

【논문】

현대 황금률의 도덕철학적 문제

박종준

【주제분류】 윤리학, 종교철학

【주요어】 황금률, 칸트, 싱어, 거워드, 일반해석, 특수해석, 합리화된 황금률, 초과 의무

【요약문】 황금률은 보편적인 도덕원리로 적절하지 않다는 비판에 직면하여 황금률의 도덕적인 지위를 구하고자 하는 노력의 과정에서 황금률은 적지 않은 변화를 겪었고 그 과정에서 수정된 현대의 황금률들이 태어났다. 이러한 과정에서 나타난 가장 큰 변화는 황금률의 형식에 있다. 무엇보다도 현대의 황금률은 모두 일반화된 형식을 가지고 있다. 그리고 이러한 일반화된 황금률이 지향하는 도덕철학적 특징은 평등주의, 상호주의, 그리고 합리주의이다. 이 논문은 황금률에 제기된 비판에 직면하여 수정되어온 현대 황금률이 심각한 도덕철학적 위험성을 안고 있다는 것을 논변한다. 그리고 이러한 도덕철학적 위험성은 고대의 황금률이 추구하는 가치와 모순된다는 점을 논증함으로써 전통적 비판으로부터 황금률을 구원하려는 노력이 실패할 수밖에 없음을 논변한다.

I. 황금률의 연대기

황금률이란 무엇인가? BC. 700년경부터 전해진, 어쩌면 그보다 더 오래 전으로 거슬러 올라가는, 동서양의 주요 종교들뿐만 아니라 철학과 정치사상에서 다양하게 가르쳐진 황금률의 연대기가 존재한다. BC 600년 경, 탈레스는 “우리가 비난하는 타인의 행동을 결코 하지 않음”, 그리고 기원전 428-347년경 플라톤의 “나의 재산을 다른 사람이 마음대로 다루기를 원하지 않듯이 나도 다른 이들의 재산을 마음대로 다루지 말 것”이나, 그리고 이후 384-322년 경 아리스토텔레스의 “덕이 있는 사람은 자신에게 하듯이 친구를 대한다.”¹⁾ 등은 모두 **탁월함(덕)**의 방법론이다. BC 500년 경 자이나교는 “**수도자는** 모든 존재자들을 자신이 대우받기를 원하는 대로 대우해야 한다.”는 말로 **비폭력, 공감, 그리고 삶의 신성함**을 가르쳤다. 기원전 563-483년 경으로 알려진 붓다의 황금률 “당신에게 소중한 것은 다른 사람에게도 소중한 것이니 당신에게 고통스러운 것으로 다른 이를 고통스럽게 하지 말라.”²⁾와 고대 힌두교의 가르침인 “모든 창조물을 자신처럼 생각하고 마치 자신에게 하듯 그들에게 하는 사람”은 아힘사(ahimsā) 즉, 모든 **피조물들에 대한 사랑**의 원리를 표방한다. 우리는 이러한 모든 것을 소위 ‘황금률’이라고 부른다. 이러한 황금률은 다음과 같이 때로는 적극적인 때로는 소극적인 일반형식으로 재생산 되어왔다.

- 적극적 황금률: “다른 사람들이 당신에게 행하기를 원하는 것을 그들에게 하라.”
- 소극적 황금률: “다른 사람들이 당신에게 행하기를 원하지 않는 것을 그들에게 하지 말라.”

이러한 일반적 형식의 황금률은 1785년에 이르러 도덕원리로서 자격미달이라는 칸트의 비판 이후로 유럽지성사에서 황금률에 관한 논의는 최소

1) *Nicomachean Ethics* 9.

2) *Dhammapada, Northern Canon*, 5: 18.

한 150여 년 간 거의 ‘침묵에 가까운 공백기’에 들어간다.³⁾ 그리고 현대에 이르러서야 비로소 칸트의 비판으로부터 황금률을 구하고자 하는 여러 노력들이 시도되고 있다. 대표적으로, 싱어(Singer, 1963)와 거워드(Gewirth, 1978) 등은 일반적 형식의 황금률에 대해 그 형식을 수정하거나 내용을 보완하는 방식으로 황금률을 구하고자 한다. 또한 블랙스톤(Blackstone, 1965)과 라이니카이넨 (Reinikainen, 2005) 등은 황금률을 메타 도덕원리로 규정함으로써 그 내용의 보완이나 수정 없이 전통적 황금률 자체를 구하려고 한다. 한편, 현대의 게임이론에서는 전통적인 황금률의 일반적인 형식이 축자적으로 활용되고 있기도 하다.

황금률이 변화할 수밖에 없었던 이유는 고대의 황금률이 일반화 된 형식으로 표현되었을 때 심각한 도덕철학적 난점을 가진다는 비판 때문이었다. 그리고 그러한 변화의 역사 끝에 현대의 황금률들이 탄생하였다. 이러한 비판과 옹호의 과정에서 현대 황금률은 도덕철학적으로 더 발전했을까? 결론은 그와 정반대이다. 문제는 현대의 일반화된 형식의 황금률과 현대 도덕철학의 주요 이념인 상호주의, 평등주의, 그리고 합리주의의 결합에 있다. 일반화된 형식의 황금률과 현대 도덕철학의 주요이념의 결합은 황금률에 심각한 도덕철학적 위험성을 심어놓고 있다. 따라서 이러한 방식, 즉 현대 황금률의 일반화된 형식에 도덕철학의 주요 이념들의 결합을 통해 고대 황금률의 정신을 구하고자 하는 시도는 실패할 수밖에 없다. 이 논문은 황금률에 대한 전통적인 비판에 맞선 옹호의 과정에서 현대의 황금률이 어떻게 현대 도덕철학의 주요 이념들인 상호주의화, 평등화, 합리주의화 되는지를 분석한다. 그리고 현대 일반화된 형식의 현대 황금률에 내재하는 도덕철학적 위험성을 분석함으로써 그것이 황금률을 구원하는데 실패할 수밖에 없는 이유가 된다는 것을 논변한다. 마지막으로 황금률의 다양한 출처들과 맥락들을 분석함으로써 진정으로 황금률을 구원하는 길은 일반화를 포기하고 그 특수한 맥락에서의 가치를 보존하는 것임을 논변한다.

3) Wattles (1987), p. 107.

II. 황금률의 주요 출처에서 나타나는 도덕철학

황금률의 주요 출처는 주로 종교 텍스트이며 그것도 역사적으로 철학 텍스트에 등장하는 것보다 더 이른 시기에 나타난다.⁴⁾ 그런데 소위 ‘황금률’의 역사에서 현대의 형식화된 황금률 이 문자 그대로 언급되는 경우는 거의 없다. 즉 황금률의 주요 출처에서 보면, 황금률은 일반적인 형식보다는 구체적인 상황과 구체적인 행위의 종류나 내용이 명료하게 표명되고 있다. 따라서 황금률은 특정한 맥락이나 상황에서 특정한 문제에 대한 해법으로 제시된 것이다. 고대의 황금률이 이렇게 특수한 맥락에서의 (때로는 특수한 행위자에게 요구되는) 가치를 표방하기 때문에 일반화 된 명제로 형식화되는 것은 적절하지 않다. 그렇다면 특수한 맥락에서 특수한 관계에 있는 행위자에게 요구되는 것은 무엇인가? 그것은 평등주의, 상호주의, 또는 합리화 된 기성 도덕철학에서 원자화된 개인들 간의 관계들에서는 기대하기 어려운 초과무적 행위에 대한 강조이다. 이 장에서는 황금률이 어떤 특수한 맥락에서 제시된 것이며, 어떤 가치를 표방하는 것인지 살펴 본 후 그러한 가치들이 현대 도덕철학의 개념들에서는 모두 도덕의 영역을 초월하는 초과무적 가치들로 이해되고 있다는 것을 논의한다.

1. 유대인의 황금률: 신과 인간의 관계 규정

기록으로 남아있는 종교적 전통에서 황금률의 역사는 기원전 450년경으로 추정되는 유대인의 역사서에서 발견되는 다음과 같은 언급이다. “이방인을 억압하지 말라, 너희도 이집트에서 이방인이었기 때문에 이방인이 된다는 것이 어떤 것인지 잘 알기 때문이다.”⁵⁾ “너의 이웃을 너 자신과 같이 사랑하라.”⁶⁾ 그리고 기원전 30-10년경 탈무드⁷⁾에 알려진 랍비 힐렐의 “스

4) Wattles (1987), p. 110.

5) *Exodus* 23: 9.

6) *Leviticus* 19:18.

7) *Sanhedrin of the Babylonian Talmud* 56a.

