에게 기부하길 꺼려할 수 있다. 즉 내가 따르려는 이유에는 경쟁하는 이유가 있다. 이중에서 오직 도덕 이유만이 그 자체 권위가 있음을 보증하는 근거는 없다. 도덕 이유가 진정으로 행위를 촉구하는 결정적인 이유가 되려면 나의 현실적인 목적, 욕구, 소망, 이해도 계산에 들어가야 한다. 그리하여 도덕 이유는 무엇보다 해당 행위에 대한 나의 감정과 연결되어야 한다. 이제 문제는 하나다. 흄적인 이유에 근거한 도덕 판단은 여전히 다른 이유를 압도하는 성격을 가지는가, 내가 왜 도덕적이어야 하는지 알려주는가, 그리하여 도덕 의무를 다하지 못한 사람에게 책임을 물을 수 있는가? 프린츠는 그렇다고 답했다. 이 글은 프린츠가 제시한 논거에 일부 동의한다. 그러나 프린츠의 감수성 이론보다 수정된 감수성 이론이 도덕 행위의 정언적 성격과 압도성을 더 잘 설명한다.

감수성 이론에 따르면 부자인 내가 다른 욕구 때문에 기부를 거절한다고 해도 기부는 도덕적으로 옳은 속성을 가지며 기부를 종용하는 도덕 요구도 사라지지 않는다. 현재 내가 도덕 감정을 느끼지 못한다는 사실은 문제가 아니다. 기부의 옳음은 승인의 감정을 부를 수 있는 성향으로 규정되기 때문이다. 프린츠는 이러한 감정 반응을 나타내는 주체가 현실의 관찰자라고 생각한다. 그렇다면 현실의 관찰자는 언제 어떤 조건에서 기부에 응하는가? 다시 말해, 언제 어떤 조건에서 기부를 하는 데 승인의 감정을 느끼는가? 현실의 행위자는 도덕 판단을 둘러싼 외적 사실에 관한 충분한 정보가 없거나 기타 사실에 무지하여 자신과 타인에게 좋은 행위가 무엇인지 모를 수 있다. 실제 행위자는 이미 돈이 많은데도 돈을 모으면서는 계속 즐거움을 경험하지만, 기부에는 아무런 감정도 느끼지 못할 수 있다. 따라서 현실의 나는 돈을 모으는 기쁨이 더 올바르다고 판단한다. 내게 돈에 대한 욕구는 기부보다 더 강하다. 요컨대

현실의 관찰자가 가진 감정적 성향만으로는 특정 상황에서 우리가 바라는 도덕 행동을 하리라고, 즉 우리가 바라는 도덕 감정을 느끼리라고 기대할 수 없다. 아무런 조건 없이 기부를 하려는 도덕적 욕구가 이기적 욕구를 이길 수 있다는 주장은 너무 허황되다. 프린츠는 바로 그런 주장을 하고 있다. 그러나 단지 가능성만으로 일상적인 도덕 행동을 온전히 파악하기는 역부족이다.

수정된 감수성 이론은 왜 도덕적 욕구가 이기적 욕구를 이길 수 있는지 설명할 수 있다. 누누이 얘기했듯, 우리 진화 역사에서 도덕 행동은 생존과 번식 적합도를 높이는 데 중요했다. 도덕적인 사람은 안정적인 협력자이며 신뢰할 만한 배우자였다. 도덕은 특히 조상들이 헤쳐 나가야 했던 특정한 적응 문제에서 요구되었다. 삶의 중심인 이런 영역들에서 인간 종이 공유하는 기초 가치가 생겼다. 기초 가치에는 문제 해결에 유용한 행동을 산출하는 적절한 감정 반응이 연결되었다. 승인의 감정을 부르는 행위는 성취해야 할 좋은 가치를, 그 반대는 금지해야 할 나쁜 가치를 신호한다. 현대 도덕 심리학과 뇌신경과학이 지지하듯 감정과 동기 사이에는 필연적인 관계가 있다. 개체는 긍정적인 감정을 유발하는 행동은 취하고 부정적인 감정을 유발하는 행동은 피하려 한다. 기초 가치와 감정이 선천적으로 연결돼 있다는 사실은 도덕적이고자 하는 욕구가 인간 본성의 일부임을 입증한다. 인간은 감정적으로 그렇게 해야 한다고 느끼기에 도덕 행동을 한다.

한데 앞서 보았듯이 우리가 늘 도덕 행동에 승인의 감정을 느끼지는 못한다. 개체는 현실이 부과하는 여러 제약 때문에 어떤 욕구에 잘못된 감정을 느끼거나 아예 무심할 수 있다. 따라서 현실의 관찰자는 오직 자신을 벗어난 관점에 섰을 때만 진정으로 좋은 욕구에 승인의 감정을 느낄 수 있다. 즉 충분한 정보를 알고, 실천적으로 합리적인 이상적 관찰자라면 승인의 감정을 일으키는 욕구가 무엇이고 불승인의 감정을 일으키는 욕구가 무엇인지 알 수 있다. 그리하여 내가 승인하거나 불승인하는 욕구는 이상적 관찰자의 눈으로 평가된다. 이상적 관찰자 또한 승인의 감정을 느끼는 욕구는 그 자체로 추구할 만한 목적이다. 동기 내재주의적 관점에서 나는 필연적으로 해당 욕구를 추구한다.

돈만 모으는 수전노는 자신의 행동이 지나친 욕심이라는 사실을 모른다. 그는 끝없는 축적에 삶을 낭비하며 공동체의 존경과 사랑도 받지 못한다. 그는 재산을 노리는 아첨꾼이나 범죄자의 표적이 된다. 그는 정말로 행복을 알지 못한다. 그러므로 이상적 관찰자는 무지한 부자의 맹목적인 탐식을 불승인한다. 가련한 부자의 이기적 욕구는 이상적 관찰자의 관점에 설 때 비로소 도덕적 욕구로 대치된다. 이는 단지 추측이 아니다. 인간은 자신의 관점을 객관화할 때 도덕적으로 변모한다. 한 예로, 사람들은 가상의 눈만 접해도 더 이타적으로 행동한다. 슈퍼마켓 계산대에 놓여 있는 기부함에 사람 눈과 비슷한 그림을 붙이자 그렇지 않았을 때보다 기부 금액이 59%나 뛰었다(Powell et al. 2012). 이렇게 내 밖에 존재하는 눈은 자신의

사고와 감정을 객관화하도록 만든다. 편협한 시각에서 벗어날 때 더 도덕적인 사람이 된다는 연구는 도덕 욕구가 인간 본성임을 알려주면서, 그 욕구에 따르게 됨을 시사한다.

