

월리스의 양립론과 도덕적 책임의 조건^{*†}

이 병 덕

월리스는 결정론과 도덕적 책임 사이의 양립론을 주장한다. 그에 따르면 도덕적 책임은 일차적으로 반성적 자기 통제력의 문제이고, 이러한 이성적 능력은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유를 선제하지 않는다. 또한 우리에게 이러한 이성적 능력이 있다는 것과 결정론이 양립하지 않음을 비양립론자들이 지금껏 성공적으로 보여주지 못했다고 주장한다. 월리스의 양립론은 왜 결정론이 사람들에게 도덕적 책임을 묻는 사회실천의 정당성에 심각한 위협이 되지 않는지를 잘 설명해 준다. 그러나 월리스가 자신의 양립론을 옹호하기 위해 호소하는 논거들 중 정당화하기 어려운 것들이 있다. 이 논문의 목적은 이러한 문제를 피할 수 있는 새로운 양립론을 제시하고 옹호하는 데 있다. 특히 도덕적 책임의 기본조건과 면제조건을 통해 양립론을 보다 성공적으로 옹호할 수 있음을 주장한다.

주 제 윤리학

주요어 결정론, 자유, 양립론, 도덕적 책임, 월리스, 칸트

* 투고일: 2022. 05. 15, 심사 및 수정 완료일: 2022. 06. 12, 게재 확정일: 2022. 06. 13

† 이 논문은 성균관대학교의 2021학년도 삼성학술연구비에 의하여 연구되었음.

1. 들어가는 말

결정론(determinism)에 따르면, 한 주어진 시점에서 세계의 모든 사실들이 주어지면, 이 사실들과 자연법칙들에 의해 그 이후에 발생할 모든 사건들이 결정된다. 이러한 결정론이 옳다면, 사람들에게 도덕적 책임을 묻는 우리의 사회실천은 그 정당성을 잃게 되는가?

전통적인 견해에 따르면 도덕적 책임은 의지의 자유를 선제(presuppose)하고, 또한 의지의 자유는 달리 할 수 있는 능력에 의존한다. 월리스(Jay Wallace)는 1994년에 출판된 그의 유명한 책 『책임과 도덕 감정』(Responsibility and the Moral Sentiments)에서 도덕적 책임에 관한 양립론(compatibilism)을 옹호한다. 그에 따르면 도덕적 책임은 의지의 자유가 아니라, 반성적 자기통제(reflective self-control) 능력에 의존한다. 후자는 도덕적 이유들을 파악하고 적용하는 능력, 그리고 그러한 이유들에 따라 자신의 행동을 통제할 수 있는 능력이다. 다시 말해 월리스는 도덕적 책임이 달리 할 수 있는 능력의 문제가 아니라, 반성적 자기통제 능력의 문제라고 주장한다. 또한 우리에게 이러한 이성적 능력이 있다는 것과 결정론이 양립하지 않음을 비양립론자들이 지금까지 성공적으로 보여주지 못했다고 주장한다.

월리스의 양립론은 왜 결정론이 사람들에게 도덕적 책임을 묻는 우리의 사회실천에 심각한 위협이 되지 않는지를 잘 설명해 준다. 그러나 아쉽게도 월리스가 자신의 양립론을 옹호하기 위해 의존하는 논거들 중에는 정당화하기 어려운 것들이 있다. 이 논문의 목적은 이러한 문제를 피할 수 있는 새로운 양립론을 제시하고 옹호하는 데 있다. 특히 도덕적 책임의 기본조건과 면제조건을 통해 양립론을 보다 성공적으로 옹호할 수 있음을 주장한다.

2. 도덕적 책임에 관한 반응적 설명과 의지의 자유

우선 도덕적 책임은 윤리학의 가장 핵심적인 개념이다. 그렇다면 이 핵심 개념을 이해하기 위해서 어디로부터 출발해야 하는가? 표준적 설명에 따르

면 **책임이 있음**(being responsible)은 **책임을 물음**(holding responsible)에 분석적으로 선행한다. 다시 말해 어떤 행위주체에게 책임을 묻는 것이 공정하기 위해서는 그에게 실제로 책임이 있어야 한다. 그러나 윌리스는 설명의 순서가 반대여야 한다고 주장한다. 즉 책임이 있음의 개념을 책임을 물음의 개념을 통해 이해해야 한다고 주장한다.

그렇다면 책임을 물음의 개념을 어떻게 이해해야 하는가? 윌리스에 따르면 이러한 근본 개념을 설명하기 위해 우리가 출발점으로 삼아야 하는 것은 그 어느 누구도 부인하기 어려운, 우리에게 설명해야 할 데이터로 주어지는 현상이다. 또한 도덕적 분노, 분개, 죄책감 같은 우리의 반응적 태도들(reactive attitudes)은 우리가 실제로 관찰할 수 있는 현상이다. 그리고 스트로슨(Strawson 1962)에 따르면 어떤 사람에게 책임을 묻는 조건은 그러한 반응적 태도로 설명돼야 한다. 이런 이유에서 윌리스는 **도덕적 책임에 관한 반응적 설명**(the reactive account of holding people responsible)을 주장한다. 그리고 그러한 설명에 따라 책임을 물음의 개념을 이해하는 데 우리의 반응적 태도가 핵심적이라고 주장한다. 다시 말해 누군가 우리의 도덕적 기대를 어겼다고 생각할 때 우리가 보이는 반응적 태도와 사람들에게 도덕적 책임을 묻는 태도 사이에 밀접한 관계가 있다고 주장한다.

그런데 사람들에게 도덕적 책임을 묻는 우리의 사회실천과 관련하여 스트로슨과 윌리스 사이에 중요한 차이점이 있다. 스트로슨은 책임을 묻는 우리의 실천을 **반응적 감정**(reactive emotions)의 차원에서 설명한다. 반면 윌리스는 이에 대해 **규범적 접근**(a normative approach)을 해야 한다고 주장한다.

내 제안에 따르면, 우리는 도덕적 책임에 관한 논쟁을 규범적 용어로 해석한다. 일차적인 쟁점은 사람들에게 책임을 묻는 우리의 실천이 그러한 실천 전체에 선행하여 또한 독립적으로 존재하는, 책임에 관한 사실들에 대응하는지 여부가 아니다. 그것은 오히려 다음 질문에 달려 있다. 어떤 조건 하에서 사람들에게 책임을 묻는 태도를 택하는 것이 도덕적으로 공정한가? ... 이 규범적 해석은 두 가지 중요한 귀결을 가진다. 첫째, 우리가 어떤 사람에게 책임을 묻는 태도— 판단

되는 행위자라기보다는 도덕적 평가자의 태도—를 먼저 이해하지 않으면 도덕적으로 책임 있는 행위자임이 무엇인지 확립할 수 없다. 둘째, 책임의 조건들이 무엇인지 결정하기 위해서 규범적 도덕 이론에 대해 논할 수밖에 없다. 공정함에 관한 우리의 원리들이 사람들에게 책임을 묻는 것을 공정하게 하는 조건들에 관해 함축하는 바를 보기 위해, 우리는 그러한 원리들에 대해 탐구할 필요가 있다.¹⁾

사람들이 보이는 반응적 태도들이 모두 공정(fair)한 것은 아니다. 따라서 우리는 사람들의 반응적 태도들에 대해 그것들이 공정한지, 아닌지를 평가해야 한다. 그리고 그러한 평가를 위해 불가피하게 공정함에 관한 도덕원리들에 호소할 수밖에 없다. 그리고 그러한 도덕원리들은 양립론과 비양립론 논쟁에서 중립적이어야 한다. 더 나아가, 양자 사이의 논쟁을 해결하기 위해서 궁극적인 도덕원리들에 관한 논쟁에 가급적 개입하지 않는 것이 바람직하다. 궁극적인 도덕원리들을 결정한 이후에야 양립론과 비양립론 논쟁을 해결할 수 있다면, 언제 궁극적인 도덕원리들에 대해 합의할 수 있을지 요원한 현 상태에서, 양립론과 비양립론 중 어느 쪽이 더 설득력이 있는지에 관해 유의미한 철학적 진전을 이루기 어렵기 때문이다. 이런 이유에서 궁극적인 도덕원리들에 호소하기 어렵다면, 공정함에 관한 도덕원리들을 어떻게 결정할 수 있는가?