스로 싫어하는 것을 다른 이들에게 하지 말라, 그것이 토라의 전부이고 나머지는 모두 해설일 뿐이니 가서 배우라.”라는 등이 있다. 이러한 유대의 역사적 전통은 BC 4-27년 경 나사렛 예수의 가르침을 기록한 책에서의 “대우 받고자 하는 대로 다른 사람을 대우하라, 그것이 율법이며 예언서(또는 선지자)이다.”⁸⁾ 라는 일반화된 형식의 황금률로 표명된다.

그런데 현대의 일반화된 형식의 황금률과 예수의 사상은 형식 말고는 유사성이 거의 없다. 특히 예수의 가르침과 그 맥락을 살펴보면 각자가 말하는 내용과 관계 규정은 전혀 다르다. 이것은 현대의 일반화된 황금률과 형식이 동일하지만 그 추구하는 가치가 동일하다고 보기는 어렵다. 왜냐하면, 마태복음 5장에서 7장에 이르는 산상수훈 ‘그러므로’를 중심으로 살펴보면 이 구절은 “신에게 대우받기를 원하는 대로 다른 사람을 대우하라”로 해석되어야 하기 때문이다. 이 구절이 ‘그러므로’ (therefore)라는 결론 지시어를 통해 도출되도록 하는 산상수훈의 다른 내용들은 다음과 같다.

“불쌍히 여기는 자는 복이 있으니, 그들이 (신으로부터) 불쌍히 여김을 받을 것이다.”⁹⁾

“다른 이들을 용서하면, 신도 용서할 것이고, 용서하지 않으면 신도 용서하지 않을 것이다.”¹⁰⁾

“비관당하기 싫거든 다른 이를 비판하지 말라, 왜냐하면 신이 다른 이에 대한 너의 판단과 동일하게 너를 판단할 것이다.”¹¹⁾

그 외 산상수훈의 주요 내용들은 *Matthew* 5: 3-12, 22, 48; 6:1, 4; 7장에서 반복적으로 언급되며, 그리고 이러한 모든 내용은 *Matthew* 7: 12에서 황금률이라고 불리는 일반 형식으로 나타난다. 특히 *Matthew* 6:12에서는 “우리가 우리에게 죄지를 자를 용서한 것 같이 우리를 용서 하소서.”와 잇달아 강조되는 앞의 *Matthew* 6:14-15의 언급 즉 “너희가 다른 사람의 잘못

8) *Matthew* 7: 12; *Luke* 6: 13.

9) *Matthew* 5:7.

10) *Matthew* 6:14.

11) *Matthew* 7:1.

을 용서하면 하늘의 아버지(신)도 너희를 용서하겠지만, 너희가 다른 사람의 잘못을 용서하지 않으면, 하늘의 아버지께서도 너희를 용서하지 않을 것이다.” 등은 신과 인간의 관계에 관한 규정이다.

그렇다면 이러한 형식의 황금률은 황금률에 제기되는 거의 모든 비판들의 내용과는 무관한 것이 된다. 왜냐하면, 예수의 황금률은 사람들 사이에서의 평등성, 상호성, 합리성에 기반 한 행위를 명령하는 것이 아니라, 일차적으로 그리고 궁극적으로, 신과 인간 간의 관계에 대한 규정이기 때문이다. 사람들 간의 행위에 대한 명령들은 결국 신과의 관계 회복이 목적이며 그 수단이 황금률로 알려진 사람들에게 대한 태도(명령)이기 때문이다. 즉 유대의 황금률은 일차적으로 사람들 간의 관계에 대한 규정이라기 보다는 신과 인간의 관계에 대한 규정이다. 이렇게 볼 때만, 우리가 “아무리 하찮은 사람에게 한 것이라도 신에게 한 것이 된 셈이다. 즉, 목마른 자에게 마실 것을 주고, 배고픈 자에게 먹을 것을 주며, 헐벗은 자에게 입을 것을 주고 병든 자를 돌본 것, 다시 말해, 너희들 가운데 지극히 작은 자 하나에게 한 것이 곧 신에게 한 것이듯, 너희 형제들 중 지극히 작은 자에게 하지 않은 것은 곧 신에게 하지 않은 것이다.”는 주장이 타당한 맥락을 얻을 수 있게 된다.¹²⁾

2. 동양의 황금률: 자기중심주의의 극복과 군자의 예

황금률의 또 다른 출처는 불교의 경전이다. “자신을 사랑하는 사람은 남들도 해치지 말아야 한다.”는 붓다식의 황금률은 암스트롱에 따르면, 역설적이게도 니르바나(Nirvana), 즉 자신에 대한 부정의 정신 또는 자기중심주의의 탈피라는 독특한 종교적 가치와 관련된다. 니르바나란 ‘꺼졌다’는 의미로서 ‘타타가타’(tathagata) 즉, ‘사라짐’을 의미하기도 한다. 사성제(四聖諦)의 마지막 진리인 니르바나는 ‘오점이 없음’, ‘약해지지 않음’, ‘허물어지지 않음’, ‘침해할 수 없음’ 또는 ‘괴로움이 없는 상태’로서 ‘무’(無)를 의미하며 이는 철저한 자기중심적 자아의 부정이다. 아트만(atman)의 실재에 대한 부정인 아나타(anatta, 무아 無我)는 니르바나에 이르는 대중

12) *Matthew* 25: 35-46.

적인 명상의 방법인데, 여기서는 일반적인 자아가 부정된다. ‘자아’라는 개념은 ‘나’, ‘나의 것’에 관한 해로운 생각들을 낳으며, 자아가 앞세워질 경우 질투, 경쟁자에 대한 증오, 자만, 오만, 잔인함이 생기며, 자아가 위협을 받는다고 느낄 때면 폭력이 생긴다. 즉, 붓다는 남을 희생하면서까지 옹호하고 부풀리고 구슬리고 드높여야 하는 ‘자아’는 없다는 사실을 깨닫게 하려고 노력했다는 것이다.

기원전 551-479 경 공자의, “다른 사람이 당신에게 행하기를 원치 않는 바를 그들에게 하지 말라.”¹³⁾는 언급은 현대에는 소극적 황금률로 형식화 되어 불리고 있다. 암스트롱에 따르면, 기원전 7세기의 혼란한 상황과 소국들 사이의 끊임없는 전쟁에서, 예(禮)는 폭력을 일정한 테두리 안에 가두어 두고, 당시 사회를 절제와 자제라는 우아한 삶을 살았던 군자(君子)들의 사회를 건설하려는 노력의 일환이었다.¹⁴⁾ 예는 집단적인 복수를 쉽게 유발하는 애국주의를 억제하기 위해 고안된 것이었다.¹⁵⁾ 특히 군신, 부자간의 관계에 대한 규정들은 서(恕), 즉 ‘자기 자신에게 견줌’이라는 공감의 덕에 기초하고 있었으며, 이는 곧 공자가 말하는 황금률의 기반이 되는 개념이며 그 정신은 배려이다.¹⁶⁾ “한마디 말로서 매일 실천해야 할 것이 있습니까?”라는 자공의 물음에 대한 답으로서 서(恕)는 배려, 즉 남이 너에게 행하기를 원하지 않는 일을 남에게 하지 말라는 것이다. 이것은 늘 자신의 경험을 타인의 경험과 연결시켜, 매일 종일도록 우리 자신의 마음을 들여다보고, 우리에게 고통을 유발하는 것을 발견하고, 모든 상황에서 다른 사람들에게 그런 고통을 주는 일을 삼갈 것¹⁷⁾을 요구하는 고양된 도덕적 가치를 담고 있는 것이다. 중국 남조(南朝) 양(梁)의 주흥사(周興嗣:470~521)가 지은 것으로 알려진 사언고시(四言古詩) 250구(句)의 천자문(千字文) 중에는 복희씨와 신농씨와 같은 고대 성인들과 순(舜)과 우(禹) 등의 고대 제왕들을 칭송하는 구절들이 등장한다. 그리고 인간의 기본적인 덕목들이 뒤따라 나열

13) 己所不欲 勿施於人

14) Armstrong (2006), p. 147.

15) Armstrong (2006), p. 152.

16) 풍우란 (1993), p. 65.

17) Armstrong (2006), pp. 208-209.

되는데 특히 45번째 구(句)에 해당하는 “자신의 단점이 드러나기를 원하지 않는다면 동시에 남의 단점도 드러내지 말라¹⁸⁾” 역시 공자의 소극적 황금률의 정신을 담고 있다.