그렇다면 우리는 수정된 감수성 이론으로 도덕 명령에 따르지 않은 사람을 단죄할 수 있을까? 사람들은 흔히 강간범이 저지른 행동이 성욕에서 비롯됐다면 죄가 면제된다고 생각한다. 욕구는 합리적 비판의 대상이 아니라는 직관 때문이다. 왜 욕구는 옳고 그를 수 없는가, 우리 통념이 잘못된 건 아닐까? 수정된 감수성 이론은 그렇다고 주장한다. 누군가 자신의 권력이 무소불위며 처벌받지 않을 가능성이 있다고 믿어도 성욕이 없다면 강간을 실행할 동기는 생기지 않는다. 행위 이유에 욕구가 관련된다면 욕구 또한 비판의 대상이 되며, 되어야 한다. 하면 어떤 욕구가 옳고 그른가? 진화한 기초 가치에 대한 욕구이다. 진화한 기초 가치는 행위자가 욕구하는 도덕적 목적이다. 진화한 기초 가치는 도덕적 올바름과 그름이, 긍정적인 감정을 주는 욕구와 부정적인 감정을 주는 욕구와 관계됨을 보여준다. 나는 약자를 괴롭히려는 욕구에 불승인의 감정을 느끼며, 보호하려는 욕구에 승인의 감정을 느낀다. 역시나, 현실의 관찰자는 무지에서 벗어나 진화한 기초 가치로 특정 욕구에 대한 자신의 감정을 평가할 위치에 설 필요가 있다. 즉 내가 충분한 정보를 알고, 실천적으로 합리적인 이상적 관찰자의 감정을 따른다면 강제로 타인을 범하려는 욕구를 불승인하리라. 하나 강간범은 이상적 관찰자가 승인할 수 있는 적절한 욕구를 갖지 않았다. 그는 타인에 대한 무자비한 폭력인 강간을 승인하는 감정을 느꼈다. 그는 자신의 그릇된 욕구를 이상적 관점에서 평가하지 않았다. 이런 욕구에 느끼는 감정은 과오이며 비난받아야 한다. 우리는 때로 어떤 도덕 행동을 평가할 때 행위자가 적절한 감정을 느꼈느냐를 중요하게 생각한다. 우리는 시체를 훼손하며 아무런 감정의 동요도 느끼지 않는 살인자를 보며 경악하고, 그런 살인자에게 분노를 느끼지 않는 사람을 힐난한다. 이는 진화한 기초 가치와 연결된 정당한 욕구가 있기 때문에 가능하다. 부정적인 감정을 주는 욕구는 도덕적으로 그르다. 따라서 수정된 감수성 이론은 도덕 위반자에게 책임을 물을 수 있다.

외려 조이스의 진화적 폭로 논증은 그가 강조한 도덕의 실천성을 설명하는데 부적합하다. 조이스는 도덕이 자연선택이 만든 착각에 불과하다는 진실을 목도해도 그 중요성은 사라지지 않을 거라 자신한다. 도덕은 사회적 관계를 규제하는 기제로서 여전히 실천적으로 유익하기 때문이다. 도덕은 '유용한 허구'이다. 그러나 도덕 명령이 불가피하다고 진정으로 믿는 사람과 단지 쓸모가 있다고 생각하는 사람 중 누가 더 충실히 도덕 규칙을 이행하겠는가, 누가 더 도덕적인 사람인가? 물론 도덕을 정말로 믿는 사람이다. 도덕이 거짓임을 아는 사람은 무언가 이득이 있을 경우에만 도덕적으로 행동하거나 도덕적인 외양을 꾸밀 수 있다. 그렇기에 누구도 속임수지만 도움이 되니까 따르는 행동을 진짜 '도덕'이라 보지 않으리라. 더구나 조이스는 허구에도 실천적 힘이 있음을 보여주려고 감정을 끌어들인다. 조이스는 소설을 예로

든다. 우리는 소설의 내용을 믿지 않으면서도 진실한 감정을 느낀다. 우리는 이야기와 인물에 깊이 공감한다. 소설을 읽으며 느끼는 감정은 독자를 변화시킨다. 마찬가지로 조이스는 도덕 판단에서 경험하는 감정이 우리 믿음이 참인지 거짓인지에 상관없이 행위를 일으키는데 작용한다고 주장한다. 하나 이렇게 감정의 동기적 측면을 부각하는 방식은 감수성 이론과 큰 차이가 없다. 조이스는 도덕 판단의 감정적 요소보다 인지적 요소를 더 강조했다. 살인은 나쁘다는 판단은 단지 분노를 느끼는 상태가 아니라 정언적 이유가 있다는 믿음을 표현한다고 말이다. 따라서 회의적 결론을 피하려 도덕 감정을 전면으로 내세우는 방식은 자기 부정이다. 진화적 폭로 논증은 예기치 않게, 조이스가 그토록 중요하게 여겼던 도덕의 정언적 힘을 되레 무력하게 만든다.

조이스가 드러낸 실패는 수정된 감수성 이론을 지지하는 통찰을 준다. 도덕과 소설은 다르다. 도덕 판단의 참과 행동은 깊은 관계가 있다. 판단의 참을 믿지 않으면서 도덕 행동을 일으키기는 어렵다. 조이스는 회의주의에서 벗어나려고 감정을 선택했다. 그렇다면 거꾸로 감정과 도덕적 참을 연결하지 못할 이유가 없다. 도덕 사실이 평가적 태도와 독립적인 존재가 아니라면, 도덕적 참이 우리 반응으로 구성된다면, 감정과 도덕적 참을 짝지을 수 있다. 조이스의 비판과 달리 감정적 성향과 도덕 속성의 관계는 잉여가 아니다.

수정된 감수성 이론은 이상적 관찰자를 도입하여 어떻게 도덕 이유가 다른 이유를 압도할 수 있는지 설명하고 아울러 감정이 도덕 판단의 참과 연결됨을 보여준다. 이상적 관찰자의 관점은 일종의 '유령공 상상하기'다. 이상적 관찰자라는 유령 공은 내가 목적해야 할 바를 가리킨다. 그 방향은 내가 승인하는 본질적으로 좋은 가치, 진화한 기초 가치에 닿는다. 그러면 도덕 판단은 필연적으로 내가 의도하는 행동을 이끈다. 요컨대 도덕 명령은 심리적으로 정언적이다. 도덕은 그 자체로 가치 있는 목적이다. 그러나 형이상학적으로는 가언적이다. 감정 없이 행위 이유는 생기지 않는다(Prinz 2007: 136). 감정은 인간 종이 공유하는 보편 가치에 대한 참된 반응이다. 나는 기초 가치에 승인의 감정을 느끼고 도덕적이고자 한다.

