한국심리학회지 : 문화 및 사회문제
Korean Journal of Culture and Social Issues
2026, Vol. 32, No. 2, 295~327.
http://dx.doi.org/10.20406/kjcs.2026.5.32.2.295

심리적 구성개념으로서의 일반의지: 일반의지의 21세기적 소환*

윤 황†

본 연구는 데이터리즘(Dataism)이 개인의 자유의지를 대체하고 공동체적 유대를 해체하는 오늘날의 실존적 위기를 배경으로, 18세기 정치철학적 개념인 Rousseau의 일반의지(General Will)를 21세기 심리학적 구성개념으로 전용하고자 하는 이론적 시도이다. 본 연구에서는 Rousseau의 자연상태 이론, 자기사랑과 이기심의 분기, 탈자연화 및 공동자아 개념을 심리학적 언어로 재해석하고, 이를 토대로 공공선 지향성, 자기입법적 주체성, 정서적 연대성의 세 하위요인으로 구성되는 새로운 심리학적 구성개념으로 일반의지 지향성(General Will Orientation: GWO)을 제안하였다. GWO의 독자성을 논증하기 위해 Kant의 정언명령, Rawls의 공정으로서의 정의 등 도덕철학 이론들 및 Kohlberg의 도덕발달이론, 긍정심리학의 VIA 체계 등 심리학적 유사 변인들과 비교하였다. 또한 인간 도덕성의 부정적 극단인 정신병질(Psychopathy)과의 대칭적 구조를 통해 GWO가 구성적 선(善)으로서 동일한 행동 영역 위에서 반대 방향으로 작동하는 측정 가능한 심리학적 구성개념임을 논증하는 한편, 규범윤리와 덕윤리의 통합적 대안으로서 GWO의 이론적 위상을 제시하였다. GWO는 알고리즘 시대에 인간의 주체적 도덕 역량을 보호하는 심리학적 대항 기제로서 도덕심리학의 외연을 철학적 깊이로 확장하는 데 기여할 것으로 기대된다.

주요어 : 루소, 일반의지, GWO, 데이터리즘, 정신병질

* 이 논문은 2026학년도 배재대학교 교내학술연구비 지원에 의하여 수행됨.

† 교신저자: 윤황, 배재대학교 심리상담학과 교수, 대전광역시 서구 배재로 155-40, Email: andsalt@pcu.ac.kr

오늘날 현대인의 일상은 데이터리즘(Dataism)이 구축한 정교한 알고리즘 체계와 분리되어 생각하기 어렵다. 우리는 끊임없이 쏟아지는 정보의 홍수 속에서 알고리즘이 제공하는 선택의 자유와 기술적 편리 속에서 살아가고 있다. 알고리즘이 건네주는 선택지 속에서 유튜브를 시청하고 신문기사들을 읽으며, 쇼핑도 하고 식사 메뉴도 결정한다. 심지어 정치적으로는 진영논리를 강화하고, 자신과 관계된 모든 내집단과 외집단에 대한 편향을 심화해나간다(Pariser, 2011). 이러한 문명의 이기와 선택의 자유는 인간이 환경과 상호작용하며 저마다 조형해가던 세상 인식의 틀(schema)을 알고리즘 시스템에 넘겨주는 역설을 낳고 있다. 데이터리즘은 환경적 조건의 카운터 파트너로서 개인의 자유의지를 대체하고 도식을 조형해 나가고 있는 중이다(Harari, 2017/2023). Hobbes(1651/2018)가 묘사한 '자연상태'가 물리적 생존을 위협받는 불안의 시대였다면, 오늘날의 디지털 환경은 데이터리즘이 제공하는 어항 속에서 선택의 자유를 누리며 불안을 해소하는 '실존적 아노미' 상태에 비유해 볼 수 있다. 이러한 변화는 타인과의 관계 맺기 방식에서도 뚜렷하게 관찰된다. Rousseau (1755/2003a)가 일찍이 경고했던 '타인의 시선에 종속된 삶'은 오늘날 SNS라는 가상공간으로 전이되어 재연되고 있는 중이다(장경원 등, 2022). 본래의 선량한 '자기사랑(Amour de soi)'이 타인과의 끊임없는 비교 속에서 '이기적 자기애(Amour-propre)'로 변질되어 가는 과정은 현대인이 겪는 고립과 소외의 심리적 배경을 이룬다. 디지털 공간에서 타인의 승인과 비교를 전제로 조형된 자아는 주체적 결단력을 상실한 채 파편화된 개인주의로 침잠되는 결과를 초래하고 있다.

본 연구에서는 이러한 실존적 위기에 대한 대안으로서, Jean-Jacques Rousseau가 제안한 '일반의지(General Will; Rousseau, 1762/ 2011)'의 심리학적 전용을 제안하고자 한다. 18세기 정치철학의 산물인 일반의지를 21세기 현대 심리학으로 소환하는 것은, 일반의지가 단순한 정치적 합의를 넘어, 실존자로서의 주체성 회복과 파편화된 개인주의를 극복할 수 있는 강력한 '동기적 기제'로서 기능할 가능성을 탐색하기 위함이다. 이를 위해 본 연구에서는 먼저 정치철학적 개념인 일반의지를 소개하고 심리학적 구성개념으로 재해석할 것이다. 이어지는 논의에서는 일반의지가 Kant나 Rawls의 개념 및 Kohlberg의 도덕발달 5단계(사회계약 지향) 개념, 기존 긍정심리학의 유사 변인들과 차별화되는 지점을 고찰할 것이다. 그리고 공동체적 선을 지향하는 심리적 지표로서의 일반의지를 그 대척점에 있는 정신병질과 대비시킴으로써, 형이상학적 개념인 선(善)이 일반의지를 통해 심리학적으로 구성 가능한 실체로서 구현될 수 있음을 논증하고, 일반의지가 데이터리즘 시대의 주체성 회복을 위한 새로운 심리학적 구성개념으로서 지니는 이론적 타당성을 검토할 것이다. 마지막으로는 선을 반영하는 실체적 개념으로서의 일반의지가 현대 윤리학의 난제인 규범윤리와 덕윤리 간 통합적 대안이 될 수 있음을 논의할 것이다.

Rousseau의 일반의지와 이론적 서사

자연상태와 자연인

사회계약론으로 대표되는 Rousseau의 정치사상은 인간 본성에 대한 심리학적 통찰에서

출발한다(박호성, 1993). 그는 인간의 마음이 사회적 관계 속에서 어떻게 달라지고 있으며, 그러한 마음의 상태가 어떠한 사회상을 형성하게 되는가에 따라 인류의 문명사를 자연상태, 사회상태, 대안적 사회상태(정치상태)로 구분한다(임의영, 2020; Rousseau, 1755/2003a).

사회계약론의 대표적인 사상가들인 Hobbes와 Locke, Rousseau는 태초 인간의 상황을 자연상태로 가설하고 있다. Hobbes(1651/2018)의 자연상태는 자기보존을 위해 각자가 서로를 겨누는, 소위 만인에 의한 만인의 투쟁으로, 공포와 허영이 존재의 동인을 이루게 됨으로써 절대권력과의 사회계약을 필요로 한다. Locke(1689/2022)의 자연상태는 자연법이 지배하는 자유롭고 평등한 세상을 가정한다. 그 속의 자연인은 자연법을 인식할 수 있는 합리적인 존재이지만, 각자의 권리가 상충하는 불안정한 상태를 극복하기 위해 시민주권의 사회계약을 추구하게 된다.

Rousseau의 자연상태는 Hobbes나 Locke와는 다르다. Rousseau의 자연상태는 Hobbes처럼 자기보존을 위해 각자가 서로를 겨누는 전쟁상황이 아니라, 자유롭고 평등한 Locke의 자연상태와 유사한 상황을 가정한다. 그러나 그 속에서 살아가는 자연인은 자연법을 인식할 수 있는 합리적인 존재가 아니라는 점에서 Rousseau의 자연상태는 자연법이 작동하지 않는 세상이다. Rousseau는 인간불평등기원론(1755/2003a)을 통해 그가 그리는 자연상태와 자연인을 아래와 같이 묘사하고 있다.