위 물음에 답하기 위해 윌리스는 **반성적 평형(reflective equilibrium)**의 방법을 제안한다. 이 방법에 따르면 공정함에 관한 도덕원리들을 “유능한 도덕적 평가자들이 높은 단계의 자신감을 갖는 숙고된 확신들”을 토대로 결정할 수 있다.²⁾ 그리고 이와 같은 숙고된 확신에 따르면, 어린 아이들이나 단순한 동물들의 경우처럼 반성적 자기 통제력을 지니지 못한 존재에게 책임을 묻는 것은 공정하지 않다. 다시 말해 오직 특정 수준 이상의 반성적 자기 통제력을 지니는 존재들에게만 책임을 묻는 것이 공정하다. 이러한 논의를 통해 윌리스는 도덕적 책임이 일차적으로 반성적 자기 통제력의 문제

1) Wallace 1994, pp. 3-4.

2) Wallace 1994, p. 113.

라고 주장한다.

그런데 전통적으로 도덕적 책임과 자유를 연관시킨 이유는, 선택의 자유가 없는 사람에게 책임을 묻는 것은 불공정하다고 생각했기 때문이다. 월리스는 이런 의미의 자유를 ‘의지의 강한 자유’(strong freedom of the will)라고 부른다.³⁾ 그리고 이런 의미의 자유는 자연법칙들과 과거 사실들이 고정된 상황에서 우리가 선택할 수 있는 진정한 대안의 존재를 요구한다. 그러나 월리스는 도덕적 책임의 조건이 이런 의미의 강한 자유에 의존하지 않는다고 주장한다. 앞서 언급한 바대로, 도덕적 책임은 일차적으로 반성적 자기 통제력의 문제이고, 또한 이러한 반성적 자기 통제력은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유를 선제하지 않는다. 예컨대 어떤 사람에게 도덕적 책임을 묻는 것이 공정하지 않다면, 그 이유는 그가 달리 할 수 있는 능력을 결여해서라기보다는, 그가 특정 수준 이상의 반성적 자기 통제력을 결여하기 때문이다. 요컨대 월리스에 따르면 도덕적 책임의 조건은 의지의 강한 자유가 아니라 반성적 자기 통제력이라는 이성적 능력의 차원에서 이해돼야 한다.

3. 월리스의 양립론

양립론과 비양립론 사이의 근본적인 논쟁은 다음과 같다. 우리는 사람들에게 도덕적 책임을 묻는 사회실천(social practice)을 한다. 그렇다면 결정론이 참일 경우에도 그러한 실천이 정당화될 수 있는가? 비양립론에 따르면 결정론이 참일 경우에 도덕적 책임을 묻는 사회실천은 정당화되지 않는다. 왜냐하면 결정론과 도덕적 책임은 양립하지 않기 때문이다. 그렇다면 그 논거는 무엇인가?

비양립론을 옹호하는 가장 유망한 전략은 월리스가 ‘일반화 전략’(generalization strategy)이라고 부르는 것이다.⁴⁾ 앞서 언급한 바대로, 전

3) Wallace 1994, p. 3.

4) Wallace 1994, p. 115를 보기 바람.

통적으로 도덕적 책임과 자유를 연관시킨 이유는, 선택의 자유가 없는 사람에게 책임을 묻는 것이 불공정하다는 생각 때문이었다. 비양립론자들이 일반화 전략을 통해 보여주고자 하는 것은, 결정론이 참이면 책임을 공정하게 묻기 위해 요구되는 조건이 어린아이나 사이코패스의 경우처럼 특수한 경우들뿐만 아니라 우리 모두에게 성립하지 않는다는 것이다.

그러나 윌리스는 위와 같은 일반화 전략은 성공적이지 않다고 주장한다. 먼저 다음의 원리를 고려해보자.

- (1) 한 행위주체가 달리 할 수 있는 능력을 결여한다면, 그의 행위에 책임을 묻는 것은 공정하지 않다.

많은 비양립론자들은 (1)을 받아들인다. 그러나 프랭크퍼트(Frankfurt 1969)는 (1)에 대해 다음과 같은 반례를 제시한다.

존즈는 스미스에게 의도적으로 총을 쏜다. 그런데 존즈는 모르지만, 블랙은 다음과 같은 만반의 조치를 해 놓았다. 만약 존즈가 스미스를 쏘지 않는 경우에는 존즈의 심적 상태를 조작함으로써 결국은 존즈가 스미스를 쏘게 만든다.

위 경우에서 존즈는 블랙의 존재로 인해 스미스를 쏘는 것을 피할 수 없었다. 즉 달리 할 수 있는 진정한 대안이 없었다. 그러나 직관적으로 존즈는 스미스를 쏘는 것에 대해 여전히 책임이 있다. 왜냐하면 존즈는 그 어떤 간섭도 없이 스미스에게 의도적으로 총을 쏘았기 때문이다. 이러한 반례가 보여주는 것은 도덕적 책임이 달리 할 수 있는 능력에 의존하지 않는다는 사실이다.

그렇다면 도덕적 책임은 무엇에 의존하는가? 앞서 언급한 바대로 윌리스에 따르면, 어떤 주체에게 도덕적 책임을 묻는 것이 공정하기 위해서는 책임조건(accountability condition)이 성립해야 한다. 그리고 그 주체가 특정 수준 이상의 반성적 자기 통제력을 소유하는 경우에만 이 책임조건이 성립

한다. 반면 이 책임조건이 성립하지 않는 경우에 그 주체는 도덕적 책임에서 면제(exempt)된다. 예컨대 특정 수준 이상의 반성적 자기 통제력을 결여한 어린 아이들이나 단순한 동물들에게는 책임조건이 성립하지 않는다. 다시 말해 면제조건이 성립한다. 윌리스는 다음과 같이 말한다.

스트로슨에 따라, 우리는 면제조건들의 두 하위 집단을 구분할 수 있다. 첫째, 최면, 극심한 스트레스 또는 신체적 고통, 그리고 특정한 약물에 의한 단기적 영향들의 경우처럼 한 사람에게 책임을 묻는 것이 그 사람의 삶의 한 제한된 기간 동안에만 공정하지 않게 만드는 조건들이 있다. 둘째, 한 사람의 정상적 조건이 그에게 도덕적 책임을 묻는 것이 공정하지 않도록 만들어 주는 좀 더 체계적이고 지속적인 상태들이 있다. 여기에는 예컨대 정신이상, 정신병, 매우 어릴, 사이코패시, 그리고 체계적 행동 통제 또는 조건화의 영향들이 포함된다.⁵⁾

우선 어린 아이들은 명백히 반성적 자기 통제력을 충분히 발전시킨 것으로 여겨지지 않는다. 인지적으로, 그들은 도덕원리들과 그 원리들이 포함하는 (해악 또는 상호이익과 같은) 많은 경우에 매우 복잡한 개념들을 적용하는 방법을 배우는 단계에 있다. 정서적으로, 그들은 도덕원리들에 따라 신빙성 있게 자신의 행동을 통제하는 능력을 아직은 획득하지 못했다. ... 아이들은 [반성적 자기 통제력]을 결여하거나 또는 여전히 그러한 능력을 획득하는 과정에 있기 때문에, 우리가 받아들이는 도덕적 의무들에 관해서 전적으로 책임을 묻는 것은 합당하지 않다.⁶⁾

그런데 어떤 주체에게 책임조건이 성립한다고 해서 그가 한 행동을 비난하는 것이 항상 공정한 것은 아니다. 왜냐하면 그에게 정당한 변명거리(valid excuse)가 있을 수 있기 때문이다. 이처럼 정당한 변명거리가 있는 경우에는 비난받지 않을 수 있다. 따라서 그가 한 행동을 비난하는 것이 공정하기 위해서는 비난조건(blameworthiness condition)이 성립해야 한다. 다시 말해 변명조건(excusing condition)이 성립하지 않아야 한다. 요컨대 윌리스는 도덕적 책임과 관련하여 책임조건과 비난조건을 구분한다. 이 구분에 따르면

5) Wallace 1994, p. 155.

6) Wallace 1994, pp. 166-167.

어떤 행위자에게 책임을 묻는 것이 공정하기 위해서는 우선 책임조건이 성립해야 한다. 즉 그가 특정 수준 이상의 반성적 자기 통제력을 소유해야 한다. 또한 그가 한 행동을 비난하는 것이 공정하기 위해서는 비난조건이 성립해야 한다. 즉 그의 행동에 대해 정당한 변명거리가 없어야 한다.