3. 황금률이 표방하는 특수한 정신과 가치

현대 도덕철학에서 논의되는 일반화된 형식의 황금률에는 구체적 행위자와 행위가 분명하게 언급되지 않는 반면 황금률의 주요 출처에서는 그 원리를 실천해야 하는 구체적인 대상과 목적 그리고 때로는 추구하는 가치들이 명료하게 언급된다. 그렇다면, 황금률이 추구하는 가치란 어떻게 이해될 수 있는가? 앞서 논의했듯이 황금률이 특수한 맥락에서의 관계들에 대한 것이기 때문에 황금률의 정신과 가치를 일반화하기는 어렵다. 우리가 확인할 수 있는 것은 서로 다른 가치들이 각자의 황금률을 통해서 표방된다는 것이다.

그리스의 철학에서는 언급되는 황금률의 원리는 모두 *탁월함(덕)*을 성취하는 방법으로 제시되는 것들이다. 그런데 의무론과 결과론으로 대변되는 현대 도덕철학에서 탁월함 또는 덕은 여전히 그 도덕적 가치로서의 지위가 완전하게 정립되지 않고 있다. 즉 그리스의 황금률에서 요구되는 탁월함은 합리성, 상호성 그리고 평등주의로 대변되는 현대 도덕철학에서는 고유한 도덕적 의무의 영역을 차지하지 못하고 있는 것이다.

동양에서는 약간 다른 정신이 황금률을 통해 드러난다. 자이나교의 가르침인 “수도자는 모든 존재자들을 자신이 대우받기를 원하는 대로 대우해야 한다.”는 말은 *비폭력*, *공감*, 그리고 *삶의 신성함*을 강조한다. 기원전 563-483년 경으로 알려진 붓다의 황금률 “당신에게 소중한 것은 다른 사람에게도 소중한 것이니 당신에게 고통스러운 것으로 다른 이를 고통스럽게 하지 말라.”¹⁹⁾와 고대 힌두교의 가르침인 “모든 창조물을 자신처럼 생각하고 마치 자신에게 하듯 그들에게 하는 사람”은 아힘사(ahimsā) 즉, 모든 *피*

18) 罔談彼短

19) *Dhammapada, Northern Canon*, 5: 18.

조물들에 대한 사랑의 원리를 표방한다. 그런데, 공감, 삶의 신성함, 그리고 아힘사에서 표방되는 불살생의 정신은 현대 도덕철학의 가치들을 초월하는 초과의무적인 행위들이다. 현대도덕철학에서 비공감이나 아힘사 정신 등은 바람직한 것일 수는 있지만 도덕적으로 요구되는 것은 아닌 것으로 여겨지기 때문이다.

암스트롱(Armstrong, 2006)에 따르면, 기원전 900여 년 전부터 200년 사이의 세계의 네 지역에서 탄생한 위대한 전통들, 붓다, 공자, 그리고 그리스의 철학적 합리주의와 유대의 전통-에서 현자들은 각각 그 나름의 방식으로 황금률을 정리해냈다. 이 후 발전한 유교와 도교, 힌두교와 불교, 그리고 유대교, 기독교, 이슬람교 등은 모두 ‘축의 시대’에 발전한 전통이 나중에 개화한 것이다. 이들은 모두 비폭력에 기초한 영성을 장려하여 자기 시대의 폭력적 잔혹함과 공격성에 맞서려 했다. 이 시대는 모두 전쟁과 폭력의 공포, 그리고 절망과 좌절이 지배하던 시대였으며, 이 시대의 현자들은 폭력의 악순환을 막기 위한 대응으로서 자비의 윤리를 발전시켰고, 우리 안의 폭력의 원인인 자기중심주의와 적대감 그리고 증오를 체계적으로 억제하기 위해 노력했다. 그들은 서로 다른 경로를 거쳤음에도 깊은 수준에서는 서로 비슷한 해결책을 제시했는데, 그것이 바로 황금률로 불리는 공감의 윤리이다.²⁰⁾ 따라서, 그들에게 황금률이란 가장 숭고한 도덕이자 종교적 신앙의 대상으로서 상호주의, 평등주의, 합리주의라는 현대 도덕철학의 주요 개념들과는 상당히 다른 가치를 구현하고 있다.²¹⁾

황금률, 유대인에게는 그것이 율법이자 예언서이고 예수에게는 사랑이었다. 붓다에게는 자기초월의 요구였고 그리스 철학에서는 덕이자 옳음, 선

20) Armstrong (2006), pp. 390-392.

21) 칸트가 언급하는 황금률도 이러한 특수 관계에 대한 것임을 알 수 있다. 세베루스의 황금률이 명령하는 대상은 종교들 간의 갈등 상황에서의 개별 종교인들이다. 칸트의 비판의 대상이 되는 “Quod tibi non vis fieri, alteri ne feceris.”는 로마의 황제 알렉산더 세베루스가 종교들 간의 갈등을 막고 평화를 증진하기 위해 공공장소에 게시한 것으로, 그것이 황제의 황금벽에 씌어졌기 때문에 ‘황금률’이라고 불리게 되었다고 전해진다 (Gensler, 2013, p. 80). 칸트가 비판한 황금률이란 결국 로마 황제가 종교 간의 평화를 위해 도입한 하나의 정치적 수단이라는 점, 즉 그것이 여러 문화권에서 다른 가치를 위해 표명된 여러 황금률의 버전들 가운데 특수한 하나라는 것이다.

함, 또는 공정함이었다. 자이나교에서는 수도자들의 삶을 규정하는 것이었고, 조로아스터에게는 성격적 자질이였다. 중국의 전통철학에서 황금률은 군자의 도덕인 인, 즉 자애로움을 의미한다. “다른 사람들이 당신에게 행하기를 원하는 것을 그들에게 하라.”로 형식화되는 형식의 황금률은 그 자체로는 황금률의 연대기에서 거의 찾아 볼 수 없을 정도로 일반화된 것이다. 지금까지 논의한 것처럼 황금률의 주요출처들에서의 언급들을 참고로 할 때 이러한 일반화는 지나친 것이다. 뿐만 아니라 이렇게 문화, 종교, 공동체의 도덕등과 무관한 방식으로 일반화된 형식의 황금률로는 원래 출처들에서 강조되는 가치와 정신을 표현할 수 없으며, 그러한 맥락으로부터의 괴리가 곧 황금률의 독특한 정신을 훼손할 수 있다. 다음 장에서는 일반화된 형식의 황금률을 둘러싼 비판과 옹호의 노력에서 어떻게 황금률의 특수한 정신들이 훼손되고 있는지 논의한다.

III. 황금률에 대한 전통적 비판

1. 칸트의 비판

칸트는 “다른 사람들이 너에게 행하기를 원치 않는 것을 다른 사람들에게 하지 말라.”는 소극적 황금률은 도덕의 기반이 되는 기준이나 원리일 수 없다고 말하는데, 왜냐하면, 소극적 황금률은 자신에 대한 의무(예를 들어 자살금지)나 타인에 대한 의무도 담지하지 않고 있을 뿐만 아니라, 상호간의 의무도 담지하지 못하기 때문이다.²²⁾ 예를 들어, 칸트는 소극적 황금률을 “타인이 너를 돕기를 원하지 않으면 너도 타인을 돕지 말라.”는 것과 같이 해석하면서, 이것이 타인을 도울 의무를 부정하고 있다고 평가한다. 한편, 칸트는 “나의 인격에서나 다른 모든 사람의 인격에서 인간을 항상 동시에 목적으로 대하고, 결코 한낱 수단으로 대하지 않도록, 그렇게 행위 하

22) Kant (1959), p. 48.

라.”라는 최상의 실천명령을 제시한다. 그리고 칸트의 유명한 ‘판사와 범죄자의 사례’는 황금률이 인격을 수단화하는 것을 보여준다. 범죄자는 소극적 황금률을 이용하여 판사에게 다음과 같이 호소할 것이다. “타인이 당신을 처벌하기를 원치 않는다면 당신도 타인을 처벌하지 말라.” 이렇게 되면 범죄자는 판사를 오직 수단으로서만 대할 것이다. 따라서 황금률은 명백한 도덕적 의무를 규정하고 있지 않을 뿐만 아니라 인간을 수단으로 여기기 때문에 최상위 실천명령을 위반하며 또한 정언명령이 아니라 가언명령이기 때문에 보편법칙으로서 자격이 없다는 것이다.²³⁾ 전통적으로 황금률에 대한 비판들은 행위자 중심적 해석과 수용자 지향적이라는 두 범주로 구분되는데, 칸트의 비판은 이러한 두 범주 모두에 기반하고 있다.²⁴⁾

2. 행위자 중심적 또는 수용자 지향적 황금률

바이스(Weiss, 1941)에 따르면, 황금률은 다음의 조건에서만 합당한 윤리원리가 될 수 있다.