> 원시의 인간은 일도 언어도 거처도 없고, 싸움도 교제도 없으며, 타인을 해칠 욕구가 없듯이 타인을 필요로 하지도 않고, 어쩌면 동류의 인간을 개인적으로 단 한번도 만난 적 없이 그저 숲속을 떠돌아다녔을 것이다. 그는 얼마 안 되는 정념의 지배를 받을 뿐 스스로 자족하면서 자신의 상태에 맞는 감정과 지적 능력만을 갖고 있었다. 원시의 인간은 자신의 진정한 필요만을 느꼈고, 눈으로 보아 흥미롭다고 여겨지는 것만을 쳐다보았다. 그의 지능은 그의 허영심과 마찬가지로 발달하지 못했다. 우연히 그가 어떤 발견을 한다 해도 그는 자신의 지식조차 기억하지 못하기 때문에 그것을 전수할 수 없었다. 기술은 발명자와 더불어 소멸했다. 교육이란 것은 존재하지 않았으며, 아무런 진보도 없이 세월이 흐름에 따라 세대가 이어질 뿐이었다. 그리고 각각의 세대는 언제나 똑같은 지점에서 출발했으므로, 최초의 시대의 모든 조야함 속에서 수백 년이 되풀이 되며 흘러갔다. 종은 이미 늙었으나 인간 개체는 항상 어린애로 머물러 있었다(Rousseau, 1755/2003a; 임의영, 2020 재인용)

Rousseau의 자연상태는 이성적 판단이나 권리의 추구 이전에 감정과 본능의 심리적 평온함이 지배하는 세상이다. Rousseau의 자연인은 고립되어 살아가지만, 그 내면에는 생존을 위한 '자기사랑(amour de soi)'과 타인의 고통에 직관적으로 반응하는 '연민(pitié)'이라는 두 가지 원초적 정념이 균형을 이루고 있다(박찬영, 2022; 박호성, 1993; Rousseau, 1755/2003a). 이 중 연민은 타자의 고통에 반응함으로써 심리적 불편감을 해소하고 평온함을 유지하는 항상성 기제인 바, 도덕적으로 학습된 상태라기보다는 본능적 기제에 가깝다(임의영, 2020; Rousseau, 1755/2003a). 이는 타인의 고통을 자

신의 정서적 체계 안으로 수용하는 원초적 수준의 상호성을 내포하고 있다고 볼 수 있는데, 이는 Rousseau의 자연인이 단순히 고립된 원자가 아니라, 타자와 정서적으로 공명할 수 있는 심리적 토대를 이미 갖추고 있음을 시사한다. 따라서, Rousseau의 자연인은 자기보존과 심리적 평온함이 존재의 동인이며, 그 본성은 자기사랑과 연민을 통해 살아가는 자족적 선량함이라 할 수 있다(임의영, 2020). 그는 Hobbes의 인간처럼 타인을 정복하여 허영을 채울 필요도, Locke의 인간처럼 권리간 상충으로 긴장할 필요도 없다. 오직 내면의 소리에 귀 기울이며 평온함 속에서 살아가는 자율적 존재인 것이다.

사회상태로 진행 서사

Rousseau는 동물과 구별되는 인간만의 고유한 특성으로 '자유의지(agent libre, free agent)'와 '완성가능성(perfectibilité, perfectibility)'을 든다. 비둘기가 고기 그릇 옆에서, 고양이가 과일 더미 위에서 본능의 명령을 어기지 못해 굶어 죽는 것과 달리(Rousseau, 1755/2003a), 인간은 자연의 명령을 거스를 수 있고 때로는 방종할 수 있는 자유를 지닌 존재다. 이러한 자유의지는 주어진 본능의 경계를 넘어 환경에 적응하고 스스로를 변화시키는 완성가능성으로 이어진다. 본래 이 잠재력은 고정된 본능에 머물지 않고 선량함을 향해 나아갈 수 있는 가능성이었으나, 인구가 늘고 인간이 한 곳에 모여들기 시작하면서 그 가능성은 왜곡된 방향으로 자라나게 된다. 자기 내면의 목소리에만 귀 기울이던 자연인이 타인과 빈번하게 마주치고 서로를 의식하고 비교하게 되면서 자족적 평온함엔 균열이 발생한다. 생존을 뒷받침하던 건강한 '자기사랑(Amour de soi)'은 타인과의 비교와 인정을 갈구하는 배타적 '자기애(Amour-propre)'로 변질되고, 이 과정에서 자기사랑과 연민이 유지하던 정서적 균형이 무너지면서 인간의 내면에는 이기심과 탐욕이 자리 잡게 된다. Rousseau는 이러한 정념의 전도가 소유의 개념과 사유재산의 발생을 고착시켰으며, 이것이 곧 인간 불평등의 기원이 되었다고 주장한다(Rousseau, 1755/2003a). 결과적으로 Rousseau가 지각한 사회상태는 심리적 평온함을 자신의 내부에서 찾지 못하고, 타인과의 비교를 통해 갈구하게 된 상태로서, 타인의 인정 없이는 스스로를 긍정할 수 없는 '의존적 존재'로 전락해 버린 시대이다. 자족적 선량함이 사라진 자리에는 평온함 대신 타인보다 우월해지려는 욕망과 결핍에 따른 불안이 들어서며, 인간은 이제 끊임없는 비교 속에서만 자신의 존재를 증명해야 하는 타락의 시대를 살아가게 된다(박호성, 1993; 임의영, 2020; Rousseau, 1755/2003a). 그리고 이 타락의 서사는 제도적 불평등이 구조적으로 공고화된 사회로 귀착된다.

정치철학적 개념으로서의 일반의지

Rousseau는 당시 18세기 사회를 인간의 본성이 왜곡되고 타인의 시선에 종속된 '타락의 시대'로 규정했다(김은주, 2023; Rousseau, 1755/2003a; 임의영, 2020). 그리고 이러한 타락한 사회상태를 극복하기 위한 대안적 사회상태(정치상태)로서, 자신의 저서 사회계약론(Rousseau, 1762/2011)을 통해 '일반의지(General Will)'에 기초한 새로운 사회계약의 체결을 제안한다. 이때 일반의지와 사회계약은 상호간의 단순한 인과물이라 할 수 없는데, 이를 이

해하기 위해서는 '일반의지(Volonté générale, General Will)'에 대한 개념적 이해가 선행될 필요가 있다.

우선, 일반의지의 '일반(General)'은 의지가 지향하는 보편적 대상과 범위를 의미한다. Rousseau에게 있어 일반적이라는 것은 통계적 다수를 의미하는 것이 아니라, 공동체 구성원 모두에게 차별 없이 적용되는 보편적 가치 지향을 뜻하며, 절차적으로는 구성원 모두의 합의를 전제로 한다(Sreenivasan, 2000). 한편 '의지(Will)'는 외부의 강요가 아닌 개인의 주체적이고 자발적인 입법적 결단을 의미하는데, 앞서 살펴본 바와 같이, Rousseau는 인간을 다른 동물과 달리 '자유의지'를 가진 존재로 간주한다. 즉, 의지는 수동적인 욕구나 충동이 아니라, 자신이 따를 삶의 원칙을 스스로 결정하는 주체적이고 능동적인 행위자성(Agency)의 발현을 반영한다(Rousseau, 1762/2011). 따라서 '일반의지(General Will)'는 타율적인 구속이 아니라, 자발적인 선택을 통해 도덕적 주체로 거듭나겠다는 의지적 결단으로서, 그것이 일반적이기 위해서는 능동적 주체자로서 구성원의 합의가 요구된다. 그런데, 그것이 어떻게 가능한가? 능동적 주체자로서 내린 결정이 구성원 모두에 의해 합의되기 위해서는 그 합의된 내용이 즉각적으로는 구성원 각자에게 이해의 차이가 발생할 순 있어도 궁극적으론 구성원 모두에게 이익이 될 수 있는, 이른바 공공선으로서의 조건을 충족시켜야만 한다. 즉, 공동체의 주권자로서 구성원 모두가 참여하여 내린, 구성원 모두에게 이익이 될 수 있는 자발적 결단의 총체가 곧 일반의지이며, 모두의 참여와 모두의 이익을 요구한다는 점에서 철저하게 자유와 평등을 전제하고 있다. 또한 Rousseau가 '주권은 양도할 수 있어도 의지는 양도할 수 없다'고 천명한 바와 같이, 구성원 각자가 주권자로서 갖는 고유한 의지라는 점에서 일반의지는 타인에게 양도할 수 없으며(김용민, 2016; Rousseau, 1762/2011), 비록 구성원 각자의 판단은 흐려질 수 있고, 합의된 결과가 틀린 것일 수는 있어도 의지 그 자체는 항상 공공선을 향해 올바르게 존재한다는 점에서 무오류성의 실체로서 존재한다(임의영, 2020; Rousseau, 1762/2011). 따라서 일반의지는 공동체 구성원의 '공동자아(Moi commun, Common Self)'라 할 수 있으며, 이는 단순히 계약서에 서명하는 행위를 넘어, 구성원들이 정서적, 이성적으로 결속되어 공공선(Common Good)을 추구하는 하나의 도덕적, 집단적 신체를 형성함을 의미한다(임의영, 2020; Rousseau, 1762/2011). 그리고 공동체 구성원은 각자의 모든 권리를 일반의지에 위임하고 일반의지의 강제에 스스로 복종하는 사회계약을 맺는다는 점에서 소위 '자유롭도록 강제함(on le forcera d'être libre, forced to be free)'을 수용한다[1](오근창, 2013; Rousseau, 1762/2011). 이러한 관점에서 일반의지는 사적 이익의 총합인 '전체 의

1) Rousseau의 '자유롭도록 강제한다'는 표현은 흔히 전체주의적 억압을 정당화하는 수단으로 오해받아 왔다. 그러나 Rousseau의 체계 내에서 이 강제는 외적인 폭력이 아니라, 시민이 공동체의 구성원으로서 사회계약에 참여할 때 이미 자발적으로 동의한 '일반의지의 행사'를 의미한다. 즉, 개인이 사적인 이해관계나 일시적인 충동에 매몰되어 자신이 스스로 입법한 보편적 원칙을 어기려 할 때, 공동체는 그가 본래 추구했던 '입법자로서의 주체성'을 유지할 수 있도록 그를 원칙으로 회귀시키는 것이다. 결국 이 강제는 개인을 억압하는 것이 아니라, 오히려 그가 타인의 의지나 자신의 정념에 예속되지 않고 본연의 자유로운 상태를 유지할 수 있도록 보장하는 논리적 장치로서 기능하게 된다.