그렇다면 결정론이 참인 경우 면제조건이나 변명조건이 성립하는가? 먼저 변명조건이 성립하는지 살펴보자. 우리 각자는 우리가 한 선택에 대해 책임을 져야 한다. 그런데 어떤 사람의 행동이 그가 의도한 것이 아니었다면, 그는 자신의 의도가 아니었다고 변명할 수 있다. 그렇다면 결정론은 이와 같은 변명거리가 될 수 있는가? 다시 말해 결정론이 참이면, 우리의 모든 행동은 우리의 선택에 의한 것이 아니라고 말할 수 있는가?

윌리스에 따르면 우리는 의도적 행위를 할 수 있고, 또한 우리 자신 및 타인의 많은 행위들을 우리 자신 및 타인이 무엇을 선택했는지의 차원에서 설명할 수 있다.⁷⁾ 그리고 비록 결정론이 참인지 여부에 대해 잘 몰라도 여전히 이러한 사실을 받아들일 수 있다고 주장한다.

책임과 자유와 같이 복잡한 쟁점들을 다루기 위해서는 당연한 것으로 여겨져야 할 것들이 있다. 이 책의 목적상 단순히 가정하고 시작하는 것들 중에는 믿음, 욕망, 감정과 같은 명제적 태도들의 존재와 설명적 유효성이 있다.⁸⁾

이러한 이유에서 윌리스는 결정론이 우리의 선택을 배제하지 못한다고 주장한다.

이제 책임의 면제조건이 성립하는지 살펴보자. 앞서 언급한 바대로, 어떤 주체에게 도덕적 책임을 묻기 위해서는 그가 특정 수준 이상의 반성적 자기 통제력을 소유해야 한다. 그리고 표준적인 책임의 면제조건들은 이러한 책임조건이 훼손되거나 결여된 경우에 성립한다. 그렇다면 결정론은 책임조건이 훼손하거나 결여되도록 만드는가? 다시 말해 결정론이 참이면 그 어떤 행위자도 반성적 자기 통제력을 소유하지 못하게 되는가? 윌리스에 따르면

7) Wallace 1994, p. 149를 보기 바람.

8) Wallace 1994, p. 6.

그러한 귀결은 성립하지 않는다. 앞서 언급한 바대로 결정론에 따르면, 한 주어진 시점에 세계의 모든 사실들이 주어지면, 이 사실들과 자연법칙에 의해 그 이후에 발생할 모든 사건들이 결정된다. 그런데 이 주장이 옳은지 여부와 한 행위자가 이성적 능력을 소유하는지 여부 사이에 직접적인 관련이 없다. 후자는 언어를 구사하는 능력과 같은 심적인 능력이다. 결정론이 그와 같은 심적 능력을 박탈시킨다는 것은 매우 이상한 가정이다. 이와 관련해 주목할 점은 비양립론자들도 자신들이 이성적 능력을 갖고 있음을 부정하지 않는다는 사실이다. 그들은 결정론과 이성적 능력이 양립할 수 없다고 주장하는 대신에, 결정론과 달리 할 수 있는 능력이 양립하지 않는다고 주장한다.⁹⁾ 따라서 결정론이 우리의 이성적 능력을 박탈한다고 볼 필요가 없다. 최면, 정신이상, 세뇌 등에 의해 한 행위자의 반성적 자기 통제 능력이 현저하게 훼손된 경우에는 책임의 면제조건이 성립할 수 있다. 그러나 결정론과 반성적 자기 통제 능력이 양립하는 한, 결정론은 우리 모두를 도덕적 책임에서 벗어날 수 있게 해주는 면제조건이 아니다. 요컨대 비양립론의 일반화 전략이 성공적이지 않다고 윌리스는 주장한다.

필자는 윌리스의 다음 네 주장에 전적으로 동의한다.

첫째, 도덕적 책임은 일차적으로 이성적 능력의 문제이다.

둘째, 도덕적 책임은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유에 의존하지 않는다.

셋째, 비양립론을 위한 일반화 전략은 성공적이지 않다.

넷째, 우리에게 이성적 능력이 있다는 우리의 숙고된 확신과 결정론이 양립하지 않음을 비양립론자들이 지금까지 성공적으로 보여주지 못했다.

그렇지만 필자가 여전히 윌리스에 동의하지 않는 중요한 논점들이 있다. 다음 절에서 필자는 자유에 관한 칸트적 견해(a Kantian view of freedom)를 간략히 소개할 것이다. 그리고 이를 기반으로 5절에서 윌리스의 견해에 동

9) Wallace 1994, pp. 182-183을 보기 바람.

의하지 않는 네 가지 논점들에 대해 논의할 것이다.

4. 자유에 관한 칸트적 설명

우선 칸트에 따르면 우리는 자율적 존재(autonomous being)이다. 그리고 우리가 단순한 동물들과 달리 자율적인 존재인 이유는 도덕법칙에 따라 행위할 수 있는 이성적 능력을 가지고 있기 때문이다. 그런데 칸트에 따르면 자율성(autonomy)과 자유(freedom)는 동등한 개념(reciprocal concepts)이다.¹⁰⁾ 따라서 우리는 자유의 개념을 자율성의 개념을 통해 이해할 수 있다. 다시 말해 자유를 도덕규범에 따라 행위할 수 있는 이성적 능력을 통해 이해할 수 있다. 이것이 칸트가 말하는 ‘긍정적 의미의 자유’(freedom in the positive sense)이다. 즉 X가 긍정적 의미에서 자유로운 존재라고 말하는 것은 X가, 단지 원인들에 의해 제약되는 단순한 동물에 불과한 것이 아니라, 도덕규범에 따라 행위할 수 있는 이성적 존재라고 말하는 것이다. 이와 관련하여 브랜덤은 다음과 같이 말한다.

인간의 자유의 문제에 관해 가장 시사적인 반응들의 하나는 칸트적 전통, 즉 자유는 존재하는 것(what is)뿐만 아니라, 존재해야 하는 것(what ought to be)에도 부응하는 것이기 때문에 단지 원인들에 의해서라기보다는 규범들에 의해 제약됨에 의해 정확히 성립한다는 견해에 의해 발전되어 왔다.¹¹⁾

칸트의 규범적 전회의 한 불가결한 부분은 자유에 관한 그의 근본적으로 독창적인 개념이다. 그의 이론은 부정적 자유라기보다는 긍정적 자유의 개념을 제시한다는 점에서 특별하다. 즉, 어떤 종류의 제약으로부터의 자유(freedom from some sort of constraint)라기보다는 무언가를 할 수 있는 자유(freedom to do something)의 개념이다.¹²⁾

10) Kant 1996a, 4:447; 4:450.

11) Brandom 1979, p. 187.

12) Brandom 2009, p. 58.

그리고 칸트는 위와 같은 긍정적 의미의 자유와 부정적 의미의 자유(freedom in the negative sense)를 구분한다. 그리고 후자는 ‘외적 원인에 의해 결정되지 않음’(not being determined by alien causes)을 통해 이해될 수 있다.¹³⁾ 긍정적 의미의 자유와 부정적 의미의 자유에 대한 이와 같은 칸트적 구분에 대해 가이어(Paul Guyer)와 앨리슨(Henry Allison)은 다음과 같이 말한다.

부정적으로 이해되는 자유, 즉 외적 원인들에 의해 의지가 결정되는 것으로부터의 자유는 오직 긍정적으로 이해되는 자유, 즉 우리의 의지를 우리 스스로 부여하는 법칙에 의해 결정하는 것을 통해 성취될 수 있다.¹⁴⁾

[자유] 부정적 개념은 감각적 충동들에 의한 결정으로부터의 독립이고, 반면 긍정적 개념은 그 자체로 실천적인 순수 이성의 능력이다. ... 순수 이성이 실천적이라는 주장은 의지가 자율적이라는 주장과 동등하다. ... 그러한 자유의 긍정적 개념은 순수 실천이성이 명하는 바에 따라 행하는 능력, 즉 정언명령들이 요구하는 바대로, 보편 법칙에 부합하도록 준칙들을 선택하는 능력이다.¹⁵⁾

따라서 우리는 자유를 자율성을 통해 이해할 수 있고, 또한 후자는 도덕규범들에 따라 행위할 수 있는 능력으로 이해할 수 있다. 그런데 칸트가 말하는 부정적 의미의 자유와 관련해 주목할 점이 있다.

부정적 의미의 자유에 관한 칸트의 생각은 자유에 관한 스토아학과(Stoicism)의 견해와 유사하다. 스토아학파에 따르면, 자유의 가치는 통제할 수 없는 욕망을 피함으로써 얻을 수 있는 마음의 평온(tranquility)에 있다. 따라서 이들은 자유를 ‘자신의 욕망 또는 다른 행위자에 의해 지배되지 않음’을 통해 이해한다.¹⁶⁾ 그런데 이러한 자유 개념은 행복의 성취와 밀접히 관련되어 있다. 따라서 칸트가 말하는 부정적 의미의 자유 개념이 도덕적

13) Kant 1996a, 4:446.