4.3. 도덕 내용의 일치

감수성 이론에 대한 조이스의 세 번째 비판은 '내용 문제(content problem)'이다. 프린츠의 감수성 이론은 감정적 성향, 즉 감성이 도덕 속성을 구성한다고 주장한다. 주관적 감정에 바탕을 둔 감수성 이론은 도덕 상대주의를 지지한다. 그렇다면 조이스는 감수성 이론과 상식 도덕관이 나쁘다고 판단하는 사태가 서로 일치하지 않을 수 있다고 말한다. 상식 도덕관은 인종 청소를 반대한다. 하나 감수성 이론은 인종 청소를 찬성할 수 있다. 어떤 개인이나 집단이 인종 청소에 승인의 감정을 느낀다면 말이다. 이상적 관찰자는 이런 반직관적인 결론을 막으려고 도입됐다. 조이스도 이상적 관찰자가 하나의 대안이 될 수 있음을 안다. 그래서 조이스는 다시 묻는다. 완전한 정보를 가진, 실천적으로 합리적인 관찰자를 도입한다 해도 그가 인종 청소에 반대하리라 어떻게 알 수 있는가? 이상적 관찰자를 특별히 규정하지 않는 이상, 모든 이상적 관찰자가 상식 도덕과 동일한 규범 내용을 승인하리라 자신할 수 없다. 문제를 해결하려고 이상적 관찰자가 도덕적 덕을 갖춘 관찰자라고 정의하면 악순환에 빠진다. 애초에 감수성 이론은 도덕 속성의 자연화를 목적으로 삼았다. 그런데 지금은 자연 속성으로 설명되지 않는 도덕 속성이 이미 존재한다고 전제하게 된다(Joyce 2008: 261-262).

조이스의 내용 문제는 이른바 '에우티프론 딜레마'라고 할 수 있다. 플라톤의 대화편 『에우티프론』에서 소크라테스는 에우티프론에게 경건함에 대해 질문한다. 어떤 행위는 신이 사랑하기에 경건한가 아니면 경건하기에 신이 사랑하는가. 똑같은 문제가 수정된 감수성 이론에도 적용된다. 어떤 행동은 이상적 관찰자가 불승인의 감성을 갖기에 그른가 아니면 그르기에 불승인의 감성을 갖는가. 감수성 이론은 관찰자의 반응이 도덕 속성을 만든다고 주장한다. 이때 어떤 행동이 그릇돼서 이상적 관찰자가 불승인한다면 마음 독립적인 도덕 속성을 미리 상정하게 된다. 결국 이상적 관찰자는 도덕적인 관찰자라는 악순환에 빠진다. 반면 어떤 행동이 이상적 관찰자가 불승인해서 그르다면 해당 자연 속성은 자의적일 우려가 있다. 이상적 관찰자마다 불승인하는 가치는 다를 수 있기 때문이다. 따라서 해결해야 할 문제는 이상적 관찰자의 감정 반응과 연결된 자연 속성이 무엇이며 감정의 수렴을 어떻게 보증할 수 있느냐다. 수정된 감수성 이론을 지키려면 이상적 관찰자가 자연 속성으로 정의된다고 주장해야 한다. 물론 이상적 관찰자는 자연 속성으로 정의될 수 있으며 상식 도덕과 일치하는 판단을 내릴 수 있다. 어떻게 가능한가? 진화한 기초 가치 덕분이다.

반복해서 말하는 바, 호모 사피엔스 조상의 후손인 인간 종은 보편적으로 공유하는 기초 가치가 있다. 기초 가치는 인간의 진화 역사를 반영한다. 세계를 판단하는 인간 종의 인지적 능력과, 생존과 번식에 기본적인 필요와 욕구는 동일하다. 이런 조건에서 세계와 인간은 상호작용하여 기초 가치를 생성한다. 기초 가치는 개체의 적합도를 증진하는 데 매우 중요하므로

즉각적인 반응과 연결된다. 현대 심리학은 인간의 뇌가 신속하고 자동적으로 세계에 대한 좋음-나쁨, 접근-회피 반응을 산출하도록 설계되어 있음을 발견했다. 많은 인지 심리학 연구들은 빠르고 노력이 필요 없는 통제된 과정과, 느리고 의식적이며 언어적 사고에 깊이 의존하는 과정이 분리된다고 본다. 인간의 마음은 재빠르게 외부나 내부 환경을 평가하고, 보고 듣는 모든 대상을 좋고 나쁨이라는 선호로 판단한다. 예를 들어 사람들은 단지 자신과 이름이 비슷하다는 이유로 이성에게 호감을 느끼거나 자신의 이름을 닮은 도시로 이주하는 경향이 있다(Haidt & Bjorklund 2008: 186-187). 인간 경험의 이런 감정적인 요소는 매우 큰 선택적 이점을 주었다. 기민한 판단은 예측 불가능한 환경을 익숙하게 유형화하고 중요한 정보를 쉽게 획득하도록 만들기 때문이다. 이렇게 인간 본성과 세계는 영향을 주고받아 기초 가치를 만들며, 시간이 지나면서 가치와 반응의 연결은 더 정교해 진다. 즉 인간을 이루는 생물학적 사실은 도덕 속성을 구성하는 재료이다.

도덕 속성이 형성되는 과정은 다음과 같다. 일단 좋고 나쁨을 판단하는 기본 감정이 환경과 상호작용하여 기초 가치를 생성한다. 기초 가치는 감정을 일으켜 특정 행동을 산출한다. 이런 연결이 반복되면 기초 가치는 더 효과적으로 인간 행동을 규제하려 우리가 도덕이라 이름붙이고 사고하는 영역이 된다. 그리하여 기초 가치는 기본 감정을 도덕 감정으로 전용하고, 도덕 감정은 기초 가치를 도덕 가치로 만든다. 다시 말해, 우리 반응과 도덕 속성은 공진화한다. 감정은 도덕 속성을, 도덕 속성은 감정을 창조하고 강화한다.

우리 조상의 진화사에서 누군가에게 불필요한 고통을 주는 행위는 분노를 야기했다. 분노는 기본적으로 위협에 대한 회피 반응이다. 그래서 조상들은 감정적 동요를 일으킨 죄인을 비난하고 처벌했다. 공동체에서 배제된 죄인은 자신의 행동에 슬픔을 느꼈다. 고통이 한쪽에는 분노를, 다른 한쪽에는 슬픔을 유발한다는 사실이 반복적으로 일어나면 고통을 일으키는 행동은 규칙으로 규제된다. 그예 도덕적 측면이 없던 감정은 새로운 의미를 획득한다. 슬픔은 단순히 잃어버린 대상에 대한 반응이 아니라 규칙을 위반했을 때의 느낌과 연결된다. 분노는 위협에 대한 반응을 넘어 규칙 위반자에 대한 도덕적 분노로 확장된다. 죄책감과 정의로운 분노라는 감정도 탄생한다. 감정의 전환과 함께 규칙도 새로운 의미를 부여받는다. 규칙은 도덕 감정으로 강화되고 비로소 도덕 규칙이 된다(Prinz 2007: 118-119).

고통에 대한 불승인만이 도덕적 그름을 구성하는 자연 속성은 아니다. 아이 양육하기, 협동과 동맹 맺기, 사회적 위계 유지하기, 배우자 지키기 등 조상들이 해결해야 했던 적응 문제는 여러 영역과 관련된다. 상술한 기초 가치와 반응의 연결은 하이트가 말한 여섯 가지 도덕 기반, 즉 피해/배려, 공평성/호혜성, 내집단/충성, 권위/존경, 순결/신성, 자유/압제 내에서 전형적으로 일어났다고 생각한다. 이는 우리가 선천적으로 보유한 도덕 가치가 다원적임을 시사한다. 도덕 기반이 다양한 만큼 도덕적 그름과 동일시되는 자연 속성이 반드시 하나일 필요

는 없다. 진화를 이용하여 도덕적 올바름이 사회적 조화 혹은 공동체의 좋음이라는 단일한 자연 속성으로 설명될 수 있다는 주장은 틀렸다. 단일한 가치는 일상적인 도덕 행동과도, 도덕 심리학의 경험 연구와도 맞지 않는다. 일원론적인 도덕 자연주의는 자연주의의 오류를 범할 수밖에 없다.