지(Will of all)'나 개인의 욕망인 '개별 의지(Particular will)'와는 엄격히 구별된다(Kierstead, 1974; Rousseau, 1762/2011). 전체 의지가 다수결이라는 양적 논리라면, 일반의지는 다수결 이전에 존재하는 '무엇이 옳은가'에 대한 질적 합의인 것이다.

따라서 일반의지와 사회계약의 관계는 현대 국가에서 '헌법과 제도'의 관계에 비유해 볼 수 있는데, 헌법이 국가 구성의 원리이자 가치 지향(일반의지)이라면, 제도는 이를 실현하기 위한 구체적인 시스템(사회계약)이다. 일반의지가 사회계약이라는 형식에 생명력과 정당성을 불어넣는 구심점이자 동력이 되는 셈이다.

일반의지의 심리학적 전이와 재해석

일반의지는 단순한 정치적 합의가 아니라, 개인이 고립된 '자연인'에서 공동체적 주체인 '시민'으로 이행할 때 발생하는 구성원 각자의 '마음의 변화'에 그 기반을 두고 있다. Rousseau는 사회계약론과 같은 해 출간된 『에밀』에서 유아기부터 성인기에 이르는 5단계의 발달과정을 통해, 인간의 자연적 감각과 욕구를 억압하지 않으면서도 이를 점진적으로 사회적 양심과 도덕적 이상으로 고양시키는 구체적인 '자연적 선성의 도덕적 승화과정'을 제시한 바 있다(Rousseau, 1762/2003b). 이는 일반의지의 실현이 법률이나 제도와 같은 외부 시스템의 구축을 넘어서, 개인의 내면에서 발현되는 심리적 역량을 통해 완성되는 것임을 시사한다. '자연으로 돌아가라'는 Rousseau의 주장은 마음의 변화 과정을 상징적으로 함축하고 있는데, 여기서의 자연은 원시 자연상태로의 회귀를 의미하지 않는다. 자연으로 돌아간다는 것은 사회상태를 극복하고 일반의지를 실천할 수 있는 마음 상태로의 변화를 의미하며(곧, 사적 욕망에 오염되지 않은 주체적 입법이 가능한 상태로의 변화), 이 변화과정을 Rousseau는 '탈자연화(Dénaturation, Denaturalization)'라 명명한다. 그리고 Rousseau는 탈자연화된 마음 상태의 개인을 절대적 단독자(1)였던 자아가 전체의 일부(1/n)로 재구성되는 과정이라 설명하고 있는데(Rousseau, 1762/2011), 이는 개인이 가진 본성의 파괴를 의미하지 않는다. 심리학적으로 환언하면, 이는 자아의 분열이나 축소가 아니라, 협소한 자기중심성을 탈피하여 타인과 공동체로 자아를 넓히는 '자아의 확장(étendre son existence, extend one's existence)'을 의미하며, 확장된 자아들의 총체가 곧 '공동자아(Moi commun, Common Self)'로서의 일반의지이다(Rousseau, 1762/ 2011). 표 1은 정치철학적 개념인 일반의지를 개인의 심리상태를 반영하는 심리학적 구성개념으로 전이한 내용을 정리하고 있다.

우선, 정치철학적으로 일반의지의 주체는 주권자로서의 시민이고, 그 대상에게 부여되는 일반성은 나를 포함한 공동체 구성원 모두에게 적용되는 법률의 보편성이다(Rousseau, 1762/ 2011). 이를 심리학적 개념으로 옮겨보면, 일반의지의 주체는 확장된 자아로서, 일반성은 나의 정체성을 타자와 공동체 전체로 확장하여 인식하는 상호연결성, 즉 공동자아의 상태를 의미한다(임의영, 2020; Rousseau, 1762/2011). Rousseau가 사회계약을 통해 형성된 결사체를 공동자아라 명명했듯이, 심리학적 공동자아는 나와 타인, 나와 공동체의 상호의존성을 인식하고 타인의 안녕을 나의 정체성 안으로 통합해내는 고차원적으로 연결된

표 1. 정치철학적 개념으로서의 일반의지와 심리학적 구성개념으로서의 일반의지 비교

구분	정치철학적 개념	심리학적 구성개념
주체	주권자로서의 시민	확장된 자아
일반성(General)	법 적용의 보편성	타인 및 공동체와의 상호연결성
의지(Will)	주체적 입법행위	주체적 행위자성
양도 불가능성	대표 불가능	의지의 주체성
무오류성	공공선의 지향	도덕적 지향성
핵심기제	사회계약과 투표	자기사랑의 확장을 통한 공동체적 결속
자유의 성격	자유롭도록 강제함	자율적 자기조절

상태라 할 수 있다.

앞서 언급한 바와 같이, Rousseau는 '권력은 양도될 수 있어도 의지는 양도될 수 없다'고 천명하면서 일반의지의 양도 불가능성을 강조한 바 있다(Rousseau, 1762/2011). 현대 심리학적 맥락을 통해 이는 '의지적 주체성'으로 번역해 볼 수 있다. 즉, 알고리즘과 데이터가 최적의 선택을 제시하는 데이터리즘 시대에 자신의 도덕적 판단과 선택권을 외부 시스템으로 외주화하지 않고, 스스로 입법자가 되어 결단하려는 강력한 내적 동기를 일컫는다. 따라서, 심리학적 개념으로서의 의지는 외부환경의 압력이나 유혹에 굴하지 않고 자신이 세운 원칙을 고수하는 이른 바 '마음의 주권'이라 할 수 있다(Ryan & Deci, 2000).

정치철학적 개념으로서의 일반의지는 '항상 옳으며 공공의 이익을 지향한다'는 무오류성의 특징을 갖는다. 심리학적으로 이러한 정치적 무오류성은 일반의지가 공동체의 선을 향하고 있다는 '도덕적 지향성(Moral Orientation)'으로 해석된다. 일반의지는 구성원 모두가 참여하여 내린, 구성원 모두에게 이익이 될 수 있는 자발적 결단이기 때문이다. 따라서 비록 개인의 인지적 판단이나 구성원들 간의 합의는 오류를 범할 수 있을지라도, 도덕지향으로서의 일반의지는 이와는 별개의 실체로서 항상 올바른 곳을 가리키게 된다. 또한 이러한 지향성을 현실화하는 정치적 기제는 사회계약과 투표가 되는데, 사회계약을 통해 사적 개인은 공적 시민으로 다시 태어나며, 투표는 단순한 다수결의 승패를 가리는 것이 아니라 무엇이 공동체의 이익인가를 식별해내는 절차로서의 의미를 갖는다. 이때 Rousseau는 '자신을 생각하지 않고서는 전체를 생각할 수 없다'면서, 투표 행위의 이면에는 자기 이익의 추구가 전제되어 있음을 인정한 바 있다. 본 연구에서는 일반의지가 갖고 있는 이러한 정치적 기제를 심리학적 구성개념으로 전이하기 위해 Rousseau의 '자기사랑(Amour de soi)'과 자기심리학 이론가인 Kohut(2009)의 '성숙한 자기애(Mature Narcissism)' 간의 개념적 유비에 주목한다. Rousseau가 자기사랑을 타인과의 비교나 우월감을 추구하는 '이기심(Amour-propre)'과 구별하여 자기 보존과 질서 형성의 긍정적 원동력으로 규정했듯이(박찬영, 2022; Rousseau, 1755/2003a), 현대 자기심리학 역시 병리적 나르시시즘과 구별되는 '성숙한 자기애'를 자아통합과 창조성의 원천으로 정의한다(Gabbard,

2014; Kohut, 2009). 특히 Rousseau가 자기사랑의 확장을 통해 타인을 자신의 일부로 느끼는 공동체적 결속을 설명했듯이, 성숙한 자기애는 타인을 단순한 대상이 아닌 자신의 자아를 확장시키는 '자기대상(Selfobject)'으로 포용하는 심리적 기제를 제공한다(Gabbard, 2014; Kohut, 2009). 따라서 사회계약과 투표가 사익을 공익으로 전환시키는 정치적 절차라면, 성숙한 자기애는 고립된 자아가 타자와의 연대를 통해 확장된 자기로 나아가게 하는 심리적 기제로서 일반의지의 동력으로 작용한다.