14) Guyer 2016, p. 138.

15) Allison 2020, p. 455.

16) 부정적 의미의 자유에 관한 칸트의 생각이 자유에 관한 스토아학파의 견해와 연관되어 있다는 주장에 대한 보다 자세한 논의를 위해서는 Guyer 2016, p. 7 그리고 Allison 2020, p. 46을 보기 바람.

책임에 대해 말해주는 바가 분명치 않다. 이런 이유에서 필자는 부정적 의미의 자유를 ‘외적 원인에 의해 결정되지 않음’으로 이해하는 칸트의 견해 대신에 책임의 면제조건(the exemption condition of responsibility)을 통해서 이해하는 것이 좀 더 바람직하다고 생각한다. 다음은 이 조건에 대한 설명이다.¹⁷⁾

우선 우리는 앞서 언급한 바대로 도덕규범에 따라 행위할 수 있는 이성적 존재이다. 그렇기 때문에 우리의 행위들은 도덕규범에 따라 평가될 수 있다. 다시 말해 우리는 단순한 동물들과 달리 우리의 행위들에 대해 도덕적 책임이 있다. 우리는 이 사실을 토마스 홉스(Hobbes 1651)가 언급한 이른바 ‘자연상태’(the state of nature)를 고려함으로써 좀 더 명확히 이해할 수 있다. 홉스에 의하면 자연상태는 옳고 그름에 관한 도덕규범이 없는 상태이다. 도덕규범이 없는 자연상태에서는 각자가 자신이 원하는 바대로 행동할 수 있다. 즉 자신에게 이익이 된다면 언제든지 다른 사람의 것을 약탈할 수도 있고, 심지어 다른 사람의 생명도 빼앗을 수 있다. 그런데 그러한 자연상태는 강자가 약자를 잡아먹는 것이 허용되는, 정글의 법칙에 의해 지배되는 동물의 세계와 다를 바가 없다. 그러나 우리는 우리의 상호관계를 도덕규범에 의해 규제함으로써 자연상태와는 다른 방식으로 공존할 수 있는 이성적 행위자들이다.¹⁸⁾ 예컨대 다른 사람을 아무 이유 없이 해치는 사람이 있다

17) 책임의 면제조건에 관한 이러한 주장은 물론 칸트의 주장이 아니다. 필자의 관심은 칸트의 견해를 충실하게 따르는 데 있지 않고, 그 대신 넓은 의미의 칸트적 관점에서 책임에 관한 적절한 이론을 제시하는 데 있다.

18) 이 논점과 관련하여 한 가지 유념할 점이 있다. 칸트의 견해와 홉스의 견해 사이에는 간과해서는 안 되는 근본적인 차이점들이 있다. 예컨대 홉스의 견해에 따르면, 자기 보존이나 욕망의 충족 등의 좋은 것들(good things)은 도덕규범에 선행하여 존재한다. 그리고 우리가 도덕규범을 따르는 것이 합리적인 이유는 도덕규범이 유지되는 사회에 사는 것이 우리에게 이롭기 때문이다. 따라서 홉스는 도덕규범을 수용하는 것을 그렇게 하는 것이 자기 보존에 유리하다는 이유에서 정당화한다. 반면 칸트(Kant 1996b, 5:63)에 따르면, “선악의 개념은 도덕법칙에 선행하여 결정되지 않는다. 오직 그것에 후행하여 또한 그것에 의해서 결정된다.” 또한 ‘좋은’과 ‘나쁜’과 같은 용어들은 규범적 용어들이다. 그리고 행위를 좋은 것 또는 나쁜 것으로 평가하기 위해서는 평가의 기준이 되는 규범이 있어야 한다. 따라서 그의 견해에 의하면 도덕규범에 선행하여 또한 그것과 독립적으로 도덕적 좋음은 존재하지 않는다. 이런

면, 우리는 그를 도덕규범을 위반한다는 이유에서 처벌할 수 있다. 따라서 우리가 단순한 동물과 달리 이성적 존재인 이유는 우리의 상호관계를 도덕 규범에 의해 규제함으로써 정글의 법칙이 지배하는 동물의 세계와 구분되는 방식의 삶을 살 수 있는 존재이기 때문이다.

위와 같은 이유에서 어떤 행위자 X가 이성적 존재이면, 바로 그 사실에 의해 X는 자신의 행위에 대해 도덕적으로 책임이 있는 존재이다. 따라서 X가 이성적 존재이면, 그가 자신의 한 일에 대해 도덕적 책임이 있다는 것이 **책임의 기본조건(the default condition of responsibility)**이다.

그런데 X에게 적절한 면책사유가 있는 경우에 그러한 기본조건은 무력화될 수 있다. 이것이 **책임의 면제조건(the exemption condition of responsibility)**이다. 다시 말해 한 행위자가 도덕규범에 어긋나는 행위를 했더라도, 그 행위에 대해 적절한 면책사유가 있으면 (적어도 어느 정도) 도덕적 책임에서 벗어날 수 있다.¹⁹⁾ 한 가지 예는 다음과 같다. 예컨대 어떤 아이가 물놀이를 하다 실수로 물에 빠져 익사할 상황에 있다고 하자. 그런데 나 또한 물에 빠져 허우적거리고 있고, 설상가상으로 수영도 못한다고 하자. 그렇지만 다행히 나는 다른 사람의 도움으로 간신히 목숨을 건졌지만, 불행하게도 그 아이는 물속에서 오래 버티지 못해 구조되지 못했다고 하자. 이 경우 나는 그 어린 아이의 근처에 있었지만, 그 아이의 익사에 대한 도덕적 책임에서 벗어날 수 있다. 왜냐하면 내게는 적절한 면책사유가 있기 때문이다. 즉 내게는 그 아이를 구할 능력이 없었다.

앞서 언급한 바대로, 우리는 단순한 동물들과 달리 도덕규범에 규제되는

이유에서 칸트에 따르면 우리가 도덕규범을 준수해야 하는 이유는 그렇게 하는 것이 우리에게 이롭기 때문이 아니다. 또한 그래야만 자연상태를 피할 수 있기 때문도 아니다. 다른 동물들과 달리, 상호관계를 도덕규범에 의해 규제함으로써 자연상태와 다른 방식으로 공존하는 것이 우리의 고유한 이성적 본성이기 때문이다. 그리고 이와 같은 고유한 이성적 본성이 도덕규범의 규범성의 원천이다. 칸트의 견해와 흄스의 견해의 차이점들에 대한 좀 더 자세한 논의를 위해서는 필자의 2018년 논문 “The Moral Law as a Fact of Reason and Correctness Conditions for Moral Law”와 2018년 논문 “도덕법칙의 보편성과 인간가치체계의 상대성”을 보기 바람.

19) 책임의 기본조건과 면제조건에 대한 보다 자세한 논의를 위해서는 필자의 2020년 논문 “의무·능력 원리와 인식적 규범성”을 보기 바람.

이성적 존재이다. 그리고 우리는 도덕규범들에 따라 행위할 수 있는 능력을 갖고 있다는 점에서 긍정적 의미에서 자유로운 존재이다. 이러한 의미의 자유를 칸트는 도덕법칙의 존재근거(*ratio essendi*)라고 말한다.²⁰⁾ 그 이유는 대략적으로 다음과 같다. 그에 따르면 의무(*ought*)는 능력(*can*)을 함축한다. 다시 말해 우리에게 어떤 것을 해야 하는 도덕적 의무가 있다는 것은 우리가 그것을 할 수 있음을 함축한다.²¹⁾ 할 수 없는 것을 하라고 요구하는 것은 부질없는 일이기 때문이다. 따라서 우리는 도덕규범에 따라 행위할 수 있는 능력, 즉 긍정적 의미의 자유를 갖고 있는 한, 도덕규범을 따를 의무를 지닌다. 다시 말해 우리가 단순한 동물들과 달리 도덕규범을 따를 의무를 지니는 이유는 긍정적 의미에서 자유로운 존재이기 때문이다. 다시 말해 긍정적 의미의 자유는 도덕규범의 존재근거이다. 따라서 X가 긍정적 의미에서 자유로운 존재이면, X가 자신의 행위에 대해 도덕적 책임이 있다는 것은 책임의 기본조건이다. 그렇지만 X에게 적절한 면책사유가 있는 예외적인 상황에서는 책임에서 벗어날 수 있다. 여기서 한 가지 주목할 점은 그와 같은 예외적 상황이 성립하는지 여부에 관한 입증 책임은 행위의 주체 또는 그 주체를 옹호하고자 하는 사람에게 있다는 것이다. 따라서 기본조건을 무력화시킬 수 있는 면책사유가 제시되지 않는 한, 기본조건은 유지되고, 따라서 주체는 자신의 행위에 대해 책임을 져야 한다.