- 우리는 우리가 원하는 것을 안다.
- 우리가 원하는 것은 우리가 욕구해야 하는 것과 일치한다.
- 우리에게 좋은 것은 다른 이들에게도 좋은 것이다.

이러한 조건들이 만족되지 않는 경우 황금률은 행위자 중심적 또는 수용자 지향적 특징 때문에 선보다는 악을 명령할 수 있을 것이다. 이러한 황금

23) “여기서 ‘사람들이 네게 함을 바라지 않는 것을 남에게 하지 말라’는 등의 상투적인 말이 준거나 원리로 쓰일 수 있을 것이라는 생각은 하지 말아야 할 일이다. 왜냐하면 이런 말은, 비록 여러 가지 제한과 함께 이기는 하겠지만, 단지 저 원칙으로부터 파생된 것일 뿐이기 때문이다. 그것은 보편적 법칙일 수가 없다. 왜냐하면 그것은 자기 자신에 대한 의무들의 근거도, 타인에 대한 사랑의 근거도 함유하고 있지 않으며, (타인들에게 자선을 베푸는 일을 그가 하지 않아도 된다면, 타인들이 그에게 자선을 행하지 않는 것에 많은 사람들이 기꺼이 동의할 터이다) 드디어는 서로 간의 당연한 의무들의 근거도 함유하고 있지 않기 때문이다.” 칸트 『윤리형이상학 정초』, 백종현 옮김, p. 150.

24) Reinikainen (2005), p. 157.

률의 특징으로 발생하는 문제는 다음과 같다. 첫째, 다양한 견해와 취향 그리고 욕구들이 존재하는 복합적인 사회에서 개인들이 타인들에게 어떻게 행위 해야 할지를 고려할 때, 황금률은 행위자 자신의 견해와 욕구들을 참고해야 한다는 지나치게 강한 요구를 하는 것으로 보인다.²⁵⁾ 즉, 만일 우리가 황금률에 따라 행위 하면, 다른 사람에 대한 우리 행위의 도덕적 적합성을 위한 기준은 우리 자신에게 있게 된다.²⁶⁾ 이렇게 보면, 일반화된 형식의 황금률은 행위자 중심적 또는 수용자 지향적으로 해석될 수 있는 여지가 있으며, 따라서 이해관심이 상대적으로 동질적이고 단순한 사회에서 잘 작동할 것이다. 그런데 수용자로서의 자신의 욕구와 희망을 투사할 때 그것이 실제 수용자의 욕구와 희망과 다를 수 있다. 그래서 황금률은 통일된 인간 본성이라는 믿음에 근거할 수밖에 없다는 것이다.²⁷⁾ 이런 근거에서 쇼우(B. Shaw)는 “다른 사람이 원할 것으로 믿는 것을 행하지 말라, 왜냐하면 그들의 취향은 다를 수 있기 때문”이라고 주장하기에 이른다.²⁸⁾ 둘째, 수용자로서 행위자의 희망은 비도덕적일 수 있어서, 행위자와 수용자의 희망이라는 기준은 때로 정당한 사회적, 법적, 경제적 규범들에 반하는 행동을 요구할 수 있다는 것이다. 시즈윅(Sidgwick)이 말한 것처럼 “사람들은 범죄에서 다른 사람의 협조를 희망할 수 있고 기꺼이 협조할 것이다.”²⁹⁾ 이렇게 해석된 황금률은 심지어 이타적인 원리조차도 아니게 된다. 이것은, 예를 들어, 욕구를 알 수 없는 상대방과의 관계에서 드러나는데, 우리는 동물들과 아이들 또는 정신적으로 결함이 있는 사람들에 대해서 그들이 우리에게 원하는 것처럼 그들이 우리에게 행하기를 기대할 수 없다.

25) Russell, (1942), pp. 109-110.

26) 적극적 황금률은 행위의 도덕적 적합성을 위한 기준을 행위자 자신의 선호에 두기 때문에 여타의 메탈룰(metal rules) 특히, 소극적 황금률과 소위 ‘황동룰’(cooper rule)보다 열등하다고 주장된다. (Huang 2005, pp. 394-425) 참조.

27) Lippmann (1922), pp. 121-2.

28) Bernard Shaw (1930), p. 217.

29) Sidgwick (1907), p. 380.

IV. 전통적 비판으로부터 황금률을 구하려는 시도들과 그 한계

1. 싱어의 일반해석

싱어(M. Singer, 1963)는 황금률이 인간의 본성의 동질성을 가정한다면 그야말로 불합리한 것이 되겠지만, 사실은 황금률이 그러한 가정에 기반하지 않는다는 것을 보이고자 한다. 싱어에 따르면 이러한 오해는 기본적으로 황금률을 추자적으로 해석함으로써 황금률의 정신을 잘못 이해하는데 기인한다. 싱어에 따르면, 다음의 두 진술은 구분되어야 한다.

- 특수해석: “다른 이들이 당신에게 해주기를 원하는 것을 그들에게 하라.”
- 일반해석: “다른 이들이 당신에게 해주기를 원하는 것과 같이 그들에게 하라.”

황금률은 예를 들어 상대방이 자신에게 행하기를 **원하는 것이 아니라 원하는 것과 같이** 행하기를 요구한다. 이 말은 상대방이 적용할 행위 원리와 동일한 원리에 기반 하여 상대방을 대하라는 것을 의미 한다.³⁰⁾ 따라서 황금률은, 예를 들어, 대지 임대인이 임차인에게 무료로 땅을 임대하기를 명령하는 것이 아니다. 즉 이타적인 행위가 아니라 동일한 기준에 따라 행위 하라는 것이다. 이것은 임차인의 입장에서 생각해보아도 마찬가지이다. 임차인이 황금률의 특수해석을 따른다면 임대인이 원하는 만큼의 (아마도 더 높은, 두 배 내지는 세배) 임대료를 주려고 할 것이다. 반면, 일반해석에 따르면 황금률은 평등성이라는 정의원리의 정신을 요구 한다. 결국 황금률의 특수해석은 불합리하게 되고, 일반해석이 합리적이게 된다. 그러나 싱어 스스로 말하듯이 황금률이 그 기준이나 원리의 구체적 내용이 어떤 것이어야 하는지 명료하게 말하는 것은 아니다. 즉 황금률은 동일한 기준이 적용되기를 명령하는 것이지만, 그 기준 자체를 규정하지는 않는다.³¹⁾ 그렇다면, 황

30) Singer (1963), p. 310.

금률은 동일하지 않은 취향, 이해관심, 욕구 등과 양립가능하며, 인간 본성의 동질화에 근거하지 않는다는 것이 입증될 수 있다. 따라서 황금률의 특수해석에는 가능한 이러한 비판이 일반해석에는 가능하지 않게 된다.

결국, 싱어에 따르면, 황금률에 제기된 문제점은 황금률의 진짜 문제가 아니라 황금률에 대한 잘못된 해석으로 생긴 문제라는 것이다. 우리가 황금률을 도덕원리로서 받아들지 않고 구체적 상황에서 특정 행위의 옳고 그름을 판단하는 도덕원칙으로 받아들이는, 즉 황금률이 각 행위에 대한 도덕판단을 해주기를 원하는 기대로 인하여 특수해석을 요구하는 문제가 발생하게 된다는 것이다. 싱어에 따르면, “다른 사람들과의 관계에서 그들이 당신에게 적용하기를 원하는 동일한 원리나 기준에 근거하여 행위 하라.”라고 기술 될 수 있는 황금률의 일반해석은 정의의 근본적인 요구인 “모든 사람들의 행위는 동일한 기준에 의거하여 판단되어야 하며, 어떤 사람도 특별하거나 특권적인 입장을 주장할 수 없다.”는 것을 형식화한 것이다.

이렇게 보면 싱어의 일반해석은 ‘불편부당한 관망자’의 견해에서 판단하는 것과 크게 다르지 않은 것으로 보인다. 즉 각 행위자는 자신을 구체적 상황으로부터 추상화하여 어떻게 판단할지를 결정해야 한다.³²⁾ 그렇다면, 일반해석은 다음과 같이 형식화 될 수 있다. “불편부당한 관망자가 당신에게 하기를 원하는 것처럼 다른 사람들에게 하라.” 그렇다면 불편부당한 관망자는 우리 서로가 무엇을 행하기를 원하는가? 구체적 행위를 명령하는 고대 황금률과 달리 일반해석의 황금률은 특정한 행위를 명령하지 않는다. 이런 점에서 이것은 우리의 일상적인 도덕적 결정들로부터 동떨어진 것으로 그리고 그 주창자들, 즉 주요 종교들의 해석과 매우 다른 것으로 믿게 만든다.³³⁾ 물론, 행위지침을 주지 않는 것이 일반해석의 치명적인 문제라고 보기도 어렵다는 입장이 있을 수 있다. 예를 들어, 블랙스톤(Blackstone)은 싱어의 일반해석을 지지하며 황금률이 메타도덕률, 즉 규칙들을 형식화하는 메타도덕률이기 때문에 우리는 그것이 실제적인 도덕적 문제들을 해결하

31) Singer (1963), p. 313.