끝으로, Rousseau에게 '자유롭도록 강제함'은 사적 욕망으로부터 자유롭지 못한 개별의지를 극복하고 스스로 제정한 법, 즉 일반의지에 복종함으로써 획득되는 도덕적 자유의 실현을 의미한다. 이를 심리적 구성개념으로 전이시켜 보면, 자신의 충동과 유혹을 억제(개별의지)하고 내면의 보편적 도덕률(일반의지)에 자율적으로 복종하는 자기조절기제(self-regulation)로 연결된다(Bandura, 1986). Rousseau에게 있어서 '덕(la vertu, virtue)'은 단순한 선함이 아니라, 자신의 사적인 욕망(개별의지)을 억제하고 일반의지에 일치시키는 '의지의 힘'으로써, 그의 '유덕한 시민(le citoyen vertueux, virtuous citizen)'은 개별의지와 일반의지가 일치하는 인간상을 일컬으며, 유덕한 시민만이 진정으로 자유롭다(Rousseau, 1755/2003a, 1762/2011). 따라서 심리적 구성개념으로서의 일반의지는 사적 욕망에 이끌리는 상태가 아니고, 그렇다고 외적 강제나 억압에 의해서도 아니라, '원칙을 세우고 실현하는 나'의 의지적 자기조절 능력을 의미하며, 이것이 곧 Rousseau가 꿈꾸었던 '유덕한 시민'의 심리학적 초상이다.

심리학적 구성개념으로서 일반의지의 독자성

이제부터는 심리학적 구성개념으로서의 일반의지와 철학적 유사 개념 및 심리학적 유사 변인들과의 비교를 통해 일반의지가 갖는 경계를 명확히 하고 그 독자성을 밝히고자 한다.

일반의지와 Kant의 정언명령

Rousseau의 일반의지와 Kant의 정언명령은 현대 윤리학의 핵심인 '자율성'의 토대를 공유한다. Kant가 Rousseau의 『에밀』과 『사회계약론』으로부터 깊은 철학적 영감을 받았음은 주지의 사실이며(김시형, 2015), 이는 두 사상가가 '자유'를 정의하는 방식에서 명확히 드러난다. Rousseau에게 있어서 자유는 '자신이 입법한 법에 자발적으로 복종하는 것'이라면, Kant에게 있어서 자유는 '이성이 스스로 입법한 도덕 법칙에 따르는 것'으로서 '자율적 자기입법'이라는 공통점을 갖는다. 즉, 두 사상가 모두 도덕적 행동의 근거를 외적 권위나 맹목적 충동 등 타율이나 우연에서 찾지 않고, 행위주체의 자기 입법과 이에 대한 자발적 복종에서 찾았다는 점에서 궤를 같이한다. 또한, Rousseau의 일반의지가 개별적 이해관계를 넘어 공동체 전체에 적용되는 법 적용의 '보편성'을 전제하듯, "네 행위의 준칙이 동시에 보편적 법칙이 될 수 있도록 행위하라"는 정언명령의 목소리 역시 도덕 법칙의 '보편성'을 향하고 있다는 점에서 그 논리적 출발점은 동일하다고 볼 수 있다. 그러나 이러한 형식적 유사성에도 불구하고, 도덕적 행동을 유발하는 심리적 기제가 무엇인가는 두 이론을 명백히 다른 이론으로 구분한다.

Kant의 도덕철학은 이성과 경향성을 엄격히 준별하는데, Kant에게 있어서 경향성이란 감각적 욕구와 본능적 충동을 일컬으며, 이는 가변적이고 주관적인 속성을 갖는다(Kant, 1788/2009). 그러나 도덕법칙은 언제 어디서나 보편타당성을 가져야 한다는 점에서 경향성은 도덕의 기초가 될 수 없으며, 오로지 이성이 명령하는 도덕적 의무에 따를 때만이 도덕적 행동으로서의 지위를 갖게 된다. 즉 Kant에게는 어떤 행위가 도덕적 결과를 초래할지라도 그 이유가 도덕적 의무에서 비롯되지 않았다면 도덕적 행동이라 볼 수 없다. 상인이 아무리 정직하게 장사를 한다 할지라도 그것이 내 마음 편하자고 정직한 것이라면 도덕적인 것이 아니며, 오로지 정직해야 한다는 의무감에서 비롯되었을 때만이 도덕적인 것이 된다(Kant, 1785/2018). 이러한 Kant의 도덕 모델은 논리적 정합성과 도덕적 순수성을 확보하는 데는 탁월하지만, 심리학적 관점에서는 중요한 난제를 남긴다. 도덕적 행동을 위해 자연적 정서와 동기, 즉 경향성을 통제해야 한다는 전제는 행위자에게 끊임없는 내적 긴장을 발생시키는데, 결국, '해야 한다'가 '하고 싶다'를 극복할 것을 요구한다는 점이다.

반면, Rousseau의 일반의지는 이성과 감정의 분리가 아닌 통합을 지향한다. 도덕적 행동을 위해 자연적 본성을 통제하는 것이 아니라, 본성을 도덕적 차원으로 고양시키는 방식을 취하는데, 이는 곧 인간의 자연적 본성인 자기사랑이 일반의지로 승화되는 과정을 일컫는다. 그 과정을 들여다보면, 먼저, '자기사랑'은 인간의 가장 원초적이고 자연스러운 감정인 자기보존 욕구로서, Kant에 따르면 이는 경향성에 해당 되지만, Rousseau에게는 도덕적 에너지의 원천이 된다. '이성'은 확장의 도구로서 자기사랑을 맹목적인 상태에 머물지 하지 않고, 타인과의 관계 속으로 확장시킨다. 이성을 통해 "나의 보존이 중요한 만큼 타인의 보존도 중요하다"는 상호성이 발생하며, 사적 이해관계를 넘어 공공의 이익을 사고할 수 있는 능력을 갖추게 된다(박찬영, 2022; Rousseau, 1762/2003b). '양심'은 동기적 기제로서 "이성이 선을 알려주면, 양심은 선을 사랑하도록 안내한다(Rousseau, 1762 /2003b)." 즉 이성은 무엇이 선인지 판단하지만, 그 선을 행하고 싶게 만드는 힘은 양심을 통해 발휘된다는 것으로, 양심은 확장된 자기사랑을 공공선과 일치, 유지시키는 내면의 힘이라 할 수 있다(김용민, 2016). 마지막으로, 자발적 의지로서 일반의지의 발현이다. 자기사랑, 이성, 양심의 단계를 거쳐 형성된 일반의지는 더 이상 의무가 아니다. 그것은 확장된 자아가 공동체의 안녕을 위해 자발적으로 선택한 의지적 열망으로서 '해야 한다'를 넘어선 '하고 싶다'의 승화된 모습이며, '해야 한다'는 당위보다 훨씬 강력하고 지속적인 행동의 동력을 제공한다(오수웅, 2017).

요컨대, Kant와 Rousseau는 모두 자율적인 도덕적 주체로서 도덕 법칙의 보편성을 지향하지만, 그 도달 경로는 상이하다. Kant의 모델이 도덕적 순수성을 위해 자연적 경향성을 통제해야 하는 자기극복의 윤리라면(김시형, 2015; Kant, 1788/2009), Rousseau의 일반의지는 자연적 본성인 자기사랑을 이성과 양심을 통해 사회적 수준으로 확장시키는 통합의 윤리라 할 수 있다. 이러한 Rousseau의 관점은 의무와 욕구, 혹은 이성과 정서를 대립적인 것으로 보지 않고 통합하려 한다는 점에서, 현대 심리학이 지향하는 통합적 도덕성 모델에 부합한다. 또한 두 이론이 갖는 메커니즘의

차이는 데이터리즘의 시대에 더욱 극명한 시사점을 제공한다. Kant의 정언명령은 입력된 준칙이 보편화 가능한지를 검증하는 일종의 알고리즘적 절차로 환원 가능하다. 만약 도덕이 순수한 논리적 형식과 엄격한 규칙 준수라면, 인간보다 더 완벽하게 편향 없이 계산하고 통제할 수 있는 고도화된 AI가 인간보다 더 도덕적일 수 있다는 역설에 직면하게 된다. 그러나 Rousseau의 일반의지는 단순한 논리적 정답 찾기가 아니다. 그것은 타자의 고통에 공명할 수 있는 신체성과 공동체의 운명에 자신을 투기하는 주체적 결단을 요구한다. 즉, 일반의지는 차가운 '데이터'나 엄격한 '이성'의 통제를 넘어, 뜨거운 '마음'의 영역에 속한다.

이상과 같이 두 이론이 갖는 개념적 차이는 심리측정적 차원으로도 적용해 볼 수 있다. Kant의 정언명령을 심리학적으로 조작화한다면, 그 측정 차원은 도덕적 의무에 대한 인식, 자신의 행위 준칙을 보편적 법칙으로 검토하는 형식적 추론 능력, 경향성을 통제하고 의무를 이행하려는 자기극복적 동기 등으로 구성될 것이다. 즉 Kant의 도덕 모델에 대한 조작적 정의는 "보편화 가능한 도덕법칙을 얼마나 의무로서 인식하고 이행하는가"를 구체화하게 될 것이다. 반면 일반의지가 갖는 자기입법적 주체성은 의사결정의 외주화를 거부하고 스스로 도덕적 원칙을 입법하려는 능동적 동기와 자기 입법에 자율적으로 복종하려는 의지를 측정 차원으로 한다. 양자는 모두 자율적 자기입법의 이론적 토대를 공유하지만, 측정의 초점이 전자는 보편 법칙의 형식적 인식과 의무 이행에 있는 반면, 후자는 데이터리즘적 환경에서 입법 주체성을 능동적으로 발휘하려는 결단에 있다는 점에서 명확히 구별된다.