5. 윌리스 견해의 문제점들

이제 앞 절에서 간략히 소개한 자유에 관한 칸트적 견해를 토대로 윌리스의 견해를 다시 살펴보자.

20) Kant 1996b, 5:4.

21) 이 원리에 대한 칸트의 언급에 대해서는 다음을 보기 바람. Kant 1963, A548/B576, A807/B835.

5.1. 책임이 있음과 책임을 물을

2절에서 언급한 바대로, 표준적 설명에 따르면 책임이 있음(being responsible)은 책임을 물을(holding responsible)에 분석적으로 선행한다. 즉 X에게 도덕적 책임을 묻는 것이 공정하기 위해서는 X에게 실제로 도덕적 책임이 있어야 한다. 그러나 윌리스는 설명의 순서가 반대여야 한다고 주장한다. 즉 책임이 있음의 개념을 책임을 물을의 개념을 통해 이해해야 한다고 주장한다. 그러나 이 주장이 옳지 않다고 볼 중요한 이유가 있다. 먼저 다음을 비교해 보자.

(2) X는 그가 한 일에 대해 책임이 있다.

(X is responsible for what he has done.)

(3) 우리는 X에게 그가 한 일에 대해 책임을 물을 수 있다.

(We can hold X responsible for what he has done.)

우리가 X에게 그가 한 일에 대해 책임을 묻기 위해서는 우선 X가 도덕적 판단의 대상이어야 한다. 예컨대 X가 단순한 동물인 경우에 도덕적 책임을 묻는 것은 부적절하다. 따라서 우리가 X에게 그가 한 일에 대해 책임을 묻는 것이 유의미하기 위해서는 그가 도덕적 평가의 대상임이 미리 결정돼야 한다. 또한 X가 도덕적 판단의 대상인 경우에도 그에게 책임을 묻는 것이 공정하기 위해서는 그가 자신이 책임이 없다고 주장할 경우에 그 주장을 정당하게 논박할 수 있어야 한다. 다시 말해 그에게 실제로 책임이 있다는 사실에 대해 우리가 정당화할 수 있어야 한다. 또한 그에게 실제로 책임이 있는 경우에도 그에게 피치 못할 사정이 있을 수 있다. 그리고 그 피치 못할 사정이 X로 하여금 책임에서 완전히 벗어날 수 있도록 해 줄 정도는 아니지만, 그럼에도 우리가 책임을 그에게 굳이 묻지 않고 넘어갈 수 있도록 해 줄 수 있는 것일 수 있다. 이런 이유들에서 (2)가 (3)에 분석적으로 선행한다. 그리고 비슷한 이유들에서 스미스(Smith 2007)와 피셔(Fischer 2010)는 책임을 물음이 책임이 있음에 분석적으로 선행한다는 윌리스의 주장에 반대한다.

이제 앞서 논의했던 자유에 관한 칸트적 견해를 토대로 위 논점에 대해 좀 더 부연설명을 하면 다음과 같다. 앞서 언급한 바대로 X가 이성적 존재라는 말은 X가 도덕규범들에 규제된다고 말하는 것과 마찬가지다. 따라서 윌리스의 주장과 달리, X가 도덕적으로 책임이 있는 존재라는 사실은 그가 사람들에게 책임을 묻는 반응적 태도를 소유하기 때문이 아니다. 이와 독립적으로 X가 이성적 존재이면 바로 그 사실에 의해 X는 도덕적으로 책임이 있는 존재이다. 그리고 X가 이성적 존재이면 그가 자신이 한 일에 대해 도덕적 책임이 있다는 것이 책임의 기본조건이다. 따라서 이러한 기본조건을 무력화할 수 있는 책임의 면책사유가 없는 한, 그는 자신이 한 일에 대해 도덕적 책임이 있다. 그런 경우 우리는 정당하게 그에게 책임을 물을 수 있다. 반면 X에게 적절한 면책사유가 있어서 그가 책임에서 벗어날 수 있는 경우엔 그에게 책임을 물어서는 안 된다. 따라서 어떤 행위주체에게 책임을 묻는 것이 공정하기 위해서는 그에게 실제로 책임이 있어야 한다. 이런 이유에서 책임을 물음은 책임이 있음에 분석적으로 선행하지 않는다. 또한 도덕적 책임이 일차적으로 이성적 능력의 문제임을 주장하기 위해 도덕적 책임에 관한 스트로슨식의 반응적 설명을 받아들일 필요가 없다.

5.2. 반성적 평형과 합리성에 관한 칸트적 견해

2절에서 언급했던 것처럼 윌리스는 공정함에 관한 도덕원리에 관해 반성적 평형의 방법을 사용할 것을 제안한다. 다시 말해 유능한 도덕적 평가자들이 높은 단계의 자신감을 갖는 숙고된 확신을 토대로, 양립론자와 비양립론자가 모두 동의할 수 있는, 공정함에 관한 도덕원리를 결정할 수 있다고 주장한다. 그리고 이와 같은 숙고된 확신에 따르면, 특정 수준 이상의 반성적 자기 통제력을 소유하지 못한 존재들, 예컨대 어린아이나 사이코패스에게 도덕적 책임을 묻는 것은 공정하지 않다고 주장한다.

그러나 반성적 평형에 호소하는 것에는 중요한 난점들이 있다. 첫째, 유능한 도덕적 평가자들의 집합을 결정하기 쉽지 않다. 어떤 기준에 의해 누구는 이 집합에 속하고, 누구는 이 집합에서 배제되는가? 둘째, 우리가 유능

한 도덕적 평가자들의 집합을 결정할 수 있다고 해도, 반성적 평형의 방법은 의견의 수렴을 보장하지 않는다. 다시 말해 모든 유능한 도덕적 평가자들이 위와 같은 숙고된 확신에 동의할 것이라는 보장이 없다. 셋째, 유능한 도덕적 평가자들이 공정함에 관한 어떤 도덕원리에 동의한다고 할지라도, 누군가는 그러한 원리가 왜 옳은지에 대해 정당화 요구를 할 수 있다. 유능한 도덕적 평가자들이 높은 단계의 자신감을 갖고 동의한다는 사실만으로는 그와 같은 정당화 요구에 답할 수 없다. 공정함에 관한 그 특정한 원리가 왜 옳은지에 대해 답해야 한다.

반면 필자의 견해는 위와 같은 문제들을 피할 수 있다. 필자는 셀라시언 정합성 이론을 받아들인다. 이 이론에 따르면, 정당화의 요구에 답하기 위해서 우리는 정당화를 요구하고 이에 답하는 우리의 사회실천에서 출발해야 한다. 그리고 정당화의 사회실천 속에서 주어진 주장에 대해 제기될 수 있는 모든 비판들에 답할 수 있는 경우에 그 주장은 정당화된²²⁾ 그리고 이 정당화 이론에 따르면 우리가 이성적 존재임은 대략적으로 다음과 같은 이유에서 추정적 정당화(default justification)의 위상을 가진다.

정당화를 요구하고 이에 답하는 우리의 사회실천 속에서 우리가 이성적 존재인지에 대해 정당화하라는 요구에 어떻게 답할 수 있는가? 어떤 믿음에 대한 정당화 요구에 답하기 위해서 우리는 인식규범들(epistemic norms)을 의거해서 그 믿음이 인식적으로 정당화되는지 여부를 평가해야 한다. 예컨대 어떤 사람이 전건긍정추론 규칙에 어긋나는 방식으로 어떤 것을 믿는다면, 그 믿음은 인식규범을 위반한다는 근거에서 인식적으로 정당화되지 않

22) 이 이론은 정당화되는 것으로 여겨지는 것(being taken as justified)과 실제로 정당화되는 것(being really justified)을 구분한다. t_1 시점에서 어떤 주장에 대해 제기된 모든 비판들에 답할 수 있는 경우에 그 주장은 그 시점에서 정당화의 위상을 지닌다. 그렇지만 미래 시점 t_2 에서 제기되는 비판에 의해 그 주장은 미래에 정당화의 위상을 잃을 수 있다. 그런 경우 그 주장은 실제로 정당화되는 주장이 아니라, t_1 시점에서 단지 정당화되는 것으로 여겨진 것에 불과하다. 셀라시언 정합성 이론에 대한 보다 자세한 논의를 위해선 필자의 2017년 논문 “The Truth-Conduciveness Problem of Coherentism and a Sellarsian Explanatory Coherence Theory”와 필자의 2022년 책 『현대인식론: 정당화의 사회실천에 의한 접근』을 보기 바람.