32) Singer (1963), p. 302.

33) Gould (1983), p. 74.

기를 기대해서는 안 된다고 주장한다.³⁴⁾ 그런데 블랙스톤의 이러한 옹호 역시 고대 황금률에 나타난 구체적 행위에 대한 규정, 그리고 그러한 행위 규정에 반영된 도덕적 가치들은 드러내지 못한다.

2. 거위드의 합리화된 황금률

거위드(Gewirth)는 기존의 황금률에 대한 오해가 모두 황금률에 대한 추자적 해석에서 비롯된다고 주장한다. 즉 거위드는 황금률에 담긴 정신에 대한 이해 없이 문자 그대로의 황금률을 도덕원칙으로 볼 때의 문제점을 지적하고 있는 것이다. 특히, 상어의 일반해석에서 구체적인 행위를 원리로 대체한다고 하더라도, 어쨌든 행위자가 선택한 원리 즉 행위자가 선호하거나 희망하는 것이 수용자에게 부과된다는 점에서 여전히 황금률에 대한 비판을 피할 수 없다는 것이 거위드의 생각이다. 거위드에 따르면, 상어와 같이 황금률을 상호주의와 평등주의적으로 해석하는 통상적 해석(generic interpretation)이 문제를 근본적으로 해결해주는 것은 아니다. 왜냐하면 첫째, 이 해석은 행위자와 수용자의 선호가 충돌하는 경우 어떤 행위지침을 주지 못하고 둘째, 고려해야 하는 모든 사람들의 집합과 그 고려의 내용 즉 선호의 종류들에 대해서 확정하지 않기 때문에 자칫 잘못된 선호들까지도 고려될 수 있기 때문이다. 따라서 황금률이 구체되기 위해서는 옳음에 관한 기준이 잠재적으로 자의적인 여러 욕구들로부터 구분될 수 있어야 한다는 것이다.³⁵⁾ 그렇다면, 거위드가 말하는 황금률에 담긴 정신은 무엇인가? 거위드의 해법은 황금률이 명령하는 욕구들은 모두 합리적이어야 한다는 것이다. 따라서 황금률은 다음과 같이 수정되어야 한다. “상대방이 당신에게 행하기를 합리적으로 원하는 대로 상대방에게 하라.” 이렇게 황금률은 합리화 즉, ‘합리화된 황금률’로 변모한다. 그렇다면 상호성과 평등성에 기반한 포괄적 해석은 다음과 같이 될 것이다. “당신의 수용자의 합리적인 욕구들과 당신의 욕구에 부합하게 행위 하라.” 이렇게 합리화된 황금률의 내용

34) Blackstone, (1965), p. 172.

35) Gewirth (1978), pp. 526-527.

은 통상적 권리들로 채워지고, 그러한 권리들은 주로 자유나 복지를 위한 권리로 구성된다.

거워드에 따르면, 자유나 복지를 위한 권리들이 존재한다는 것으로부터 그 권리를 침해하지 말아야 한다는 도덕적 의무가 도출된다. 이런 점에서 거워드의 황금률 해석은 ‘사실-가치’의 구분에 대한 자신의 독특한 이론을 배경으로 하고 있다. 거워드에 따르면, 합리화된 황금률은 전통적인 황금률의 문제들을 대부분 극복할 수 있다. 그렇다면 자유와 복지가 서로 충돌할 때는 어떤가? 거워드는 그러한 사례가 많지도 않을 뿐만 아니라 우선순위가 결정될 수 있다는 근거로 별 문제가 되지 않을 것이라고 간단히 언급만 하고 지나간다. 그러나 행위자와 수용자의 자유와 복지가 충돌할 때 그렇게 간단히 문제가 해결될 수 있는지는 의문이다. 그러한 사례가 많지 않다고 해서 별 문제가 아니라고 할 수도 없을 뿐만 아니라, 동일한 권리(자유-자유, 또는 복지-복지)가 충돌할 때 어떻게 우선순위를 결정할 수 있다는 것인지는 검토해보지 않는다.

거워드에 따르면 황금률은 평등주의적 특징을 가지는데, 그것은 행위자가 자신의 자유와 행복을 수용자의 행복을 위해 포기하기를 요구하지 않는다는 점에서 그렇다.³⁶⁾ 그런데, 거워드 스스로 말하듯이, 우리는 그러한 사람들을 좋은 사람들 또는 매우 도덕적인 사람으로 평가하지 무도덕적이거나 비도덕적인 사람으로 평가하지는 않는다.³⁷⁾ 그러나 합리화된 황금률은 이러한 이타적인 행위들이 최소한 무도덕적이거나 비도덕적인 것으로 평가될 여지를 분명히 남기고 있다. 결국 평등한 관계에서 기대되는 합리적인 도덕적 의무를 초과한 초과무적 행위들은 불필요하거나 무가치한 것이 될 여지가 남는데, 문제는 황금률의 고유한 정신과 가치들이 모두 평등성과 합리성을 초월하는 것이라는 점이다. 이러한 초과무적 가치들이 평등성과 합리성에 기반 한 현대 도덕철학에서 어떻게 비합리적이고 무가치한 것이 되는지는 다음 장을 통해서 논의한다.

36) Gewirth (1978), p. 140.

37) Huang (2005), p. 399.

3. 메타 원리로서의 황금률

블랙스톤(Blackstone, 1965)은 칸트의 보편화 원리와 흄의 불편부당한 합리적 관망자 모델 그리고 황금률 등은 모두 도덕적으로 중립적인 도덕에 관한 견해들이라고 주장한다. 그에 따르면, 황금률에 관한 다음의 비판들은 모두 황금률을 규칙들을 형성하는 메타규칙이 아니라 규범적 맥락에서의 도덕률로 보는데서 비롯된 혼동이다. 블랙스톤은 황금률이 구체적인 행위의 옳고 그름을 판단하는 도덕규칙이 아니라고 주장한다. 그에 따르면, 황금률의 기능은 오히려 도덕판단을 다른 여타의 판단으로부터 구별해주도록 하는 형식적 시험의 기능을 한다. 즉 황금률에 따르면, 행위자의 특권을 부정하고 다른 사람들이 적용하기를 기대하는 기준 또는 규칙이나 원리들과 동일한 것을 적용하는 어떤 판단이든 도덕적 판단일 수 있다. 이것은 황금률에 순응하는 것이 도덕적 판단이 되기 위한 필요조건이 된다는 것을 의미한다.³⁸⁾ 블랙스톤은 흄의 불편부당한 합리적 관망자나 칸트의 보편화 원리 등이 모두 이러한 메타 도덕적 특징을 가진다고 본다. 이 모든 원리들이 불편부당성이라는 도덕의 특징을 기반으로 하고 있다. 그는 이것을 ‘도덕적 관점’이라고 보는데, 도덕적 관점에서 보더라도 모든 사람들이 동일한 도덕적 결론에 이르거나 도덕적 확신이나 규범적 윤리에 귀결하지는 않는다고 말한다. 따라서 황금률에 전통적 비판들이 적절하지 않다는 것이다.

라이니카이넨 (Reinikainen, 2005)에 따르면, 황금률과 보편화의 원리 모두 공공성 테스트 즉 영향을 받는 모든 사람들의 조망으로부터 받아들여질 수 있는 것을 행하라는 요구로 귀결된다. 따라서 황금률이 보편화의 원리보다 열등한 원리라고 말할 수 없다.³⁹⁾ 보편화의 요구는 우리가 제안하는 것이 모든 관련된 유사한 경우들이나 상황들에서 적용 가능한 것인지에 대한 평가이다. 이것은 역지사지의 요구, 즉 서로 간의 입장이 바뀌었을 때 적용 가능한 것인지에 대한 평가와 공지성의 요구 즉, 확장된 역지사지의 원리로서, 영향을 받는 모든 사람들의 조망이 고려되는 것이다.

38) Blackstone (1965), pp. 172-173.

39) Reinikainen (2005), p. 155.