일반의지와 Rawls의 공정으로서의 정의

Rousseau의 일반의지가 지향하는 자유와 평등, 공공선의 이념은 Rawls의 공정으로서의 정의(justice as fairness)에 깊이 스며들어 있다. Rawls는 자신의 이론이 Kant 구성주의에 기초하고 있음을 명시적으로 밝히고 있는데(Rawls, 1995), 이는 그가 Rousseau로부터 시작되어 Kant로 이어진 '자율적 입법'의 전통을 계승하고 있음을 의미한다. Rousseau가 자유를 '자신이 스스로 제정한 법에 복종하는 것'으로 정의하고(Rousseau, 1762/2011), Kant가 이를 '이성이 입법한 도덕 법칙에 따르는 자율성'으로 정교화했다면(Kant, 1788/2009), Rawls는 이 자율적 도덕 주체의 개념을 사회 구조의 원리로 확장시켰다(Rawls, 1999/2003). 즉, 정의로운 사회의 원칙은 신의 계시나 자연의 질서 등 외부의 권위로부터 주어지는 것이 아니라, 자유롭고 평등한 이성적 존재들이 공정한 절차를 통해 스스로 구성해내는 것이어야 한다는 것이다. 이런 맥락에서 볼 때, Rawls의 기획은 Rousseau가 꿈꾸었던 '사회계약'을 21세기의 다원화된 민주주의 사회에 맞게 재설계하려는 시도라고 볼 수 있다. 이를 위해 두 사상가는 모두 사적 욕망이나 우연적인 사회적 지위, 타고난 재능 등 '나'라는 특수성을 배제하고 '모두'의 입장이 되어 보편타당한 원칙을 도출하려 한다는 점에서 동일한 문제의식을 공유한다고 볼 수 있다.

그렇다면, 서로 다른 이해관계와 가치관을 가진 현대인들이 어떻게 만장일치로 합의된 정의의 원칙을 도출할 수 있을까? Rousseau가 탈자연화를 목표로 한 교육을 통해 인간 본성

의 선함과 연민을 회복하는 '실존적 변화'에서 그 답을 찾으려 한 반면, Rawls는 인간의 불완전함을 인정한 채 합리적인 선택을 유도하는 정교한 사고실험을 제안하는데, Rousseau의 '자연상태'를 현대적으로 변주한 '원초적 입장'이 그것이다(Rawls, 1999/2003). Rawls는 최초로 정의의 원칙을 합의하는 원초적 상황에서 정의로운 원칙을 도출해내기 위해서는 '무지의 베일(Veil of Ignorance)'이라는 가상의 장막 뒤에 서야 한다고 가정한다. 이 베일 뒤에서 합의 당사자들은 자신이 사회에서 어떤 지위를 가질지, 타고난 재능이 무엇인지, 심지어 자신이 선호하는 가치관이 무엇인지조차 알 수 없는데, 이는 마치 Rousseau의 자연인이 사회적 계급이 발생하기 이전의 평등한 상태에 놓여 있는 것과 유사하다. 그러나 원초적 상황에 놓인 Rawls의 자연인은 Rousseau의 자연인처럼 내면의 평화로움 속에서 자족하며 살아가는 존재가 아니라 자신의 이익에 관심이 많은 합리적인 존재들로 가정된다. 이 불확실성의 상황에서 합의 당사자들은 매우 흥미로운 선택을 하게 되는데, 자신이 백만장자로 태어날지 최악의 조건에 놓인 빈민으로 태어날지 모르는 상황에서, 합리적인 인간이라면 모험을 걸기보다는 최악의 상황을 대비하려 할 것이다. 즉, 내가 가장 불운한 처지(최소 수혜자)에 놓이더라도 인간다운 삶을 보장받을 수 있는 안전장치를 선택하게 되는데, Rawls는 이를 '최소극대화 전략(Maximin strategy)'라 부른다. 결과적으로 무지의 베일은 이기적인 개인들로 하여금 자신이 가질 수 있는 사적 조건을 모르는 상태에서 선택하게 함으로써, 역설적으로 가장 공정하고 이타적인 결과(차등의 원칙, Difference Principle)를 도출해내는 장치가 된다. 이는 '자신을 생각하지 않고서는 전체를 생각할 수 없다'는 Rousseau(1762/2011)의 통찰을 절차적으로 구현해낸 셈이기도 하다.

Rawls의 논리는 빈틈없이 정교하지만, 동기적 차원에서 볼 때 Rousseau의 일반의지와는 명백하게 결을 달리 한다. 가장 핵심적인 차이는 합의 주체들의 심리적 태도에 있다. Rawls는 원초적 입장의 당사자들을 '상호 무관심한 합리성'을 가진 존재로 가정한다(Rawls, 1999/2003). 여기서 '상호 무관심'이란 타인을 증오한다는 의미가 아니라, 타인의 이익에 대해 질투하지도 않지만 그렇다고 헌신적인 사랑을 베풀지도 않는, 철저히 자기이익(self-interested)에 경도된 상태를 일컫는다. 원초적 입장에 놓인 합의 당사자들이 이타적 원칙(차등의 원칙)에 합의하는 이유는 최소 수혜자의 처지와 고통에 공감해서가 아니라, 그 최소 수혜자가 내가 될 수도 있다는 가능성 때문에, 즉 자신의 안전을 확보하기 위한 전략적 선택일 뿐이다. 따라서 Rawls 정의론의 기저에 깔린 핵심 정서는 타인에 대한 연대나 사랑이 아니라, 알 수 없는 미래에 대한 불안과 이를 통제하려는 회피동기라고 볼 수 있다. Rawls의 시민들은 서로 사랑하거나 연대하지 않아도 정의로울 수 있다. 그들은 단지 서로에게 해를 끼치지 않기로 약속한 영리한 계약자들이기 때문이다. 반면, Rousseau의 일반의지는 이러한 '차가운 계약'을 넘어선다. Rousseau에게 있어서 진정한 사회계약은 개인이 사적 자아의 껍질을 깨고 공동체적 자아로 거듭나는 마음의 변화, 즉 탈자연화를 전제로 한다. 이는 본성을 억압하는 것이 아니라, 자기중심적 자기애(Amour-propre)를 공동체 전체로 확장시킨 자기사랑(Amour de soi)으로 승화시키는 과정이다. 따라서 일반의지의 동력은 계산된 합리성이 아니라, 타인의 아픔을 나의 아픔으

로 느끼는 연민과 공동체를 향한 정서적 유대라 할 수 있다. Rousseau의 시민은 불안해서가 아니라, 나와 더불어 공동체를 사랑하기 때문에 일반의지에 복종한다.

이러한 심리적 기제의 차이는 정의의 실현 방식에 대한 근본적인 차이로 이어진다. Rawls에게 정의는 일차적으로 제도의 덕목으로서, 인간의 도덕성이 불완전하더라도 공정한 절차와 시스템(헌법, 법률)이 잘 설계되어 있다면 사회는 정의롭게 유지될 수 있다. 이는 도덕적 부담을 개인의 내면에서 외부의 제도로 옮겨놓은 것이다. 그러나 Rousseau에게 제도는 필요조건일 뿐 충분조건이 아니다. 아무리 좋은 법이 있어도 시민들의 마음에 그것을 따르고자 하는 '덕'이 없다면, 법은 종이조각에 불과하다. Rousseau는 "법은 대리석이나 청동이 아니라 시민들의 가슴 속에 새겨져야 한다"고 역설하며, 정의의 궁극적인 거처를 제도가 아닌 인간의 '마음'으로 보았다(Rousseau, 1997). 즉, 일반의지는 외부의 시스템이 아니라, 구성원 각자가 주체적 입법자로서 갖추어야 할 '마음의 습관'이자 실존적 결단을 요구한다. 결국 Rawls의 기획이 상호 무관심한 개인들이 공존하기 위한 최적의 제도를 만드는 것이라면, Rousseau의 기획은 파편화된 개인들을 하나의 신체로 묶어내는 실존적 작업이라 할 수 있다.

본 연구는 데이터리즘과 알고리즘이 인간의 합리적 선택을 대신해주는 21세기에, 계산과 절차에 의존하는 Rawls의 정의가 갖는 취약성을 지적한다. Rawls의 시민은 제도가 보장하는 안전을 위해 계약을 준수하지만, 공동체의 위기 상황에서 자신의 이익을 희생하며 헌신할 정서적 동기는 결여되어 있기 때문이다. 심리학적으로 환언하면, 이는 규칙의 정당성을 머리로 인정하는 인지적 판단(계약에 대한 합리적 동의)이, 곧바로 공동체를 위해 자신을 희생하는 실천적 행동(공공선의 능동적 추구)을 담보하지는 않음을 시사한다. 상호 무관심한 합리성은 공정한 분배를 계산해낼 수는 있어도, 타자와 공동체를 위해 행동하게 만드는 뜨거운 마음, 즉, 정서적 동기와 연대성을 제공하지 못하기 때문이다.