수정된 감수성 이론은 자연주의의 오류를 피할 수 있다. 도덕 속성은 기초 가치에 느끼는 감정적 성향으로 예화되는 사실이다. 도덕 속성은 여섯 가지 도덕 기반에서 나타나는 감정 반응에 수반한다. 기초 가치에 대한 감정 반응은 인간 본성이다. 우리 반응은 도덕 기반을 만드는 기초 가치를 정립했고, 기초 가치는 다시 우리 반응을 도덕으로 정교화했다. 상호 의존 과정으로 구성된 도덕 속성은 오직 인간중심적인 의미에서 참이다. 우리 반응을 벗어나는 도덕 판단은 없다. 화성에서 온 외계인은 우리가 올바르다고 생각하는 도덕 판단과 실천에 동의하지 않으리라. 도덕 속성이 반응에 좌우되는 인간중심적인 사실이라면, 기초 가치와 연관된 행동에서 느끼는 승인과 불승인은 참으로 정당화된다. “나는 살인에 대해 불승인의 감정을 느낀다. 따라서 살인을 하지 말아야 한다.”라는 추론은 전제에 ‘~해야 한다’없이 규범적 결론을 내린다. 한데 우리는 이 추론의 참을 즉시 이해한다. 정상적인 인간이라면 이를 타당한 추론이라 생각한다. 살인에 즉각적인 분노를 느끼는 불승인은 올바른 반응이다. 판단을 둘러싼 모든 정보를 알고, 실천적으로 합리적인 사람이라면 살인을 불승인한다. 따라서 살인은 도덕적으로 그르다. 이런 추론이 타당한 이유는 진화한 기초 가치와 감정적 성향이 역사적 필연성으로 연결되었기 때문이다. 그리하여 진화한 기초 가치와 관계된 우리 감정은 도덕적 참을 위한 올바른 길잡이다.

요약하자. 이상적 관찰자는 현실의 관찰자와 마찬가지로 인간 종의 진화 역사를 공유하는 관찰자이다. 이상적 관찰자의 본성은 기초 가치로 정의된다. 인종 청소는 인간 본성이 전형적으로 불승인하는 행동이다. 고로 이상적 관찰자는 인종 청소를 불승인한다. 더불어 현실의 관찰자가 인종 청소를 불승인하는 반응은 이상적 관찰자가 불승인할 수 있는 반응으로 설명된다. 따라서 둘의 반응이 일치하리라는 예측은 자연스럽다.

그렇다면 기본적인 좋고 나쁨에 대한 욕구에서 기원한 도덕 규칙이 합리적 권위를 가질 수 있을까? 물론 그렇다. 기초 가치가 도덕 규칙으로 확장된 일은 우연이 아니다. 도덕은 인간 종의 생존과 번식에 매우 중요했고 공동체의 조화와 협동을 규제하는 힘이다. 우리 감정은 도덕에 무심하지 않다. 과거 조상들은 감정을 느끼는 성향 덕에 삶의 문제를 효과적으로 해결했다. 따라서 감정 반응은 어떤 행위를 추구하거나 금지하는 규범적 역할을 한다. 살인에 대한 경멸과 분노는 살인을 금지하고 위반자를 처벌하는 규칙을 만든다. 우리는 격한 분노로 잔인한 행위를 하지 않으려는 동기를 가진다. 감정적 성향은 그 자체 규범적이다.

수정된 감수성 이론은 객관적인 도덕 가치가 다원적이라 주장한다. 여섯 가지 기초 가치는

감정과 연결되어 도덕 속성의 토대가 되며 규범적 권위를 발휘한다. 여기서 하나의 비판이 제기될 수 있다. 행위자가 따라야 할 가치가 서로 충돌할 경우에 어떻게 해야 하는가? 때로 내집단에 대한 충성은 외집단에 가혹한 피해를 끼치기도 한다. 집단의 안전과 자원 확보를 목적으로 전쟁을 벌이는 행동이 그렇다. 그런데 충성심과 해 금지 모두 우리가 추구해야 할 기초 가치이다. 지금 행위자는 이도 저도 할 수 없는 상황에 처했다. 수정된 감수성 이론의 실천적 힘은 무너지는가?

다원적인 도덕 가치가 서로 모순되는 상황은 충분히 가능하다. 각각의 가치는 소규모 수렵 채집 사회에서 겪는 고유한 적응 문제에서 발생했다. 즉 각 가치는 서로 다른 원인에서 기원했다. 그렇기에 영역이 서로 겹치는 일은 상존한다. 더구나 현대 사회는 더 이상 단순한 해결 방법이 통하지 않는 복잡한 공간이며 삶의 모습도 다양해졌다. 인간은 이제 미래 세대의 행복까지도 고민하는 존재다. 도덕적 책임이 미치는 범위는 날로 확장돼 간다. 이에 수정된 감수성 이론은 어떤 가치, 판단, 행동, 규칙이 도덕적으로 더 나은지 인도할 수 있어야 한다.

먼저 두 가지를 지적하고 시작하자. 첫째, 도덕 가치가 상충한다고 해서 다원성이 곧 도덕 상대주의로 흐르지는 않는다. 도덕 상대주의는 절대적인 도덕 규칙을 부정한다. 반면 다원적 객관주의는 여러 도덕 규칙이 다양한 맥락에서 적절하게 사용되어야 한다고 주장한다. 자주 말했듯이, 단일한 가치는 오히려 장애가 된다. 환경 변화에 유연하게 대처하지 못하는 행위자는 금방 도태되기 쉽다. 즉 객관성과 일원성을 동일시하면 잘못이다. 경우에 따라서 대의를 위해 거짓말하는 행동이 허용될 수도 있다. 그렇다면 제기된 비판의 타개책은 분명하다. 우리는 특정 맥락에서 어떤 규칙을 사용하는 게 좋을지 알려주는 평가 기준을 정해야 한다. 둘째, 수정된 감수성 이론은 도덕적 논쟁과 이성적 추론을 배제하지 않는다. 도덕 판단의 수렴은 나와 상대방 모두 도덕과 무관한 사실에 대해 충분히 알고 이해해야 이루어진다. 그래서 현실의 관찰자는 이상적 관점에 서려 판단과 관련된 정보를 모으고 습득한다. 더불어 맞은편의 사람과 토론하고 논증하여 불일치하는 쟁점을 두고 합의를 이끌어낸다. 이런 과정을 거쳐 갑론을박했던 사태는 기초 가치 목록 중 한 곳으로 들어가 즉각적인 감정 반응이 일어난다. 따라서 도덕 규칙의 평가 기준은 감정 반응 전 단계에서 사용된다.