이렇게 해석된 황금률의 정신은 황금률에 대한 칸트의 비판을 재평가하기를 요구한다. 라이니카이넨 (Reinikainen, 2005)에 따르면, 황금률은 그것의 적용에 대해서도 규정하는 메타테스트를 요구하고 있다. 황금률은 스스로가 사용되는 방식에 관한 것까지 규정하고 있는데, 그것은 황금률을 적용할 때 그러한 적용이 행위자 자신에게도 받아들여질 수 있어야만 한다는 것이다. 즉 황금률을 사용하는 것 자체가 황금률의 원리에 따라야 한다는 것이다.⁴⁰⁾ 이렇게 보면, 범죄자가 황금률을 사용하여 처벌을 면하려고 하는 것처럼 판사도 황금률을 사용하여 공정하게 처벌하려고 할 것이고, 이것은 범죄자의 황금률이 황금률을 위반하는 방식으로 사용되고 있다는 것을 보여준다. 황금률은 그 스스로가 행위를 인도하는 법칙으로서 사용되어야 하는 방법에도 적용되기 때문에, 우리가 대우받기를 원하는 대로 대우해야 한다는 황금률에 기반 한 판단은 다른 사람들의 행동에 대해 어떻게 대응해야 하는지에 대한 판단까지도 포함하고 있다. 메타테스트로서 황금률은 상대방이 우리를 대할 때 사용하는 규칙을 받아들일 수 있는 방식으로 그 규칙을 사용하기를 요구하는 것이다. 결국 황금률은 황금률을 어기지 않는 방식으로 사용될 때 비로소 받아들여질 수 있다. 그렇지 않다면 황금률이 아니라 황금률의 적용 방식이 잘못된 것이다.

황금률에 제기되는 여러 도덕적 비판들을 극복하기 위해 황금률을 구하고자 하는 많은 시도들은 대체로 유사한 전략을 채택한다. 그것은 황금률을 구체적인 도덕원리가 아닌 도덕에 관한 한 견해로 보는 것이다.⁴¹⁾ 한편으로, 황금률을 그것이 적용되는 방식을 규정하는 메타원리로 해석하는 이러한 전략들은 일견 타당해보일 뿐만 아니라 전통적으로 제기되어온 비판들을 효과적으로 피해갈 수 있는 것처럼 보인다.

그러나 이러한 전략은 황금률을 도덕원리로 보는 것이 아니라 도덕에 관한 한 견해로 볼 때만 효과가 있다. 즉 메타원리로서의 황금률은 황금률이 적용되는 방식에 관한 것이지 도덕원리로서의 황금률에 관한 것이 아니다. 블랙스

40) Reinikainen (2005), p. 159.

41) Blackstone (1965), p. 174. 블랙스톤에 따르면 Gewirth (1978)와 Singer (1963) 모두 이와 같은 합리적 관망자 모델의 일종이다.

톤이 말하듯이, 황금률이 규범적 맥락에서의 도덕률이 아니라 규칙들을 형성하는 메타규칙이어서 구체적인 행위의 옳고 그름을 판단하는 도덕규칙이 아니고, 또한 황금률의 기능이 단지 도덕판단을 다른 여타의 판단으로부터 구별해 주도록 하는 형식적 시험의 기능만을 수행한다면 황금률에 무슨 특별한 도덕적 가치가 있는지 의심스럽다. 만일 메타원리로서의 황금률이 불편부당한 관망자 모델, 또는 보편화의 요구 등과 전혀 차별화되지 못한다면, 황금률은 단지 우리가 도덕을 어떻게 적용해야 하는지에 대한, 즉 도덕원리가 아니라 도덕이 적용되어야 하는 방식에 대한 한 견해일 뿐이고, 도덕철학에서 그러한 견해들은 충분할 정도로 합의가 이루어진 상태이다. 결국, 메타원리로서 황금률은 전통적 비판으로부터 자유로워질 수는 있지만 고대 황금률이 가진 그 자체의 고유한 도덕적 규범성뿐만 아니라 초과무적 가치들을 상실한다.

V. 현대 황금률의 도덕철학적 문제들

1. 평등화, 상호주의화 그리고 합리화 된 황금률

칸트의 비판으로부터 황금률을 구하고자 하는 대표적인 논의들은 주로, 평등적 정의, 합리성, 그리고 메타테스트로서 일종의 상호주의에 기반하고 있다. 바이스(Weiss)는 칸트의 비판을 약간 순화시켜 황금률과의 화해를 꾀한다. “너 자신이나 다른 모든 이들의 인격에서 인간성을 항상 동시에 목적으로 대하고, 결코 한낱 수단으로 대하지 않도록, 그렇게 행위 하라.”⁴²⁾는 실천명령에 따른 해석에 의하면 황금률은 다음과 같은 명령이 된다. “너 자신이나 다른 합리적 존재를 단지 수단으로만 대하지 말고, 모든 경우에 있어서, 항상 동시에, 그 자체로 목적으로서 대하라.” 바이스에 따르면 이것은 황금률에 대한 하나의 좋은 해석이다. 그러나 황금률은 정언명령적이지 않다. 왜냐하면, 인간은 공동체를 이루어 서로에게 의존하고 서로를 수단으

42) Kant (1959), p. 47.

로하여 생존하기 때문이다. 따라서 황금률과 칸트의 정언명령의 차이는 단지 수단화의 정도와 방법에 관한 것이다. 그 차이는 황금률은 수단화의 정도와 방법에 있어서의 동등함을 요구한다는 것이다.⁴³⁾ 이렇게 본다면, 황금률은 이타적인 사랑의 윤리를 구체화하지 못할 뿐만 아니라 기껏해야 행위자들 간의 도덕적 동등성이나 또는 상호성만을 주장하는 것이라는 후앙(Huang)의 비판이 타당하게 제기될 수 있다.⁴⁴⁾

이것은 노만(Norman)이 말하는 증대한 자기희생을 치른다고 하더라도 타인을 사랑하라는 ‘사랑의 원리’를 충족하지도 못한다.⁴⁵⁾ 더구나 이러한 황금률은 원수까지도 자비의 영역에 포함시킬 정도로 충분히 확장되지 않음으로써 황금률의 메시지를 정확히 해석해주지 못한다. 리콰르는 예수가 황금률을 말하는 맥락으로부터 황금률이 자신을 초월하는 즉 초도덕 또는 초월적범이어야 한다는 해석을 가하는데, 그 맥락이란 다름 아닌 원수까지도 사랑하는 것이다. 그런데 황금률은 행위자의 행위가 일종의 보상을 위한 것이라는 점에서 평등성에 불과한 것이다. 리콰르(Ricoeur)에 따르면, 이런 황금률은 평등성이나 상호성과 같은 논리에 머물러 있을 뿐이기에 보복률(Lex talion)보다 약간 더 나은 것일 뿐이다.⁴⁶⁾ 그러나 리콰르의 이러한 평가는 지나치게 후한 것이다. 왜냐하면 황금률이 평등성이나 상호성과 같은 논리에 머물러 있다면 보복률 보다 약간 더 나은 것이 아니라 보복률 그 자체가 되고 말기 때문이다. 다음 절에서는 왜 일반화된 형식의 황금률이 보복률보다 결코 더 낫지 않은지를 논의한다.

2. 현대 황금률의 도덕철학적 위험성

일반적 형식으로 표현되는 황금률에, 앞에서 살펴본 평등주의와 합리주의라는 조건적 해석이 가해질 때 황금률은 심각한 문제점을 드러내게 된다. 예를 들어, 싱어는 적극적 황금률과 소극적 황금률이 다르지 않다는 것을

43) Weiss (1941), p. 430.

44) Huang (2005), p. 408.

45) Norman (1969), pp. 157-158.

46) Ricoeur (1990), pp. 352-397.

다음과 같이 증명한다.⁴⁷⁾ 두 버전이 동일한 명령을 하는 것으로 해석될 수 있고, 따라서 결국 행위에 미치는 효과가 동일하기 때문에, 감정적으로는 구분될지 몰라도, 그 논리적 구조와 도덕적 효과는 동일하다. 즉, 두 버전은 수사법적이고 심리적인 차이만 있을 뿐 논리적이고 도덕적인 차이가 없다는 것이다. 싱어는 황금률이 무엇을 할지 그리고 무엇을 하지 말 것인지를 명령하는 것으로 본다. 그런데, 예를 들어, “거짓말 하지 말라.”는 것은 “진실을 말하라.”라는 것과 동등한 것을 말하는 것으로 볼 수 있다. 이와 마찬가지로, 유비에 의해서, 황금률의 두 버전이 동등하지 않아 보임에도 불구하고 같은 것으로 볼 수 있다는 것이다. 유비를 확장하면 진실을 말하라는 것은 우리가 무엇을 할지에 대해 긍정적으로 명령하지만 어떤 것에 대해 침묵을 허용하지 않는 것처럼 보인다. 반면, 거짓을 말하지 말라는 것은 침묵을 허용하는 것으로 보인다. 그러나 이러한 차이는 모두 피상적인 것이다. 싱어에 따르면, “진실을 말하라.”는 “무엇인가를 말한다면, 진실을 말하라.”로 해석되어야 할 것이다.⁴⁸⁾

싱어는 이러한 방식을 유비에 의해 확장하여, 어떤 것을 원하지 않는다는 것은 그 반대의 것을 원한다는 긍정적인 형식으로 재 기술 될 수 있다고 주장한다. 예를 들어, “다른 사람들이 너에게 거짓말하기를 원하지 않는다면,”은 “다른 사람들이 너에게 거짓말을 안 하는 것을 원한다면”으로 재 기술 될 수 있다. 이것을 일반화하면 다음과 같다.