끝으로, Rawls의 공정으로서의 정의와 일반의지는 심리측정적 언어로 환언했을 때도 그 차이를 확인할 수 있다. Rawls의 공정으로서의 정의를 척도화하게 된다면, 절차적 공정성에 대한 인식, 무지의 베일과 같은 가상적 실험 상황에서의 분배 원칙 선호, 자신의 안전을 확보하기 위해 최소 수혜자의 처지를 고려하는 합리적 판단 등을 조작적으로 정의하게 될 것이다. 다시 말해 Rawls의 측정은 "공정한 절차와 분배 원칙에 얼마나 합리적으로 동의하는가"를 핵심 차원으로 하게 된다. 그러나 일반의지가 측정하고자 하는 바는 이와는 다른 차원이다. 그것은 정서적 연대성으로, 타인의 고통에 대한 정서적 공명을 토대로 한 공동체적 유대감의 정도를 측정 영역으로 삼는다. 결국 두 이론은 모두 공동체적 정의를 지향하지만, Rawls의 측정이 합리적 동의의 인지적 차원을 다루는 반면, 일반의지는 공동체에 대한 정서적 통합이라는 동기적 차원을 포착한다는 점에서 그 측정의 결이 명확히 구별된다.

일반의지와 Kohlberg의 도덕발달 5단계

Rousseau의 일반의지는 Piaget와 Kohlberg, Neo-Kohlbergian으로 이어지는 인지발달론의 이론적 원형을 제공한다. 이들은 도덕성을 외부 규범의 단순한 내면화가 아닌, 주체가 환경과

의 상호작용을 통해 스스로 구성하는 자율적 자기입법 과정으로 본다는 점에서 Rousseau와 Kant, Rawls와 근본적인 궤를 같이한다. 즉, Rousseau와 Kant, Rawls의 자율적 자기입법이 인간이 도달해야 할 도덕성의 규범적 원리를 철학의 차원에서 규명하고 있다면, Piaget와 Kohlberg, Neo-Kohlbergian의 도덕발달론은 자율적 자기입법의 원리에 도달하는 개인의 심리적 발달과정을 경험과학적으로 설명하는 관계라고 할 수 있다.

또한, 도덕 발달의 성숙한 지점을 '사회계약'으로 상정한다는 점에서도 두 철학적·심리학적 전통은 궤를 같이한다. Kohlberg의 5단계인 '사회계약 및 개인의 권리 지향'은 단순히 심리학적 발명품이 아니라, Rousseau에서 시작되어 Kant와 Rawls로 이어진 근대 철학의 '사회계약론적 전통'을 개인의 도덕 발달 단계로 조작화한 것이다. Kohlberg가 3수준 6단계의 도덕발달 단계 중 경험적 증명의 한계로 6단계를 가설적 단계라 인정한 바 있으며(Kohlberg, 1981), Neo-Kohlbergian은 5단계(후인습적 스키마)를 실질적인 도덕적 성숙의 최상위 지표로 설정했다는 사실은(Rest et al., 1999), 최종단계가 곧 발달적 목표가 된다는 점에서, '사회계약적 사유'가 인간의 도덕발달이 도달할 수 있는 최상위 수준임을 시사한다.

그러나 무엇보다도 Rousseau와 Piaget-Kohlberg를 잇는 가장 강력한 심리적 기제는 '탈중심화(decentering)'와 '가역성(reversibility)'에 있다. 강제의 도덕성에서 협동의 도덕성으로 이행하는 Piaget 도덕발달론의 핵심은 자기중심성에서 벗어나 타인의 관점을 수용하고 상호성을 인식하는 탈중심화에 있으며, Kohlberg의 경우, 각 단계가 반영하는 도덕추론 원리들은 최종단계에 가까워질수록 보편성이라는 완전한 가역성에 다가가게 된다(윤황, 2022, 2024a). Rousseau의 일반의지 역시 사적 욕망인 '개별의지'에 매몰된 자기중심성에서 벗어나 보편적 입법을 지향한다는 점에서 탈중심화를 요구한다. 특히 스스로 입법한 원칙이 타인뿐만 아니라 나 자신에게도 동일하게 적용됨을 수용하는 '가역적 사고'는 사익을 배제하고 일반의지를 형성하게 만드는 인지적 기초가 된다.

이상의 공통점에도 불구하고, Kohlberg 이론이 갖는 한계인 '판단과 행동의 괴리' 지점에서 일반의지와 Kohlberg 이론은 명백한 차이를 드러낸다(Blasi, 1980). 즉, Kohlberg의 5단계는 법이나 규칙 그 자체가 목적이 아니라, 구성원 간의 합의를 통해 얼마든지 변경 가능한 도구임을 인식하는 상태를 의미한다(Kohlberg, 1981, 1984). 그러나 규칙의 가변성을 인지하고 합의를 도출해내는 이 과정 역시 철저히 인지적 추론에 집중되어 있는데, 이는 앞서 논의한 Rawls 이론과 마찬가지로, 타인과 정서적으로 연대하지 않아도 합리적인 합의 규칙만 도출해 낸다면 정의롭다고 믿는 '차가운 계약'의 성격을 띤다. 실제로 Snarey(1985)의 교차문화 메타연구에 따르면, Kohlberg의 5단계 추론은 관계와 연대를 중시하는 전통적 집단주의 문화권에서는 거의 찾아볼 수 없었으며, 주로 서구화된 대도시 거주 고학력자들에게서만 제한적으로 발견된 바 있다(Edwards, 1986; Narvaez, 2005). 이는 Kohlberg의 5단계가 보편적인 도덕적 실천을 반영하는 완성된 지표라기보다는, 특정 문화권의 분석적이고 개인주의적인 인지 능력과 태도를 반영한 결과물에 가깝다는 한계를 보여준다(윤황, 2024a). 결국 이처럼 타인과의 정서적 유대가 배제된 상태에서의 분석적 인지가 갖는 치명적인 문제는,

아무리 정교한 정의 추론이라 할지라도 그것이 곧바로 도덕적 헌신을 담보하지는 않는다는 점이다. Kohlberg의 이론은 이성적 판단 능력은 설명할 수 있어도, 그 판단을 실천으로 옮기게 만드는 동기적 에너지를 간과하고 있다. Rousseau의 관점에서 볼 때, Kohlberg의 시민은 계산할 줄 아는 머리는 가졌으나 공동체를 위해 뜨겁게 박동하는 심장은 결여된 존재와 같다.

Kohlberg 이론이 갖는 한계를 극복하기 위한 시도로서, Neo-Kohlbergian은 도덕적 판단을 엄격한 단계 개념 대신 스키마(Schema)를 통해 설명하고, 인지적 측면에만 몰두하던 태도에서 벗어나 4구성 요소 모델(4 component model)을 통해 정서적, 동기적, 성격적 측면을 보강하고자 했다(Rest et al., 1999). 그러나 흥미로운 점은 이러한 시도가 본질적으로 Piaget 이론으로의 회귀를 의미한다는 사실이다(윤황, 2024b). Piaget는 “오직 인지적 요소로만 구성된 행동을 찾을 수 없는 것과 마찬가지로, 오직 정서에서만 비롯된 행동을 찾는 것 역시 불가능하다”면서, 인지와 정서의 불가분성을 주장한 바 있다(Piaget, 1981). Piaget에게 인지가 목적지를 향해 방향을 잡는 ‘구조(조향장치)’라면, 정서는 행동을 추동하는 ‘에너지(연료)’로서 이 둘은 동전의 양면처럼 평행하게 작동한다. 즉, Neo-Kohlbergian이 Kohlberg의 한계를 극복하기 위해 도덕적 민감성(정서적 요소)과 도덕적 동기화(동기적 요소)를 도입한 것은 Kohlberg가 임의로 절단해 버렸던 Piaget의 ‘통합적 도덕성’을 다시 복원하려는 시도라고 볼 수 있다. Neo-Kohlbergian은 이를 위해 도덕의 영역을 사회 시스템 차원의 공정성을 다루는 ‘거시도덕(Macro-morality)’과, 대인관계 차원의 배려와 성품을 다루는 ‘미시도덕(Micro-morality)’으로 구분함으로써 이론적 운신의 폭을 확장하고자 하였다(Rest et al., 1999). 그러나 일반의지의 관점에서 볼 때, 이 둘 간의 구분은 여전히 거시와 미시를 기능적으로 분리된 영역으로 취급하는 기계적 결합에 머물러 있다고 볼 수 있다. Rousseau의 일반의지는 이 두 층위를 완벽하게 통합한다. Rousseau에게 일반의지는 보편적 입법이라는 거시적 구조(인지)를 갖지만, 그 동력은 시민 개개인의 마음속에 자리 잡은 연민과 양심이라는 미시적 에너지(정서)에서 비롯된다. 미시적인 덕성 없이는 거시적인 정의가 결코 작동할 수 없다는 것이 Rousseau와 Piaget가 공유하는 통합의 본질이기도 하다.