과학 이론의 선택에서 이론 외적인 평가 기준은 널리 활용된다. 과학사가 토머스 쿤(T. Kuhn)은 과학자들이 오래된 이론을 버리고 새로운 이론을 결정하는 방식이 경험적 입증과 거리가 멀다고 지적했다. 경쟁하는 이론은 모두 나름의 정합성을 갖추어 관찰과 실험으로 어느 하나가 더 낫다고 판단하기 어렵다. 대개 과학자들은 반례에 부딪혀도 이론이 아니라 자신의 무지를 탓한다. 그런데도 어떻게 과학 이론은 진보할 수 있는가? 쿤은 과학자들이 좋은 이론이 무엇인지 가리는 기준을 공유한다고 주장한다. 쿤은 다섯 가지 목록, 정확성, 일관성, 넓은 적용 범위, 단순성, 다산성을 제시했다. 다섯 가지 평가 기준은 똑같은 비율로 적용되는

무조건적 진리는 아니지만 풍부한 성과를 거두게 하는 실용적 가치로 기능한다(Kuhn 1977).

프린츠도 자신의 감수성 이론을 옹호하려고 도덕 외적인 평가 기준을 도입한다. 프린츠의 감수성 이론은 도덕 상대주의에 찬동하기 때문에 더 나은 판단을 분별할 수 없다는 의혹에 빠진다. 나의 도덕적 주장은 당신보다 특별히 더 옳지 않다. 각 집단의 도덕 판단은 모두 참이다. 따라서 도덕적 진보라는 이상은 환상이다. 하나 프린츠는 감수성 이론이 도덕적 진보를 설명할 수 있다고 반박한다. 그 근거로 도덕 가치를 평가하는 다수의 기준을 열거한다.

첫째, 일관성 있는 규칙, 둘째, 정확한 지식에 바탕을 둔 규칙. 셋째, 쉽게 실행할 수 있는 규칙. 넷째, 사회적 안정성을 증진하는 규칙. 다섯째, 구성원의 효용을 늘리는 규칙. 여섯째, 개인의 행복도를 높이는 규칙. 일곱째, 더 간단하지만 일반적인 규칙. 여덟째, 더 많은 사람들에게 적용되는 규칙. 아홉째, 기원에 영향 받지 않는 규칙. 즉 선한 원인에서 발생하지 않았어도 여전히 유용한 규칙. 열째, 생물학적 본성에 잘 맞는 규칙(Prinz 2007: 291-292). 열 가지 도덕 외적인 평가 기준은 더 좋은 가치와 규칙을 선정하고 도덕적 진보를 측정하는 수단이다. 예를 들어 남성 우월과 여성의 종속을 수용하는 부족이 있다고 하자. 이곳의 여성은 낮은 행복도와 복지 수준을 견디며, 가혹한 착취로 사회적 불만을 품는다. 남성도 마냥 좋지는 않다. 소유물로서 여성을 더 많이 차지하려고 벌이는 싸움은 그들의 목숨을 위협한다. 결국 성의 우월을 나누는 규범은 여러 평가 기준에 미달한다. 이 부족의 삶의 질은 성이 평등할 때보다 더 낮다. 그러므로 부족의 남성이 외적인 평가 기준으로 자신들의 판단을 반성할 수 있다면 성차별의 가치가 나쁘다는 사실을 알게 되리라(Prinz 2007: 293). 물론 한 번의 성찰로 모든 게 뒤바뀌진 않는다. 하지만 남성의 인식 변화는 더 나은 가치로 향하는 첫걸음이다.

수정된 감수성 이론도 도덕 외적 평가 기준을 받아들여 가치 충돌의 문제를 해결할 수 있다. 어떤 판단이 내집단에 번영을 주지만 외집단에 재앙을 초래한다면 열 가지 기준에 비추어 해당 사안이 자리한 가치를 철저히 따져야 한다. 조사가 이뤄지면 처음의 예상을 깨는 결과가 나올 수 있다. 즉 집단을 보호하는 행위가 사실은 구성원의 행복과 질서를 파괴하거나, 잘못된 지식에 근거하거나, 많은 사람들이 또한 따를 수 있는 규칙이 아닐 수 있다. 그러면 우리는 문제가 된 판단에 부정적인 감정을 느껴 거부하게 된다. 이후 구성원 모두는 다시 더 나은 도덕 행동을 위해 편견을 교정하고 논쟁 중인 사태가 보편적인 가치 범주로 포섭되도록 노력하리라.

또한 수정된 감수성 이론은 프린츠의 감수성 이론보다 도덕 외적 평가 기준을 활용하기에 더 낫다. 경쟁하는 가치나 규칙을 비교하려면 적어도 둘 사이에 공유하는 개념이나 용어가 있어야 한다. 공통 요소가 있어야 이를 기반으로 특정 가치나 규칙이 평가 기준에 가까운지, 먼지 잴 수 있기 때문이다. 과학 이론도 동일하다. 토머스 쿤은 전혀 다른 세계관, 패러다임

을 가진 두 이론을 견줄 공통된 척도는 없다는 '공약불가능성'을 주장했다. 한데 어떻게 다섯 가지 이론 외적인 평가 기준을 사용할 수 있는가? 쿤은 대립하는 이론에서도 동일한 의미로 쓰이는 용어들이 있다고 대답했다. 즉 의미가 보존되는 용어들이 이론 선택을 위한 토대를 제공한다. 공약불가능성은 국소적으로만 일어난다(Kuhn 1983). 그러므로 인간 종에 보편적인 진화한 기초 가치에 바탕을 둔 수정된 감수성 이론은 도덕 외적인 평가 기준을 적용할 수 있는 공약가능성을 마련한다. 평가 기준으로 보아 상이한 두 도덕 규칙에서 한 쪽이 기초 가치에 조금이라도 더 복무하거나 덜 어긋날 때, 그 규칙을 택해야 한다. 기초 가치는 우리가 그 자체로 목적하는 도덕적 올바름이기 때문이다.

반면 프린츠의 감수성 이론은 무정부상태다. 도덕의 객관성을 부정하면 평가 기준을 써도 더 나은 가치는 알 수 없다. 먼저 무엇을 기반으로 복지, 행복, 안정성, 일관성 등을 측정하는가? 프린츠에 따르면 문화마다 바람직하게 여기는 행동은 다양하다. 다시 말해, 어떤 규칙이 더 가치를 증진하는지, 덜 하는지 헤아릴 수 있는 잣대는 없다. 또한 설사 한 규칙이 평가 기준에 부합해도 다른 규칙보다 더 낫다고 말할 수 없다. 그 규칙은 그저 한 개인이나 공동체에 좋을 뿐이다. 되레 다른 문화권에서는 기준에 미달할 수 있다. 프린츠는 도덕적 진보란 일련의 가치에서 도덕적으로 더 참된 가치로 이행하는 과정이 아니라고 단언한다. 그에게 진보란 오직 도덕 외적인 의미에서 더 나은 가치를 찾는 데 있다(Prinz 2007: 297). 그러나 공유된 기초 없이 평가 기준을 이용할 수 없으며, 더 참된 가치를 꼽을 수 없다면 진보라 부를 수도 없다. '더 나은', '더 참된', '더 좋은'은 하나가 맞고 다른 하나가 틀렸음을 말할 수 있을 때 의미 있다. 프린츠의 감수성 이론은 그럴 수 없다. 프린츠의 주장은 도덕 가치의 진보가 아니라 그냥 도덕 가치의 교체에 불과하다. 도덕적 진보는 도덕 객관주의에 서야 한다. 다시 프린츠는 많은 도덕 규칙은 상대적이지만 문화를 초월하여 인간 종으로서 공유하는 핵심 가치가 비교를 가능케 한다고 반론할 수 있다. 하나 이는 정확히 수정된 감수성 이론의 주장이다. 따라서 수정된 감수성 이론은 감수성 이론을 포괄하면서도 도덕적 진보를 보여줄 수 있는 객관적인 이론이다.