는다고 평가할 수 있다. 어떤 인식규범들에 관해서 과연 그 규범들이 옳은 규범들인지에 대해 논란의 여지가 있을 수 있다. 그러나 우리는 적어도 옳은 인식규범들이 있고, 우리가 그러한 규범들에 따라 믿어야 한다는 추정 하에서 우리의 믿음들을 인식적으로 평가한다. 행위의 경우도 마찬가지이다. 어떤 행위에 대한 정당화 요구에 답하기 위해서 우리는 도덕규범들(moral norms)을 통해서 그 행위가 도덕적으로 정당화되는지 여부를 평가해야 한다. 예컨대 어떤 사람이 아무 이유 없이 타인을 해치면 그 행위는 도덕규범을 위반한다는 이유에서 도덕적으로 정당화되지 않는다고 평가할 수 있다. 어떤 도덕규범들에 관해서 과연 그 규범들이 옳은 규범들인지에 대해 논란의 여지가 있을 수 있다. 그러나 우리는 적어도 옳은 도덕규범들이 있고, 우리가 그러한 규범들에 따라 행위해야 한다는 추정 하에서 우리의 행위들을 도덕적으로 평가한다. 이런 이유에서 우리가 단순한 동물들과 달리 도덕규범들에 의해 규제되는 이성적 존재들이라는 것은 정당화의 사회실천 속에서 추정적 정당화의 위상을 지닌다. 즉 우리가 그러한 이성적 존재라는 우리 자신에 관한 근본적 개념을 뒤엎을 수 있는 적극적 이유가 제시되지 않는 한, 우리 자신을 그와 같은 이성적 존재로 간주하는 것은 정당화의 사회 실천 속에서 정당화된다.

요컨대 필자의 견해는 논란의 여지가 많은 반성적 평형의 방법에 호소하는 대신에 셀라시언 정당화 이론을 사용한다. 따라서 사람들에게 도덕적 책임을 묻는 우리의 사회실천을 좀 더 성공적으로 옹호할 수 있다.

5.3. 이성적 능력, 긍정적 의미의 자유, 그리고 결정론

앞서 2절에서 언급했던 바대로, 전통적으로 도덕적 책임과 자유를 연관시킨 이유는, 선택의 자유가 없는 사람에게 책임을 묻는 것은 불공정하다고 생각했기 때문이다. 윌리스는 이런 의미의 자유를 ‘의지의 강한 자유’라고 부른다. 그런데 그에 따르면 도덕적 책임은 일차적으로 반성적 자기 통제력의 문제이고, 이러한 반성적 자기 통제력은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유를 선제하지 않는다. 다시 말해 도덕적 책임의 주체

인지 여부는 반성적 자기 통제력이 있는지의 문제이지, 의지의 강한 자유가 있는지의 문제가 아니라고 주장한다.

앞서 3절에서 언급했던 바대로, 필자는 도덕적 책임이 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유에 의존하지 않는다는 데에 윌리스와 의견을 같이 한다. 그렇지만 비록 이런 의미의 자유 개념은 받아들이지 않더라도, 자유와 결정론 논쟁과 관련하여 도덕적 책임 개념과 자유 개념이 완전히 무관하다고 말하기는 어렵다. 그렇다면 이 논쟁과 유관한 자유 개념은 무엇인가? 필자의 칸트적 견해는 이 물음에 적절한 답을 제시한다. 왜냐하면 앞서 4절에서 논의했던 긍정적 의미의 자유 개념은 왜 우리가 사람들에게 도덕적 책임을 물을 수 있는지를 잘 설명해 주기 때문이다. 긍정적 의미에서 자유로운 존재는 도덕규범들에 따라 행위할 수 있는 존재이다. 그리고 그런 존재에는 도덕적 책임의 기본조건이 성립한다. 따라서 긍정적 의미의 자유와 도덕적 책임은 직접적으로 관련이 있다.

또한 필자의 견해는 결정론이 참일 경우에 도덕적 책임을 묻는 사회실천이 정당화되지 않는다는 비양립론자의 도전으로부터 도덕적 책임을 묻는 현재의 사회실천을 방어함에 있어서 윌리스의 견해에 비해 훨씬 더 효과적이다. 앞서 언급한 바대로 우리는 정당화의 사회실천에 참여할 수 있는 이성적 존재이다. 비양립론자도 자신이 그러한 이성적 존재임을 부정하기 어렵다. 그것을 부정하는 것은 자신이 주장하는 비양립론을 합리적으로 옹호할 수 있음을 부정하는 것과 마찬가지로 때문이다. 그리고 이와 관련해 두 가지 점에 주목할 필요가 있다.

첫째, 앞서 지적했던 바대로, 정당화를 요구하고 이에 답하는 우리의 사회실천 속에서 우리는 어떤 믿음이 정당화되는지 여부를 인식규범들에 따라 평가한다. 예컨대 어떤 사람이 전건긍정추론 규칙에 어긋나는 방식으로 어떤 것을 믿는다면, 그 믿음은 인식규범을 위반한다는 이유에서 인식적으로 정당화되지 않는다고 평가할 수 있다. 또한 그러한 평가에 따라 그 사람에게 전건긍정추론 규칙에 부합하게 믿음을 수정할 것을 요구할 수 있다. 그런데 그러한 요구를 하기 위해서는 인식규범들에 부합하게 믿음을 수정

할 수 있는 이성적 능력이 그에게 있어야 한다. 앞서 4절에서 지적한 바대로, 할 수 없는 것을 하라고 요구하는 것은 부질없는 일이기 때문이다. 마찬가지로 어떤 사람이 아무 이유 없이 다른 사람에게 위해를 가한다면, 그 행위는 도덕규범을 위반한다는 이유에서 도덕적으로 정당화되지 않는다고 평가할 수 있다. 또한 그러한 평가에 따라 그 사람에게 도덕규범에 부합하게 행위를 수정할 것을 요구할 수 있다. 그리고 그러한 요구를 하기 위해서는 도덕규범들에 부합하게 행위를 수정할 수 있는 이성적 능력이 그에게 있어야 한다. 그런데 인식규범들을 따라 믿을 수 있는 능력과 도덕규범들에 따라 행위할 수 있는 능력은 둘 다 규범을 따를 수 있는 이성적 능력이다. 따라서 결정론이 전자의 이성적 능력과는 양립하지만, 후자의 능력과는 양립하지 않는다고 주장하는 비양립론자가 있다면, 우리는 어떻게 그럴 수 있는지에 대한 적절한 설명을 요구할 수 있다. 다시 말해 이 문제에 대한 입증의 책임은 비양립론자에게 있다.

둘째, 우리에게 도덕규범들에 따라 행위를 수정할 수 있는 능력이 있음을 부정하는 것은 우리가 우리의 상호관계를 도덕규범에 의해 규제함으로써, 정글의 법칙에 의해 지배되는 단순한 동물들과는 구분되는 방식으로 공존할 수 있는 이성적 행위자임을 부정하는 것이다. 그러나 앞서 5-2절에서 지적했던 바대로, 우리가 단순한 동물들과 달리 도덕규범들에 의해 규제되는 이성적 행위자들이라는 것은 정당화의 사회실천 속에서 추정적 정당화의 위상을 지닌다. 따라서 그러한 추정적 정당화를 무력화할 수 있는 적극적 근거를 비양립론자가 제시하지 못하는 한, 우리에게 도덕규범들에 따라 행위를 수정할 수 있는 능력이 있음을 부정하기 어렵다. 다시 말해 우리가 그러한 이성적 존재임을 부정할 수 있는 적극적 근거를 제시하지 못하는 한, 비양립론자는 결정론을 우리가 그러한 이성적 존재라는 사실과 양립할 수 없는 방식으로 주장하기 어렵다.

요컨대 비양립론자가 결정론이 인식규범들에 따라 믿을 수 있는 능력과는 양립하지만, 도덕규범들에 따라 행위할 능력과는 양립하지 않음을 보이고, 또한 우리가 단순한 동물들과 달리 도덕규범들에 의해 규제되는 이성적

행위자들이라는 우리 자신에 관한 근본 개념을 논박할 수 있는 적극적 근거를 제시하지 못하는 한, 우리는 도덕적 책임을 묻는 현재의 사회실천을 성공적으로 방어할 수 있다.