이상의 논의를 종합할 때, 도덕적 인지에 몰두해 있는 Kohlberg 이론의 한계와 Neo-Kohlbergian의 기계적 분리가 남긴 공백을 극복하기 위해 본 연구에서는 Rousseau의 일반의지를 현대 도덕심리학의 새로운 통합적 대안으로 소환하게 된다. 이는 기존 심리학 이론들의 인지적 요소와 정서적 요소를 사후적으로 결합한 단순한 혼합물이 아니다. 보편적 입법이라는 거시적·인지적 지향과, 타인에 대한 연민 및 공동체적 연대라는 미시적·동기적 에너지가 애초부터 분리될 수 없는 하나의 유기체로 작동하는 심리적 기제 그 자체라 할 수 있다. 만일 도덕성의 궁극적 지향이 Kohlberg식의 정교한 공정성 추론과 절차적 합의(거시도덕)에만 머문다면, 복잡한 사회적 이해관계를 오차 없이 연산해 내는 고도화된 AI 알고리즘은 인간의 도덕적 주체성마저 대체해 나가게 될 것이 자명하다. 그러나 Rousseau의 일반의지는 이처럼 최적화된 합의 규칙을 산출해 내는 기계적 사유의 차원을 넘어선다. 일반의지는 인공지능이 결코 흉내 낼 수 없는

인간 고유의 영역 - 즉, 타인의 경험에 정서적으로 공명하며 공동체의 운명에 주체적으로 헌신하려는 '뜨거운 실천적 의지' - 을 인지적 추론과 유기적으로 통합해 낸 실체이기 때문이다. 나아가 일반의지 내에 구현된 이러한 거시와 미시의 심리학적 통합은 현대 윤리학의 오랜 난제인 '규범윤리와 덕윤리 간의 통합'에 대한 중요한 실마리를 제공하는데, 이에 대한 구체적인 논의는 본 연구의 후반부에서 다루게 될 것이다.

한편, 두 이론의 이러한 차이는 심리측정의 차원에서도 분명히 드러난다. Kohlberg의 도덕발달 이론은 도덕판단인터뷰(MJI; Moral Judgment Interview, Kohlberg, 1981)와 DIT (Defining Issues Test, Rest, 1979; Rest et al., 1999)를 통해 측정되어 왔으며, 이들 측정 도구는 도덕적 딜레마에 대한 추론 능력, 사회계약적 사유의 보편화 가능성, 탈중심화와 가역성을 갖춘 도덕 판단의 정교성 등을 핵심 측정 차원으로 삼아 왔다. 다시 말해 Kohlberg 전통의 측정은 "도덕적 추론의 인지적 성숙도가 어느 단계에 도달했는가"를 핵심 차원으로 다뤄 왔다고 할 수 있다. 그러나 일반의지가 측정하고자 하는 바는 이와는 다른 차원이다. 그것은 공공선 지향성과 정서적 연대성으로, 사적 이익을 넘어선 공동체적 안녕에 대한 능동적 지향과 타인의 고통에 대한 정서적 공명을 측정 영역으로 삼는다. 결국 두 이론은 모두 자율적 입법을 토대로 한 도덕성의 보편적 차원을 다루지만, Kohlberg 전통의 측정이 도덕 추론의 인지적 발달을 평가해 온 반면, 일반의지의 공공선 지향성과 정서적 연대성은 인지적 추론을 넘어 공동체적 헌신과 정서적 유대로 이어지는 동기적·정서적 통합의 차원을 반영한다는 점에서 그 측정의 결이 명확히 구별된다.

일반의지와 긍정심리학적 변인들

Peterson과 Seligman(2004)이 구축한 긍정심리학의 VIA(Values in Action) 분류 체계는 고대 덕이론에서의 담론을 현대 심리학에서 재구성하고 있다. 비록 VIA 체계가 고대 덕이론이 담고 있는 실천적 지혜의 통합적 성격이나 규범적 본질을 가치중립적인 심리적 특질로 단순 환원해버렸다는 비판이 있지만, 철학적, 형이상학적 개념으로서의 '덕(virtue)'을 6개의 핵심 덕목과 24개의 하위 성격 강점으로 체계화하여 심리학적 구성개념으로 조작화해낸 공로는 결코 가볍게 볼 수 없다. 또한 오랫동안 질병과 결핍 등 병리적 현상에만 매몰되어 있던 심리학의 관심을 인간 본성의 긍정적 측면과 잠재력으로까지 그 지평을 넓히는 데 기여한 바 있다.

Aristotle(연도미상/2013)는 『니코마코스 윤리학』을 통해 인간 삶의 궁극적 목적인 에우다이모니아(Eudaimonia)가 고립된 개인의 심리적 만족을 넘어서, 폴리스(Polis)라는 정치적 공동체 안에서 훌륭한 시민으로 기능하며 '공공선(Public Good)'을 실현할 때 비로소 완성된다고 보았다(Aristotle, 연도미상/2013). 즉, 고대 덕이론에서의 인간은 본질적으로 '정치적 동물'로 규정되며, 개인의 도덕적 완성은 공동체의 번성과 분리될 수 없는 상호 구성적인 관계를 맺는다. 따라서, Aristotle의 덕은 개인이 공동체 속에서 동료 시민들과 상호작용하며 정치적 삶에 참여하기 위한 실천적 지혜인 바, 공동체와의 조화와 정의를 지향하는 거시적이고 정치적인 성격을 가지고 있었다. 그러나 VIA 체계는 덕이론을 심리학적 변인으로 전환하는

과정에서 덕이 존재할 수 있는 조건인 '폴리스적 공동체성'을 제거해버리는 환원주의적 오류를 범하고 말았다. 이는 심리학적 구성개념으로서 VIA 체계가 지닌 태생적인 방법론적 전제에 기인한다. 긍정심리학은 덕목을 과학적으로 계량화하기 위해 '방법론적 개인주의'를 채택함으로써 덕을 개인의 덕성 및 성격 강점으로 국한시켰고, 문화적 보편성을 확보하기 위해 덕이 실현되는 현실의 정치적 역동이나 사회경제적 불평등, 그리고 제도적 체계와 같은 거시적인 구조적 맥락들을 거세해 버렸다(Fowers, 2005; Kristjánsson, 2013). 그 결과 덕의 궁극적 지향점마저 공공선의 실현에서 개인의 주관적 안녕감(Well-being)과 현실 적응으로 치환됨으로써, 본래 덕이 아우르고 있던 거시적 특성과 맥락은 제거되고 미시도덕적 측면만을 다루게 되었다.

이러한 덕성의 미시도덕화는 알고리즘이 개인의 의사결정과 가치 판단에 광범위하게 개입하는 오늘날의 환경에서 충분한 도덕적 대안이 되기에는 미흡할 수밖에 없다. 데이터가 개인의 편향을 강화하고 거대한 구조적 압력을 행사하는 현대 사회에서, 공동체성을 상실한 채 미시적 강점을 통해 주관적 안녕감을 추구하는 개인은 결국 알고리즘 시스템에 순응하는 '착한 부품'으로 전락할 위험을 안고 있기 때문이다(Harari, 2017/2023).

심리학적 구성개념으로서의 일반의지는 긍정심리학의 VIA 체계가 담아내고 있는 도덕적 미시성을 거시적이고 주체적인 차원에서 재통합할 수 있는 대안이 될 수 있다. 이는 VIA 체계가 제시하는 24개의 강점들 중 일반의지와 직관적으로 유사하게 느껴지는 변인들과 일반의지의 핵심 작동 원리를 비교해 보면 선명하게 드러난다. 이 변인들은 표면적으로 일반의지와 공통된 도덕적 지향을 공유하는 듯 보이나, 그 심층적 기제와 작동 범위에서 명확한 차이를 갖고 있기 때문이다.

첫째, 인지적·의지적 차원에서 일반의지의 '자기입법적 주체성'은 VIA 체계의 '자기조절(Self-regulation)'과 표면적인 유사성을 갖는다. 두 개념 모두 외부의 자극이나 즉각적인 충동에 휩쓸리지 않고, 내면의 통제력을 발휘하여 스스로의 행동을 규제한다는 점에서는 공통된 심리적 작용을 전제로 한다. 그러나 VIA 체계의 자기조절은 Peterson과 Seligman(2004)의 정의가 시사하듯, 목표 달성과 규범 준수를 위해 자신의 충동·욕구·수행을 통제하는 개인적 차원의 행동 관리에 초점이 맞추어져 있다(Baumeister et al., 2007; Niemiec, 2018). 반면 일반의지가 담고 있는 자기입법의 작동 범위는 이를 넘어선다. 그것은 외부 유혹이나 즉각적 이익을 거부하고 스스로 설정한 보편타당한 도덕적 원칙에 자발적으로 복종하는 규범적 결단력을 의미한다. 즉, 단순한 충동의 억제가 아니라 공동체의 주권자로서 스스로 입법한 원칙에 따라 판단하고 행동하고자 하는 주체적 행위자성의 발현인 것이다. 따라서 VIA 체계의 자기조절이 개인의 목표 달성을 위한 도구적 자기통제의 영역에 속한다면, 일반의지의 자기입법적 주체성은 공동체적 가치를 실현하기 위해 스스로 제정한 도덕원칙에 자율적으로 복종하는 규범적 실천의 영역에 속한다.