“‘A는 X가 발생하는 것을 원하지 않는다.’는 ‘A는 X가 발생하지 않는 것을 원한다.’와 같고 ‘A는 X가 발생하기를 원한다.’는 ‘A는 X가 발생하지 않는 것을 원하지 않는다.’와 같다.”

따라서 황금률의 긍정적, 부정적 두 버전에는 차이가 없다는 것이다.

그런데 일반적 형식의 황금률이 이렇게 조건화되면 그 논리적 귀결은 매우 불합리해진다. “진실을 말하라.”는 정언적 명제가 “무엇인가를 말할 때

47) Reinikainen (2005)도 싱어의 논증과 유사한 방식으로 두 버전의 차이를 부정한다.

48) Singer (1963), pp. 304 - 305.

는, 진실을 말하라.”는 조건명제로 해석되듯이 일반형식으로 표현된 황금률의 두 버전도 조건명제로 해석될 때 동등한 것으로 취급될 수 있는가? 그렇다면 그 논리적 귀결은 무엇인가? 형식화된 황금률의 두 버전을 조건명제로 환원하면 일반적 형식의 황금률이 무시무시한 명령이 되고야 만다.

- 적극적 버전의 조건명제: “다른 이들이 너에게 행하기를 원한다면, 그들에게 동일하게 행하라.”
- 소극적 버전의 조건명제: “다른 이들에게 당하기 싫다면, 그들에게도 그렇게 하지 말라.”
- 적극적 버전의 대우명제: “다른 이들에게 대접받고자 하는 대로 행하지 않았다면, 다른 이들에게 동일한 대접받기를 원하지 말라.”
- 소극적 버전의 대우명제: “다른 이들에게 대접받기 원하지 않은 대로 행했다면, 동일한 대접을 받기를 원하라 (또는 기피하지 말라).”

이러한 조건화는 황금률에 대한 칸트의 비판에서 이미 나타난다. “상대방이 자신에게 행하기를 원하는 대로 상대방에게 행하라.”는 적극적 황금률은 조건적인 원리가 되고, 이 말은 곧 “상대방에게 행하지 않는 행위는 자신도 받기를 원하지 말라.”는 것이 된다. 예를 들어, 타인을 돕기를 원치 않은 사람은 타인으로부터 도움을 받기를 원하지 말아야 한다는 것이다. 또한 이 말은, “타인에게 도움을 받기를 원한다면 그들에게 도움을 주라.”는 가언명제가 된다. 따라서 칸트가 보기에 황금률이 위반하는 것은 보편화가가능성 조건이 아니라 정언명령의 조건이다. 이것은 또한 자신에 대한 의무를 부과하지 않는다는 첫째 비판의 근거가 된다. 자신이 도움을 받기 원치 않으므로 타인을 돕지 않는다는 원리는 자신을 돌볼 의무와 타인을 돌볼 의무를 동시에 부과하지 않는 것이며, 따라서 상호간의 의무를 부과하지 않는다는 것이다.

단지 도덕적 의무를 부과하지 않기 때문에 또는 조건적이라는 이유만으로 일반적 형식의 황금률이 문제가 있다고 평가할 수는 없을 것이다. 일반적 형식의 황금률의 중대한 문제는 소극적 황금률의 대우명제 때문이고, 이러한 문제는 더 이상 이것을 소위 ‘황금률’이라고 부를 수나 있는지 의심스럽게 만든다. “다른 이들에게 대접받기 싫은 것을 그들에게 행하지 말라.”

는 소극적 황금률에 따르면, 만일 당신이 상대방에게 어떤 행위를 했다면, 상대방이 그와 같은 행위를 당신에게 하는 것을 원해야 한다 (또는 최소한 기피하지 말아야 한다). 이것의 한 예는 다음과 같을 것이다. 살인자는 소극적 황금률에 따라, 동일한 대접을 받기를 기피할 수 없다. 즉 살인자는 자신이 저지른 살인에 대해 동일한 대접을 받는 것을 거부할 수 없다. 이렇게 되면 소극적 황금률은 “보복을 보다 약간 나은 것”이 아니라, 오히려 보복을 그 자체가 되고 만다. 이것이 곤란한 문제를 일으키는 이유는 적극적 황금률의 주요 출처에서 표방되는 정신뿐만 아니라, 소극적 황금률의 최초 출처인 공자의 가르침과도 불일치하기 때문이다. 특히, 이러한 가르침은 주요 종교에서 표방되는 가치들인 사랑, 증오에 대한 금지와 불일치한다. 황금률이 발견되는 주요 출처들 어디에도 보복을 승인하는 가르침은 없다.

3. 빛바랜 황금률: 어리석은 자들의 도덕

합리주의에 기반 한 황금률의 해석에서는 평등성 또는 상호주의를 초월하는 어떤 요소이든 어리석은 자들의 도덕이 될 뿐이다. 여기에는 황금률의 원래 출처에서 강조되는 도덕적 엄격함이나 고상함 또는 초월적 도덕성 등이 사라진다. 예를 들어, 싱어는 황금률의 평등주의적 이념을 강조하면서 “다른 이들보다 더 엄격한 기준을 스스로에게 부과하는 사람은 비도덕적이지는 않지만 어리석은 사람이다. 왜냐하면 그 스스로에게 그렇게 엄격해야 할 실제적인 이유가 없기 때문이다.”⁴⁹⁾라고 말한다. 이것은 거위드의 합리화된 황금률에서도 마찬가지이다. “다른 이들이 당신에게 하기를 바라는 것처럼 다른 이들에게 합리적으로 행하라.”⁵⁰⁾는 것이 합리화된 황금률의 명령이라면, 황금률은 이성적인 고려와 합리적 존엄성을 가진 것으로 상대방을 대하라고 가르치기 때문이다. 우리는 순간의 행복뿐만 아니라 장기적인 복지를 고려한다. 그리고 우리는 수단들이 그러한 목적을 달성할 수 있는지를 고려한다. 그래서 효과적인 행위를 위해서는 확실한 믿음과, 정확한

49) Singer (1963), p. 303.

50) Gewirth (1978), p. 138.

추론, 그리고 직관적인 반성 등이 필요하다. 불필요한 희생은 불필요하고, 불필요한 것은 합리적이지 않다.

이러한 사태는 현대 게임이론에서 극명하게 나타난다. 게임이론에서 황금률은 사회적 해약을 유발하는 원리로 취급을 받고 있다. 악셀로드(Axelrod, 1984)의 죄수의 딜레마 컴퓨터 게임에서 황금률은 무조건 협조하는 전략 즉, ‘항상 협조(ALL C)’ 전략⁵¹⁾으로 여겨지는데, 이것은 도덕적으로는 고상할지 몰라도 실제로는 ‘무고한 사람’에게 피해를 줌으로써 사회를 망칠 정도로 부도덕하거나 잘해야 그저 ‘명칭이’ 전략⁵²⁾일 뿐이다. 황금률에 대한 상어와 거위드의 평등주의와 합리주의적 형식화에 따르면, 우리는 도덕적 영웅들을 최소한 무도덕적이라거나 비도덕적으로 평가해야 하는데, 이것은 우리의 일반적인 도덕적 직관에 부합하지 않는다. 그렇다면, 소위 말하는 ‘명칭이’도 아닌 그렇다고 무도덕적이거나 비도덕적이라고도 말할 수 없는 가치를 구제할 수는 없는 것일까?

VI. 결론

황금률의 주요 출처들에서는 현대 도덕철학에서 논의되는 것과 같은 일반화된 형식의 황금률이 발견되지 않는다. 황금률에 대한 비판과 옹호의 역사는 모두 황금률의 일반화 또는 일반화된 형식의 황금률을 중심으로 이루어지는데, 현대의 일반화된 황금률은 맥락을 무시한 일반화에 기반하고 있다. 이것이 전통적인 비판에 대해 황금률을 옹호하려는 노력이 그다지 만족스럽지 않은 이유이다. 일반화된 형식의 황금률을 구해내고자 하는 노력 역시 결국 황금률이 기원한 특수한 맥락에서 강조된 정신을 왜곡하는 결과를 초래할 수밖에 없기 때문이다.