5.4. 도덕적 책임의 기본조건과 면제조건

월리스는 책임조건과 비난조건을 구분한다. 어떤 행위자에게 책임을 묻는 것이 공정하기 위해서는 책임조건이 성립해야 한다. 즉 그가 특정 수준 이상의 반성적 자기 통제력을 소유해야 한다. 또한 그가 한 행동을 비난하는 것이 공정하기 위해서는 비난조건이 성립해야 한다. 다시 말해 정당한 변명거리가 없어야 한다. 반면 필자는 책임의 기본조건과 면제조건을 구분한다. S가 이성적 존재이면, 그가 자신의 한 일에 대해 도덕적 책임이 있다는 것이 책임의 기본조건이다. 그러나 그에게 적절한 면책사유가 있는 예외적 경우에 그러한 기본조건은 무력화될 수 있다. 이것이 책임의 면제조건이다. 그런데 월리스의 책임조건/비난조건 구분과 필자의 기본조건/면제조건 구분 사이에 중요한 차이점들이 있다. 그 중 가장 중요한 두 가지 차이점들은 다음과 같다.

첫째, 도덕규범에 종속되는 존재들과 관련하여 중요한 차이점이 있다. 월리스에 따르면 일정 수준 이상의 반성적 자기 통제력을 가진 존재들에게만 책임조건이 성립한다. 따라서 어린 아이나 사이코패스처럼 반성적 자기 통제력에 심각한 문제가 있는 존재들에는 책임조건이 성립하지 않는다. 그렇다면 책임조건이 성립하는 반성적 자기 통제력의 수준은 무엇인가? 월리스는 이에 대한 답을 제시하지 않는다. 그리고 이 물음에 답하기 쉽지 않다. 또한 같은 이유에서 책임조건이 성립하는 주체들의 집합을 결정하기 어렵다. 반면 필자의 칸트적 견해에 따르면, 어떤 존재가 이성적 존재이면, 바로 그 사실에 의해 책임의 기본조건이 성립한다. 즉 그가 한 행위에 대해 적절한 면책사유가 없다면 책임을 져야 한다. 따라서 이 견해에 따르면 어린아이나 사이코패스와 같은 존재도 이성적 존재인 한, 도덕규범에 종속된다.

위 논점과 관련해 주목할 점은, 도덕규범은 보편규범이어서 모든 이성적

존재들에게 예외 없이 적용돼야 한다는 것이다. 그리고 어린아이들도 우리의 도덕공동체의 일원이다. 따라서 개와 같은 단순한 동물들과 달리, 어린 아이들은 도덕규범에 따라 행위할 수 있도록 훈육돼야 한다. 그리고 그렇게 훈육하는 것이 정당화되는 이유는 어린아이들도 우리의 도덕공동체의 일원이고, 따라서 도덕규범들의 예외가 아니기 때문이다. 다만 아직은 어려서 도덕규범에 따라 행위하도록 충분히 훈육된 상태가 아니라는 사실은 그 아이들이 도덕적 책임에서 어느 정도 벗어날 수 있는 적절한 면책사유가 될 수 있다.

도덕규범을 따르는 능력과 관련해 심리적 결함이 있는 사이코패스의 경우도 크게 다를 바 없다. 연쇄살인을 저지른 어떤 사이코패스가 반사회적 인격 장애로 인해 살인충동을 의지적으로 통제하지 못한다고 가정해 보자. 정당한 이유 없이 남을 해쳐서는 안 된다는 도덕규범은 보편규범이므로 이 규범은 그 사이코패스에게도 예외 없이 적용된다. 이와 관련해 주목해야 할 점이 있다. 비록 어떤 점에서 완전히 합리적이지는 않지만, 그 사이코패스도 단순한 동물과 달리 이성적 존재이다. 예컨대 그는 다른 사람을 해치는 것이 사회적으로 허용되지 않음을 이해할 수 있다. 따라서 그는 주변에 자신을 주시하는 경찰관이 있어서 범행이 발각될 수밖에 없는 특수한 상황에서는 남을 해치고자 하는 충동을 억제할 수 있다. 이런 이유에서 그도 우리의 도덕 공동체의 일원이다. 다시 말해 그도 도덕규범의 예외가 아니다. 다만 안타깝게도 그에게는 우리가 좌시할 수 없는 심각한 도덕적 결함이 있다. 그리고 우리가 그의 도덕적 결함에 대응하는 한 가지 방안은 그를 정신 병원에 보내 치료받도록 하는 것이다. 그러한 조치를 강제할 수 있는 것은 그 사이코패스도 우리의 도덕 공동체의 일원이고, 따라서 도덕규범의 예외가 아니기 때문이다.

둘째, 윌리스의 비난조건과 필자의 면제조건 사이에 중요한 차이점이 있다. 윌리스에 따르면, 비난조건이 성립하면 주체가 한 행동을 비난하는 것이 공정하다. 그렇지만 주체에게 정당한 변명거리가 있는 경우에는 그의 행위를 비난하는 것이 공정하지 않다. 그리고 그렇게 비난하는 것이 공정하지

않은 이유는 그가 아무런 잘못을 범하지 않았기 때문이다. 즉 아무 잘못이 없기 때문에 비난을 하는 것이 공정하지 않다는 것이다. 월리스는 이것을 ‘잘못이 없으면 비난도 없음 원리’(the principle of no blameworthiness without fault)라고 부른다.²³⁾ 그러므로 월리스의 견해에 따르면, 변명조건이 성립한다고 말하는 것은 주체가 특정 행동을 함에 있어서 아무런 잘못을 범하지 않았다고 말하는 것이다. 또한 주체에게 변명조건이 성립하는지 여부는 정도의 문제(a matter of degree)가 아니라, 전부 아니면 전무(all or nothing)의 문제이다. 그러나 필자의 견해는 다르다.

책임의 기본조건이 적절한 면책사유에 의해 무력화된다는 사실은 주체가 아무런 잘못을 하지 않았음을 함축하지 않는다. 한 가지 예시는 다음과 같다. 어떤 어린아이가 장난을 심하게 치다가 아무 잘못이 없는 다른 아이를 불구자로 만들었다고 가정해 보자. 정당한 이유 없이 남을 해쳐서는 안 된다는 도덕규범은 보편규범이므로 이 규범은 그 어린아이에게도 예외 없이 적용된다. 그렇지만 그 어린아이는 자신이 무슨 잘못에 했는지에 대해 충분히 이해하지 못할 수 있다. 그런 경우에 책임의 면제조건이 적어도 어느 정도 성립할 수 있다. 그렇지만 그런 경우에도 그 어린아이를 호되게 야단침으로써 비슷한 잘못을 반복하지 않도록 훈육하는 것은 허용될 수 있다. 그리고 그렇게 하는 것이 허용되는 이유는 정당한 이유 없이 남을 해쳐서는 안 된다는 도덕규범이 그 어린아이에도 예외 없이 적용되기 때문이다. 또한 그 어린아이가 비록 아직 어리긴 하지만, 그렇다고 해서 다른 사람을 불구자로 만든 책임에서 완전히 면제된다고 말하기 어렵다. 따라서 비록 우리가 그 어린아이를 도덕적으로 비난하는 대신에 비슷한 잘못을 반복하지 않도록 훈육한다고 해서, 이 사실이 그 아이가 아무런 잘못을 범하지 않았음을 의미하는 것은 아니다. 그리고 이 점에서 인간의 아이들과 개와 같은 단순한 동물들 사이에 중요한 차이가 있다.

23) Wallace 1994, p. 135.

6. 나오는 말

필자는 다음 네 가지 점들에서 윌리스에 동의한다. 첫째, 도덕적 책임은 일차적으로 이성적 능력의 문제이다. 둘째, 도덕적 책임은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유에 의존하지 않는다. 셋째, 비양립론을 위한 일반화 전략은 성공적이지 않다. 넷째, 우리에게 이성적 능력이 있다는 것과 결정론이 양립하지 않음을 보여주는 적절한 근거를 비양립론자들이 지금껏 제시하지 못했다.

그렇지만 필자는 다음 점들에서 윌리스와 견해를 달리한다.