5. 결론

이상으로 조이스가 제기한 진화적 도덕 반실재론과 그에 대응한 프린츠의 감수성 이론을 비교하고, 조이스의 재반박으로부터 감수성 이론을 구제할 수 있는 수정된 감수성 이론을 검토했다. 조이스는 진화가 도덕 판단을 정당화하는 도덕 사실이나 속성의 인식 가능성을 훼손한다고 주장한다. 도덕은 마음 독립적인 세계의 사실과 무관하게 기원했다. 도덕 개념과 믿음은 자연선택의 산물일 뿐이다. 더불어 도덕 판단은 불가피한 권위를 가진다. 즉 도덕 사실은 사적인 욕구를 들먹이며 회피할 수 없는 정언적 이유를 주어야 한다. 그러나 정언적 이유는 없다. 정언성도 자연선택의 소산이다. 따라서 도덕 사실은 없다. 모든 도덕 판단은 거짓이다. 도덕 사실을 자연 사실로 설명하려는 어떤 도덕 자연주의도 도덕의 정언성을 수용하지 못한다. 도덕 자연주의는 기본적으로 행위자의 욕구에 기반을 둔 가언적 체계이기 때문이다. 마음 독립적인 도덕 사실이 있다는 믿음은 진화가 만든 환상이다. 자연선택의 목적은 성공적인 생존과 번식을 이루는 데 있지, 세계의 진상을 탐구하는 데 있지 않다. 이를 '진화적 폭로 논증'이라 한다.

프린츠는 도덕 사실이나 속성이 마음 독립적이라는 전제에 반대한다. 도덕 사실이나 속성은 반응 의존적이다. 어떤 대상은 관찰자를 즐겁게 할 때 우스움의 속성을 가진다. 마찬가지로, 도덕적 옳음과 그름 또한 관찰자에게 특정 반응을 야기하는 속성이다. 프린츠는 도덕 속성을 감정을 일으키는 성향, 즉 감성으로 정의한다. 어떤 행동은 관찰자에게 승인, 불승인의 감정을 야기하는 감성이 있을 때 옳고, 그르다. 이를 '감수성 이론'이라 한다. 감수성 이론은 도덕 판단에는 진리치가 없다는 비인지주의와 다르다. 관찰자의 감정 표현은 자신의 반응이 옳다는 주장이다. 따라서 감정은 옳고 그름을 판단하는 기준으로 작용한다. 또한 감정은 행위 동기를 제공하는 인과적 역할을 한다. 감정은 그 자체 동기적 상태로 행위자를 움직이는 힘이다. 감수성 이론은 감정 반응이 세계의 사태를 표상하며, 도덕 행동을 추동하는 인과적 효과를 낸다는 의미에서 실재론이다. 인간은 선천적으로 감정과 감정을 느끼는 능력을 타고난다. 현대 도덕 심리학의 경험 연구는 도덕 판단에 감정이 결부됨을 지지한다. 하나 프린츠는 감정은 진화적 적응이지만 도덕은 감정과 문화가 결합하여 발생한 후천적 산물이라고 말한다. 도덕 규칙은 문화마다 다양하다. 그래서 프린츠는 도덕 상대주의에 바탕을 둔 주관적 실재론을 주장한다.

조이스는 프린츠의 감수성 이론을 세 가지 점에서 비판한다. 첫째, 불완전성 문제. 감정을 느끼는 관찰자와 특정 상황이 명시되지 않을 경우 도덕 판단은 극단적인 상대주의로 향한다. 둘째, 실천적 유관성 문제. 감정은 도덕 판단의 정언적 성격을 해명할 수 없다. 셋째, 내용 문제. 감수성 이론이 그르다고 판단하는 행동과 상식 도덕관이 그르다고 판단하는 행동이 일

치하지 않을 위험이 있다.

수정된 감수성 이론은 조이스가 제기한 난관을 해결한다. 그 전략은 첫째, 인간에게는 진화 역사에서 생성된 생물학적인 기초 가치가 있다. 배려/피해, 공평성/부정, 충성심/배신, 권위/전복, 고귀함/추함, 자유/압제가 기초 가치에 속한다. 여섯 가지 기초 가치는 도덕의 기반이다. 기초적인 도덕 가치는 조상들이 마주쳤던 적응 문제와 관련된다. 자연선택은 개체가 생존과 번식에 직결되는 역경을 효과적으로 타개하도록 가치와 감정을 연결했다. 따라서 다른 모든 조건이 동등하다면, 도덕 판단은 기초 가치에 대한 감정 반응으로 수렴된다. 이를 위해 도덕 판단을 둘러싼 완전한 정보를 알고, 실천적으로 합리적인 이상적 관찰자를 도입한다. 현실의 관찰자가 내리는 판단은 이상적 관찰자의 판단으로 제한된다. 다시 말해, 현실의 관찰자가 불승인의 감정을 느끼는 행동은 또한 이상적 관찰자가 불승인의 감정을 느낄 수 있을 때 그르다. 이상적 관찰자의 감성을 참조함으로써 도덕 판단의 객관성은 확보된다.

둘째, 도덕 판단이 욕구와 상관없이 따라야 할 정언적 이유에 대한 믿음을 표현한다는 주장은 확고하지 않다. 도덕은 정언적으로 적용 가능하지만 반드시 행위 이유나 동기를 주지 않을 수도 있다. 즉 정언성은 도덕 판단의 핵심이 아니다. 도덕적으로 행동하려는 욕구나 감정 없이 행위 이유는 생기지 않는다. 행위자의 조건을 고려하지 않은 순수 이성의 명령은 쉽게 무시될 수 있다. 그럼에도 여전히 도덕에는 실천적 힘이 있다. 도덕은 인간 삶의 안녕을 증진한다. 도덕은 인간 본성의 일부이다. 한데 현실의 관찰자는 무지와 편견으로 말미암아 자기에게 좋은 행위가 무엇인지 모를 수 있다. 도덕적으로 행동하려는 동기는 이상적 관찰자를 따를 때 발휘된다. 행위자가 이상적 관찰자의 관점에 선다면, 자신의 욕구를 이상적 관찰자의 관점에서 평가한다면, 그는 곧 무엇이 좋은지 이해하고 도덕 욕구에 승인의 감정을 느끼게 된다. 승인의 감정은 그를 행동하도록 만든다. 감정과 동기는 필연적으로 연결되기 때문이다. 이렇게 도덕 명령은 심리적으로 정언적이다. 적절한 감정이 있을 때, 행위자는 도덕 명령을 그 자체 중요한 목적으로서 따른다.