51) 죄수의 딜레마 게임의 전략 중 하나로, 상대방이 협조(C)적이든 비협조(D)적이든 항상 협조하는 전략이다.

52) 역시 죄수의 딜레마 게임의 전략 중 하나로서, 상대방이 협조(C)적이든 비협조(D)적이든 항상 협조하는 전략으로서 항상 착취당하는 ‘명칭이의 전략’이라고도 불린다. 항상 협조하는 전략은 황금률이면서도 명칭한 전략이 된다 (각주 49 참조).

황금률은 다양한 맥락에서 관계를 맺는 특수한 행위자들에게 요구하는 특수한 가치들을 표방하고 있는데, 그러한 특수한 맥락은 주로 고대에 발생한 종교와 문화의 맥락에서 이해되어야 할 것이다. 따라서 현대 도덕철학에서 평등화되고 합리화된 그리고 상호주의적인 형태로 일반화 된 황금률은 황금률의 이러한 특성을 올바르게 반영하지 못한다. 특히 황금률의 원래 출처에서 발견되는 사랑, 자기초월, 공감 등의 가치들은 현대 도덕철학에서는 의무가 아닌 선택으로 여겨질 뿐인 초과의무적인 행위들이다. 그러나 평등화, 상호주의화 그리고 합리화된 일반형식의 황금률은 이러한 행위의 가치들을 왜곡할 수 있다는 점에서 황금률에 대한 올바른 이해를 저해할 여지가 많다.

왓틀스(Wattles)가 구분하는 황금률의 여섯 단계는 고대의 황금률과 현대 황금률의 차이를 확인할 수 있는 기준이 된다.

- 1단계: “다른 이가 당신의 욕구를 채워주기를 바라는 대로 당신은 다른 사람에게 행하라.”
- 2단계: “다른 이가 당신의 감정을 고려해주기를 원하는 대로 다른 이의 감정을 고려하라.”
- 3단계: “다른 이를 당신과 같이 합리적인 존엄성을 가진 인간으로 대하라.”
- 4단계: “다른 사람이 당신에게 행하기를 원하는 것처럼, 다른 이들에게까지 형제애를 확장하라.”
- 5단계: “다른 이가 당신에게 대하기를 원하는 것처럼, 다른 이를 도덕적 직관에 따라 대우하라.”
- 6단계: “신이 당신으로 하여금 그들에게 행하기를 원하는 대로 다른 사람들에게 하라.”⁵³⁾

이 중에서 마지막 단계는 황금률의 주요 출처인 종교적 수준의 것이다. 특히 6단계의 황금률 인간들 간의 관계인 합리성 수준(3단계)과 도덕성 수준(5단계)을 초월해 있는 것이다. 이것은 종교적으로는 신과의 관계에서의 인간 행위에 대한 규정으로 볼 수 있지만, 도덕철학적으로는 신적인 차원으로의 의식의 고양이라는 초월도덕적 가치를 내포하고 있음을 보여준다. 왓

53) Wattles (1987), pp. 106-129.

틀스는 이러한 비교를 통하여 고대 황금률이 더 고상한 도덕적 가치를 내포하고 있다고 말하고 싶어 하는 것 같다. 그러나 이러한 사실이 곧 현대 황금률이 도덕철학적으로 열등하다는 것의 근거가 되는 것은 아니다. 또한, 필자의 논지에 따라 판단하더라도, 특수한 맥락에서 특수한 행위자에게 부과된 도덕명령과 보편적인 도덕명령으로서의 현대 황금률의 도덕적 지위를 단순 비교할 수는 없다. 그럼에도 불구하고 고대 황금률과 현대 황금률의 비교 분석은 고대 도덕과 현대 도덕의 주요한 특징을 잘 보여준다. 즉 평등화, 상호주의, 합리화를 주요한 가치 기준으로 보는 현대 도덕철학에서 왜 고대 도덕의 주요 가치 기준들이 도덕적으로 합당하게 요구되는 가치가 될 수 없는지를 보여준다. 이런 점에서 고대 황금률과 현대 황금률에 대한 비교 분석은 고대 도덕과 현대도덕의 핵심적 차이에 대한 비교 분석과 다름없다.

투 고 일: 2016. 03. 23
심사완료일: 2016. 04. 18
게재확정일: 2016. 04. 18

박종준
서울시립대학교 자유융합대학

참고문헌

- 풍우란. 1993. 『중국철학사』. 정인재 옮김. 형설출판사.
- 칸트. 2005. 『윤리형이상학 정초』. 백종현 옮김. 대우고전총서 016. 아카넷.
- Armstrong, Karen. 2006. *The Great Transformation*. Alfred A. Knopf. New York / Toronto.
- _____, 카렌 암스트롱. 2010. 『축의 시대』. 정영목 옮김. 교양인.
- Bernard Shaw, George. 1930. *Man and Superman*. in *Collected Works of Bernard Shaw X*: New York.
- Blackstone, W. T. 1965. “The Golden Rule: A Defense”. *Southern Journal of Philosophy*. Winter: 172-177.
- Cunningham, W. Patrick. 1998. “The Golden Rule as Universal Ethical Norm”. *Journal of Business Ethics* 17: 105-109.
- Gensler, Harry J. 2013. *Ethics and the Golden Rule*. Routledge.
- Gewirth, Alan. 1978. “The Golden Rule Rationalized”. in *Ethical Theory: An Anthology*. John Wiley & Sons. Inc. 2013. pp. 521-535.
- Gould, James A. 1963. “The Not-So-Golden Rule”. *Southern Journal of Philosophy*. Fall: 10-14.
- Hare, Richard M. 1963. *Freedom and Reason*. Oxford: Oxford University Press.
- Hennessey, Jr J. W. and Gert, Bernard. 1985. “Moral Rules and Moral Ideals”. *Journal of Business Ethics* 4: 105-115.
- The Holy Bible*: authorized King James version. Bible House. Charlotte. North Carolina. 1976.
- _____. 『성경』. 말씀보존학회. 1994.
- Huang, Youg. 2005. “A Copper Rule Versus The Golden Rule: A Daoist-Confucian Proposal for Global Ethics”. *Philosophy East and West*. Vol 55: 394-425.
- Kant, Immanuel. 1959. *Foundations of the Metaphysics of Morals*. by

- Lewis White Beck. Bobbs-Merrill Educational Publishing. Indianapolis.
- Lippmann, Walter. 1922. *Public Opinion*. New York: The Macmillan Company.
- Norman J. Bull. 1969. *Moral Education*. London: Routledge and Kegan Paul.
- Reinikainen, Jouni. 2005. "The Golden Rule and the Requirement of Universalizability". *The Journal of Value Inquiry*. 39: 155-168.
- Ricoeur, Paul. 1990. "The Golden Rule: Exegetical and Theological Perplexities". *New Testament studies* 36: 352-397.
- Russell, L. J. 1942. "Ideals and Practice". *Philosophy*. Vol. XVII: 109-110.
- Sidgwick, Henry. 1963. *Methods of Ethics*. 7th ed.. London: Macmillan & CO LTD.
- Singer, Marcus G. 1963. "The Golden Rule". *Philosophy*. Vol. 38: 293-314.
- Wattles, Jeffrey. 1987. "Levels of meaning in the Golden Rule". *Journal of Religious Ethics* 15: 106-129.
- Weiss, Paul. 1941. "The Golden Rule". *The journal of Philosophy*. Vol. 38. no. 16: 421-430.

ABSTRACT

Problems of The Contemporary Golden Rule as a Moral Principle

Park, Jong-June

The golden rule has been subjected to such criticism that it cannot be qualified as a moral principle. Against their critics, contemporary defenders of the golden rule have changed it into different forms. The contemporary golden rules can be characterized as ‘equalized’, ‘reciprocated’, and ‘rationalized’. Critics and defenders use a generalized form of the golden rule, in this case “Do unto others what you would have them do unto you” or “Do not do unto others what you would not have them do unto you,” known as ‘the passive rule’. I argue in this paper that the contemporary generalized golden rules have a serious moral problem and that the particular moral values indicated in the original texts from which the idea of the golden rule derives cannot be successfully emphasized in the general form.

Subject Class: Ethics, Philosophy of Religion.

Keywords: the golden rule, Kant, Singer, Gewirth, the general interpretation of the golden rule, the particular interpretation of the golden rule, rationalized golden rule, supererogatory.