첫째, 책임을 물음은 책임이 있음에 분석적으로 선행하지 않는다. 따라서 도덕적 책임을 일차적으로 우리의 반응적 태도를 통해 설명하는 것은 옳지 않다. 또한 도덕적 책임이 일차적으로 이성적 능력의 문제임을 주장하기 위해 도덕적 책임에 관한 스트로슨식의 반응적 설명을 받아들일 필요가 없다.

둘째, 필자의 견해는 논란의 여지가 많은 반성적 평형의 방법에 호소하는 대신에 셀라시언 정당화 이론을 사용한다. 따라서 사람들에게 도덕적 책임을 묻는 우리의 사회실천을 좀 더 성공적으로 옹호할 수 있다.

셋째, 윌리스에 따르면 도덕적 책임은 일차적으로 반성적 자기 통제력의 문제이고, 이러한 반성적 자기 통제력은 달리 할 수 있는 능력으로 이해되는 의지의 강한 자유를 선제하지 않는다. 따라서 이러한 의미의 자유 개념은 사람들에게 도덕적 책임을 물을 수 있는지의 문제와 직접적인 관련이 없게 된다. 반면 필자가 받아들이는, 칸트가 말하는 ‘긍정적 의미의 자유’는 사람들에게 도덕적 책임을 물을 수 있는지의 문제와 직접적인 관련이 있다. 왜냐하면 우리가 이런 의미에서 자유로운 존재라고 말하는 것은 도덕규범에 따라 행위할 수 있는 존재라고 말하는 것이고, 그러한 존재에게는 도덕적 책임의 기본조건이 성립하기 때문이다.

또한 필자의 견해는 결정론이 참일 경우에 도덕적 책임을 묻는 사회실천이 정당화되지 않는다는 비양립론자의 도전으로부터 도덕적 책임을 묻는 현재의 사회실천을 방어함에 있어서 윌리스의 견해에 비해 훨씬 더 효과적

이다. 우리는 정당화의 사회실천에 참여할 수 있는 이성적 존재이다. 비양립론자도 자신이 그러한 이성적 존재임을 부정하기 어렵다. 그것을 부정하는 것은 자신이 주장하는 비양립론을 합리적으로 옹호할 수 있음을 부정하는 것과 마찬가지로 때문이다. 그리고 이와 관련해 두 가지 점에 주목할 필요가 있다.

첫째, 인식규범들을 따라 믿을 수 있는 능력과 도덕규범들에 따라 행위할 수 있는 능력은 둘 다 규범을 따를 수 있는 이성적 능력이다. 따라서 결정론이 전자의 이성적 능력과는 양립하지만, 후자의 능력과는 양립하지 않는다고 주장하는 비양립론자가 있다면, 우리는 어떻게 그럴 수 있는지에 대한 적절한 설명을 요구할 수 있다. 다시 말해 이 문제에 대한 입증의 책임은 비양립론자에게 있다.

둘째, 우리에게 도덕규범들에 따라 행위를 수정할 수 있는 능력이 있음을 부정하는 것은 우리가 우리의 상호관계를 도덕규범에 의해 규제함으로써 단순한 동물들과는 구분되는 방식으로 공존할 수 있는 이성적 행위자임을 부정하는 것이다. 그런데 우리가 단순한 동물들과 달리 도덕규범들에 의해 규제되는 이성적 행위자들이라는 것은 정당화의 사회실천 속에서 추정적 정당화의 위상을 지닌다. 따라서 그러한 추정적 정당화를 무력화할 수 있는 적극적 근거를 비양립론자가 제시하지 못하는 한, 우리에게 도덕규범들에 따라 행위를 수정할 수 있는 능력이 있음을 부정하기 어렵다.

요컨대 비양립론자가 결정론이 인식규범들에 따라 믿을 수 있는 능력과는 양립하지만, 도덕규범들에 따라 행위할 능력과는 양립하지 않음을 보이고, 또한 우리가 단순한 동물들과 달리 도덕규범들에 의해 규제되는 이성적 행위자들이라는 우리 자신에 관한 근본 개념을 논박할 수 있는 적극적 근거를 제시하지 못하는 한, 우리는 도덕적 책임을 묻는 현재의 사회실천을 성공적으로 방어할 수 있다.

끝으로, 윌리스의 책임조건과 비난조건 구분과 필자의 기본조건과 면제조건 구분 사이에 중요한 차이점들이 있다. 윌리스에 따르면 일정 수준 이상의 반성적 자기 통제력을 가진 존재들에게만 책임조건이 성립한다. 그런

데 이 견해는 책임조건이 성립하는 반성적 자기 통제력의 수준을 정하는 데 어려움이 있다. 또한 윌리스에 따르면 주체에게 변명조건이 성립하는지 여부는 전부 아니면 전무의 문제이다. 그러나 주체의 책임이 어느 정도 면제되는지는 정도의 문제로 보는 것이 좀 더 설득력이 있다.

참고문헌

- 이병덕 (2018) “도덕법칙의 보편성과 인간가치체계의 상대성”, 『철학적 분석』 39호, pp. 29-56.
- 이병덕 (2020) “의무 · 능력 원리와 인식적 규범성”, 『철학』 145집, pp. 161-184.
- 이병덕 (2022) 『현대인식론: 정당화의 사회실천에 의한 접근』, 개정판. 성균관대학교 출판부.
- Allison, Henry E. (2020) *Kant's Conception of Freedom: A Developmental and Critical Analysis*. Cambridge: Cambridge University Press.
- Brandom, Robert. (1979) “Freedom and Constraint by Norms”, *American Philosophical Quarterly* 16, pp. 187-196.
- Brandom, Robert (2009) *Reason in Philosophy: Animating Ideas*. Cambridge: Harvard University Press.
- Eshleman, Andrew S. (2001) “Being is not Believing: Fischer and Ravizza on Taking Responsibility”, *Australasian Journal of Philosophy* 79(4), pp. 479-490.
- Fischer, John Martin (2010) “Responsibility and Autonomy”, in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*. Oxford: Wiley-Blackwell, pp. 309-337.
- Frankfurt, Harry (1969) “Alternative Possibilities and Moral Responsibility”, *Journal of Philosophy* 66, pp. 829-839.
- Guyer, Paul (2016) *Virtues of Freedom: Selected Essays on Kant*. Oxford: Oxford University Press.
- Hobbes, Thomas (1968) *Leviathan*. Reprinted from the edition of 1651, ed. C. B. Macpherson. London: Penguin Books.
- Kant, Immanuel (1963) *Critique of Pure Reason*. Trans. by Norman

- Kemp Smith. London: Macmillan.
- Kant, Immanuel (1996a) "Groundwork of the Metaphysics of Morals." In Mary J. Gregor (trans. and ed.), *Practical Philosophy*. Cambridge: Cambridge University Press, pp. 37 - 108.
- Kant, Immanuel (1996b) "Critique of Practical Reason" in Mary J. Gregor (trans. and ed.), *Practical Philosophy*. Cambridge: Cambridge University Press, pp. 133 - 271.
- Lee, Byeong D. (2017) "The Truth-Conduciveness Problem of Coherentism and a Sellarsian Explanatory Coherence Theory", *International Journal of Philosophical Studies* 25(1), pp. 63-79.
- Lee, Byeong D. (2018) "The Moral Law as a Fact of Reason and the Correctness Conditions for the Moral Law", *Dialogue: Canadian Philosophical Review* 57(1), pp. 47-66.
- Smith, Angela (2007) "Being Responsible and Holding Responsible", *Journal of Ethics* 11, pp. 465-484.
- Strawson, P. F. (1962) "Freedom and resentment", *Proceedings of the British Academy* 48, pp. 1-25.
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.

성균관대학교 철학과 교수

Email: bydlee@skku.edu

Wallace's Compatibilism and the Conditions for Moral Responsibility

Byeong Deok Lee

Jay Wallace argues for a compatibilist view which holds that moral responsibility is compatible with determinism. According to him, moral responsibility is primarily a matter of reflective self-control, and this rational ability does not require strong freedom of the will which is understood as the ability to do otherwise. In addition, he argues that incompatibilists have so far failed to show that this rational ability is incompatible with determinism. Wallace's compatibilist theory explains well why determinism is not a serious threat to the legitimacy of our social practice of holding people morally responsible for their actions. Unfortunately, however, his theory is problematic in many respects. The goal of this paper is to offer a new compatibilist account which can avoid problems with Wallace's theory. In particular, I argue that compatibilism can be better defended on the basis of the default-and-exemption model of moral responsibility.

[Subject] Ethics

[Keywords] Determinism, Freedom, Compatibilism, Moral responsibility,
Jay Wallace, Kant