셋째, 이상적 관찰자의 성격은 현실의 관찰자와 똑같이 진화한 기초 가치로 정의된다. 기초 가치는 인간의 진화 역사를 반영한다. 기초 가치는 인간 본성이다. 이상적 관찰자는 기초 가치에 따라 본성에 반하는 인종 청소를 불승인한다. 또한 현실의 관찰자가 내리는 도덕 판단이 이상적 관찰자로 설명되는 만큼, 둘의 반응은 분명 일치한다.

이상적 관찰자는 현실의 관찰자와 동떨어진 선험적 존재가 아니다. 이상적 관찰자는 현실의 관찰자이다. 인간은 자기객관화, 자기반성이 가능한 동물이다. 인간은 경험과 시행착오를 거쳐 자신의 실수와 편견을 교정하고 더 나은 지식을 습득한다. 인간은 자신의 마음속에 완전한 정보를 가진 가상적인 타인의 눈을 소환한다. 한편 감정은 행위자 밖에 환경과 존재자를 표상하는 신체 변화이다. 따라서 인간은 이상적 관찰자의 감정으로 세계를 평가한다. 이상

적 관찰자의 감정적 성향은 도덕 판단의 옳고 그름을 가리는 기준이 된다. 인간은 어떤 행위에 대한 자신의 감정이 적절한지 그렇지 않은지 점검한다. 그리하여 다음번에는 더 올바른 감정을 느낀다. 이런 능력은 자연선택의 산물이다. 사회적 존재인 인간에게 이상적 관점의 내면화는 적합도 증진과 밀접했다. 결론적으로 진화는 도덕 판단의 참을 정당화하는 반응 의존적 사실을 제공하며 감정은 올바른 도덕 판단을 이끈다.

참고문헌

김성한(2001),「도덕의 진화론적 기원과 다윈주의 윤리설」, 고려대학교 대학원 박사학위 논문.

윤화영(2009),「감수성 이론에서의 도덕적 진리」, 『철학적 분석』, 한국분석철학회, 19권, 65-88.

주동률(1996),「수반과 윤리적 실재론」,『철학』, 한국철학회, 48권, 303-335.

Ayer, A. J., (1936),『언어, 논리, 진리』, 송하석 역, 나남. 2010.

Brink, D. O., (1989), *Moral Realism and the Foundations of Ethics*, Cambridge University Press.

Campbell, R., (1996), "Can biology make ethics objective?", *Biology and Philosophy*, 11(1), 21-31.

Crook, J. & Crook, S. J., (1988), "Tibetan polyandry: problem of adaptation and fitness", in L. Betzig, M. Borgerhoff Mulder, & P. Turke, (eds.), *Human Reproductive Behavior: A Darwinian Perspective*, Cambridge University Press, 97-114.

Cummins, D. D., (1996), "Evidence for the innateness of deontic reasoning", *Mind & Language*, 11(2), 160-190.

Cuneo, T., (2011), "Moral naturalism and categorical reasons", in N. Susana, & S. Gary, (eds.), *Ethical Naturalism: Current Debates*, Cambridge University Press, 110-130.

Darwin, C., (1998),『인간의 유래』, 김관선 역, 한길사, 2006.

Dawkins, R., (1976), *The Selfish Gene,* Oxford University Press.

Firth, R., (1952), "Ethical absolutism and the ideal observer", *Philosophy and Phenomenological Research*, 12(3), 317-345.

Foot, P., (1972), "Morality as a system of hypothetical imperatives", *The Philosophical Review*, 81(3), 305-316.

Geach, P., (1965), "Assertion", *Philosophical Review*, 74, in P. Geach (rep.), *Logic Matters*, University of California Press, 1972. 254-269.

Haidt, J., (2012),『바른 마음』, 왕수민 역, 웅진지식하우스, 2014.

Haidt, J., & Bjorklund, F., (2008), "Social intuitionists answer six questions about morality", in W. Sinnott-Armstrong, (ed.), *Moral Psychology: The*

Cognitive Science of Morality: Intuition and Diversity, Vol. 2, MIT press, 181-218.

Harman, G., (1986), "Moral explanations of natural facts: Can moral claims be tested against moral reality?", *The Southern Journal of Philosophy*, 24(S1), 57-68.

Henrich, J., Boyd, R., & Bowles, S., et al., (2005), "'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies", *Behavioral and brain sciences*, 28(06), 795-815.

Hume, D., (1980), 『도덕에 관하여』, 이준호 역, 서광사, 1998.

James, S. M., (2011), *An Introduction to Evolutionary Ethics*, John Wiley & Sons.

Johnston, M., (2010), *Surviving Death*, Princeton University Press.

Joyce, R., (2000), "Darwinian ethics and error", *Biology and Philosophy*, 15(5), 713-732.

Joyce, R., (2001), *The Myth of Morality*, Cambridge University Press.

Joyce, R., (2006), *The Evolution of Morality*, Mit Press.

Joyce, R., (2008), "Replies", *Philosophy and Phenomenological Research*, 77(1), 245-267.

Joyce, R., (2009), "Review: Jesse J. Prinz: The emotional construction of morals", *Mind*, 118(470), 508-518.

Joyce, R., (2013), "The many moral nativisms", in K. Sterelny, R. Joyce, B. Calcott, & B. Fraser, (eds.), *Cooperation and its evolution*, MIT Press, 549-572.

Kirchin, S., (2012), *Metaethics*, Palgrave Macmillan.

Kuhn, T. S., (1977), "Objectivity, value judgement, and theory choice", in T. S. Kuhn, (ed.), *The Essential Tension*, University of Chicago Press, 320-339.

Kuhn, T. S., (1982), "Commensurability, comparability, communicability", *PSA* 1982, Vol 2, Philosophy of Science Association, 669-688.

Levenson, R. W., Ekman, P., & Friesen, W. V., (1990), "Voluntary facial action generates emotion-specific autonomic nervous system activity", *Psychophysiology*, 27(4), 363-384.

Mackie, J. L., (1977),『윤리학: 옳고 그름의 탐구』, 진교훈 역, 서광사, 1990.

McDowell, J., (1985), "Value and secondary qualities", in T. Honderich, (ed.), *Morality and Objectivity*, Routledge and Kegan Paul, 110-129.

McGeer, V., (2008), "Varieties of moral agency: lesson from autism(and psychopathy)", in W. Sinnott-Armstrong, (ed.), *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Vol 3, Mit Press, 227-258.

Mikhail, J., (2008), "The poverty of the moral stimulus", in W. Sinnott-Armstrong, (ed.), *Moral Psychology: The Evolution of Morality*, vol. 1, MIT press, 353-360.

Moll, J., de Oliveira-Souza, R., & Eslinger, P. J., (2003), "Morals and the human brain: a working model", *Neuroreport*, 14(3), 299-305.

Moore, G. E., (1903), *Principia Ethica*, Cambridge University Press.

Pigden, C., (1993),「자연주의」,『메타윤리』, 김성한 외 역, 철학과 현실사, 2006.

Powell, K. L., Roberts, G., & Nettle, D., (2012), "Eye images increase charitable donations: evidence from an opportunistic field experiment in a supermarket", *Ethology*, 118(11), 1096-1101.

Prinz, J. J., (2007), *The Emotional Construction of Morals*, Oxford University Press.

Prinz, J. J., (2008a), "Acquired moral truths", *Philosophy and Phenomenological Research*, 77(1), 219-227.

Prinz, J. J., (2008b), "Is morality innate?", in W. Sinnott-Armstrong, (ed.), *Moral Psychology: The Evolution of Morality*, vol. 1, MIT press, 367-406.

Prior, A. N. (1976), "Autonomy of ethics", in P. T. Geach, & J. P. Kenny, (eds.), *Papers in Logic and Ethics*, University of Massachusetts Press, 88-96.

Railton, P., (1986), "Moral realism", *The Philosophical Review*, 95(2), 163-207.

Richards, R. J., (1986), "A defense of evolutionary ethics", *Biology and Philosophy*, 1(3), 265-293.

Rottschaefer, W. A., & Martinsen, D., (1990), "Really taking Darwin seriously: An alternative to Michael Ruse's Darwinian metaethics", *Biology and Philosophy*, 5(2), 149-173.

Ruse, M., (1986), *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*, Prometheus Books.

Ruse, M., & Wilson, E. O., (1986), "Moral philosophy as applied science", *Philosophy*, 61(236), 173-192.

Sayre-McCord, G., (2006), "Moral realism", in D. Copp, (ed.), *The Oxford Handbook*

of Ethical Theory, Oxford University Press, 39-62.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H., (2008), "Disgust as embodied moral judgment", *Personality and social psychology bulletin,* 34(8), 1096-1109.

Smetana, J. G., (1981), "Preschool children's conceptions of moral and social rules" *Child development*, 52(4), 1333-1336.

Smetana, J. G., & Braeges, J. L., (1990), "The development of toddlers' moral and conventional judgments", *Merrill-Palmer Quarterly*, 36(3), 329-346.

Smith, M., (1993),「실재론」,『메타윤리』, 김성한 외 역, 철학과 현실사, 2006.

Smith, M., (1994), *The Moral Problem*, Basil Blackwell.

Sober, E., & Wilson, D. S., (1998), *Unto Others: The Evolution of Altruism*, Harvard University Press.

Spencer, H., (1879), *The Principles of Ethics*, University Press of the Pacific, 2004.

Stevenson, C., (1937), "The emotive meaning of ethical terms," *Mind*, 46(181), 14-31.

Tangney, J. P., Stuewig, J., Malouf, E. T., & Youman, K., (2013), "Communicative function of shame and guilt", in K. Sterelny, R. Joyce, B. Calcott, & B. Fraser, (eds.), *Cooperation and its evolution*, MIT Press, 485-502.

Trivers, R. L., (1971), "The evolution of reciprocal altruism", *Quarterly review of biology*, 46(1), 35-57.

Wheatley, T., & Haidt, J., (2005), "Hypnotic disgust makes moral judgments more severe", *Psychological science*, 16(10), 780-784.

Wiggins, D., (1987), "A sensible subjectivism", in *Needs, Value, and Truth*, Oxford University Press, 185-214.

Abstract

A Critique of the Evolutionary Moral Anti-Realism

O-Hyun Kwon

Program in History and Philosophy of Science

The Graduate School

Seoul National University

In this thesis, I will claim two things. First, moral facts that justify moral judgements are not mind-independent but response-dependent. Second, evolution provides response-dependent facts to vindicate moral judgements.

I will compare two authors for discussion: Richard Joyce and Jesse J. Prinz. Joyce argues that evolution undermines the epistemic possibility of mind-independent moral facts. This is called 'evolutionary debunking arguments'. On the other hand, Prinz argues that emotions constitute response-dependent moral facts. This is called 'sensibility theory'. I tried to analyze the debates taken place between Joyce and Prinz, and critically complement the arguments of Prinz to defend moral realism, thus I will propose the 'modified sensibility theory'. Key arguments are as follows. (1) Human have evolved basic values causing universal moral emotions. (2) Introduction of the ideal observer can show convergence of emotional responses to basic values.

Joyce criticized sensibility theory in three aspects. First, emotional responses about certain actions may vary for each individual or group. For instance, Hitler and I would have different emotions about ethnic cleansing, that is sensibility theory im-

plies extreme relativism. Second, moral judgements express the belief about categorical reason, However sensibility theory to assert that If there are no emotions then there are no reason for action cannot explain the categoricity of moral reason. Third, there is a possibility that disagreement between the acts that sensibility theory classify as wrong and the acts that moral common sense classifies as wrong. How can we be sure that sensibility theory would be disapproval of ethnic cleansing?

Modified sensibility theory can save the objectivity of moral judgements. Significant values to survival and reproduction in human evolutionary history was connected with particular emotional responses. For example, someone feels disgust for people with sexual promiscuity. These kinds of values are the foundation of mechanism we call morality. The values can be categorized that causing the feelings of approval and disapproval by the evolution. Human species share basic values. All other things being equal, moral judgments are converged to universal responses to the basic values. To this end, we assume ideal observer with fully informed non-moral fact and practically rational. Actual agents refer to responses of the ideal observer. Actual agent's good is determined by what her fully informed and practically rational agent would feel her to feel in her circumstance. Ideal observer is a good guide for the correct moral judgments. Thus, the reactions of actual agents are objective moral judgments, that we all agree to. The existence of the ideal observer is confirmed empirically: a Self-objectification, and presence of a meta-emotion.

Modified sensibility theory can embrace the inescapable practical clout of morals. Morality was extremely important to our survival and reproduction of their ancestors. Desire to act morally is part of human nature. But the actual agents may not know what good desire is, due to her ignorance. When she could stand on the viewpoint of the ideal observer, she can understand what is good and urge her to feel the emotions accordingly. If we approve of what is good, then we pursue it. Because emotions and motivations are necessarily connected.

Modified sensibility theory can explain the agreement of the moral content. The ideal observer is defined as evolved basic values and the emotional responses to it, in the same way as the actual agents. If so, Actual agents are described in the context of the ideal observer, it is natural to expect that the two would match.

The problems of Prinz's sensibility theory can be solved through the ideal

observer. Ideal observer is well suited to our intuition. People often think that their point of view is very close to the ideal point of view. The ability to assume a hypothetical observer appears as a product of natural selection. Various empirical studies to support this. In conclusion, moral facts are response-dependent facts and evolution justifies moral judgments.

keywords : evolution, moral anti-realism, evolutionary moral anti-realism, evolutionary debunking arguments, sensibility theory, Richard Joyce, Jesse J. Prinz
Student Number : 2013-20201