둘째, 정서적 차원에서 일반의지의 '정서적 연대성'은 VIA 체계의 '친절(Kindness)' 및 '사랑(Love)'과 유사한 맥락을 공유하는 것처럼 보인다. 타인을 배려하고 인간관계를 중시하며 이타적인 태도를 보인다는 점에서 이들은 공통된 정서적 지향을 갖는다. 그러나 긍정심리

학의 해당 변인들은 나와 타인이 분리된 상태에서 베푸는 시혜적 온정이거나, 혹은 Gilligan의 배려지향적 도덕이론이 설명하듯 사적이고 친밀한 관계망 안에서의 윤리에 집중한다(Gilligan, 1982/1997). 반면 일반의지가 요구하는 정서적 연대성은 특정 개인에 대한 사적 친밀감을 넘어서는 '확장된 사회적 연대'를 의미한다. 이는 타인의 고통을 관찰자의 입장에서 동정하는 것이 아니라, 공동체 전체를 자신의 확장된 자아로 받아들이고 타인의 고통에 존재론적으로 공명하는 심층적 유대감이다. 이러한 차이는 연대의 범위와 기제를 통해 명확하게 드러난다. VIA 체계의 친절과 사랑이 나와 관계 맺는 특정 타인을 향한 정서적 반응이라면, 일반의지의 정서적 연대성은 공동체 전체를 자신의 정체성 안으로 통합하는 고차원적 상호연결에 기반한다. Rousseau의 자연인이 지닌 연민이 특정 대상이 아닌 타자 일반의 고통에 반응하는 원초적 기제인 것처럼, 일반의지의 정서적 토대는 사적 관계망을 초월하여 공동체적 수준에서 작동한다(임의영, 2020; Rousseau, 1762/2011).

셋째, 행동적 차원에서 일반의지의 '공공선 지향성'은 VIA 체계의 '시민정신(Citizenship)' 및 '공정성(Fairness)'과 밀접하게 연관되어 있는 듯한 인상을 준다. 공동체나 집단 내에서의 역할 수행, 규칙 준수, 타인에 대한 동등한 대우를 강조한다는 점에서 이들은 유사한 행동적 지향성을 공유한다. 그럼에도 불구하고 VIA 체계의 시민정신은 소속된 내집단에 대한 충성이나 기존 규범 체계에 대한 순응으로 수렴될 위험이 있으며, 공정성 또한 감정을 배제한 기계적, 절차적 평등에 머무를 수 있다. Rousseau의 개념으로 환언하면, 이는 사적 이해관계의 단순한 합인 전체의지 수준에 해당하며, 다수결의 결과에 무비판적으로 수렴될 수 있다(Rousseau, 1762/2011). 반면 일반의지는 특정 집단의 사익을 배제하고 공동체 구성원 모두에게 이익이 될 수 있는 공공선만을 지향한다(임의영, 2020; Rousseau, 1762/2011). VIA 체계의 시민정신이나 공정성이 현행 규범과 집단 내 역할에 대한 동조를 강화하는 방향으로 작동할 수 있다면, 일반의지는 기존 규범의 정당성 자체를 스스로 숙의하고 판단하는 주체적 입법 과정을 포함한다는 점에서 차별화된다. 요컨대, VIA 체계의 시민정신과 공정성이 주어진 사회적 맥락 안에서의 도덕적 적응에 가깝다면, 일반의지의 공공선 지향성은 그 맥락 자체를 비판적으로 성찰하고 보편적 가치에 부합하는 방향으로 재구성하려는 능동적 도덕 지향을 의미한다.

이상의 비교에서 확인되듯이, VIA 체계의 자기조절, 친절, 사랑, 시민정신, 공정성은 표면적으로 일반의지와 유사한 도덕적 지향을 공유하는 것처럼 보이지만, 도덕적 지향의 범위와 작동 기제에서 명확히 구분된다. VIA 체계는 덕목을 개인의 성격 강점으로 환원함으로써 공동체적 맥락을 탈각하고, 도덕적 실천을 위한 주체적 입법의 동력을 충분히 담아내지 못한다.

이러한 변별은 심리측정적 차원에서도 동일하게 확인된다. VIA 체계는 VIA-IS(Values in Action Inventory of Strengths; Peterson & Seligman, 2004)를 통해 측정되어 왔으며, 24개의 성격 강점을 개인의 성격 특질로 조작화하여 평가하는 방식으로 자리 잡아 왔다. 다시 말해 VIA의 측정은 "개인이 어떤 성격 강점을 얼마나 보유하고 있는가"를 핵심 차원으로 다뤄 왔다고 할 수 있다. 그러나 일반의지가 측정하고자 하는 바는 이와 다른 차원이다. 일

반의지를 구성하는 자기입법적 주체성, 정서적 연대성, 공공선 지향성은 개인이 보유한 성격 특질이 아니라 공동체적 맥락 안에서 발현되는 도덕적 지향의 정도를 측정 영역으로 삼는다. 결국 두 이론은 모두 도덕성의 긍정적 측면을 다루지만, VIA의 측정이 개인 차원의 성격 강점을 평가하고 있다면, 일반의지의 세 하위요인은 공동체적 맥락 안에서의 능동적 도덕 지향이라는 거시적·통합적 차원을 포착한다는 점에서 그 측정의 결이 명확히 구별된다.

심리적 구성개념으로서의 일반의지(GWO)와 정신병질

앞선 논의들을 통해 Kant의 정언명령, Rawls의 공정으로서의 정의, Kohlberg의 도덕발달이론, 긍정심리학의 VIA 체계와 구별되는 심리학적 구성개념으로서의 독자성을 확인할 수 있었다. 그러나 이것만으로는 충분하지 않다. 어떤 개념이 심리학적 구성개념으로서 실체성을 인정받기 위해서는, 이론적 독창성을 넘어서 그것이 실제로 측정 가능한 변인임을 논증해야하기 때문이다. 이 지점에서 인간 도덕성의 부정적 극단을 측정하는 심리학적 구성개념으로 자리 잡아온 정신병질(Psychopathy)과의 비교는 중요한 이론적 단서를 제공한다.

정신병질이 심리학적 구성개념으로서 실체성을 획득하게 된 과정은 일반의지의 심리학적 실체화 가능성을 논증하는 데 중요한 선례로 기능한다. 정신병질 개념을 처음으로 체계적으로 기술한 것은 Cleckley(1941)의 저서 『The Mask of Sanity』인데, 그는 반복적으로 반사회적 행동을 저지르면서도 표면적으로는 매력적이고 지적인 외양을 유지하는 환자군을 관찰하고, 공감 능력의 결여, 죄책감 부재, 피상적 매력, 병리적 거짓말 등 16개의 임상적 특성들을 기술한 바 있다. 그러나 Cleckley의 기술은 임상적 직관과 질적 관찰에 의존한 것으로, 정신병질은 개연성 있는 개념으로서 광범위한 동의를 얻었음에도 불구하고 객관적으로 측정 가능한 심리학적 실체로서의 지위는 오랫동안 유보 상태에 머물러왔었다.

이러한 상태를 종식시킨 것이 Hare(1991)의 정신병질 체크리스트 개정판(Psychopathy Checklist-Revised: PCL-R)이다. PCL-R은 Cleckley의 임상 기술을 토대로 구성된 20개 문항의 반구조화된 면접 및 자료 검토 도구로서, 엄격한 신뢰도와 타당도 검증을 거쳐 정신병질을 측정 가능한 심리학적 구성개념으로 확립하였다. 이후 PCL-R에 대한 요인 분석 연구들은 정신병질의 내부 구조에 대한 수렴된 이해를 가능하게 했는데, 정서적 공감의 결여, 피상적 대인관계, 타인의 도구적 취급 등을 반영하는 Factor 1(정신병질 요인)과 정서조절 실패 및 충동성, 행동 통제 결여 등을 반영하는 Factor 2(사회병질 요인)의 2요인 구조가 널리 수용되었다(Hare, 1991, 2003; Harpur et al., 1998). 여기서 주목해야 할 것은 개념과 측정 도구 간의 관계이다. 정신병질은 Cleckley 단계에서 이미 임상적으로 유의미한 개념이었으나, PCL-R이 등장하기 전까지는 반복 가능하고 비교 가능한 연구의 토대가 될 수 없었다. 측정 도구의 개발이 곧 개념의 학술적 실체성을 완성한 셈이다. 이 선례는 일반의지의 심리학적 실체화 가능성을 논증하는 데 결정적인 함의를 갖는다. 다만 여기서 한 가지 주목할 비대칭성이 있는데, 정신병질은 임상 현장의 귀납적 관찰에서 출발하여 이론적 체계화가 이후

표 2. GWO 세 하위요인의 조작적 정의와 측정영역

하위요인	조작적 정의	측정영역
공공선 지향성	사적 이익을 넘어 공동체 구성원 모두에게 이익이 될 수 있는 방향을 능동적으로 숙고하고, 타인의 안녕을 자신의 도덕적 고려 안으로 통합하려는 지향성	사익과 공익이 충돌하는 의사결정 상황에서 공동체적 결과를 우선 고려하는 경향의 정도, 자신의 선택이 공동체 전체에 미치는 영향에 대한 책임감 인식의 정도, 공공의 이익에 대해 능동적으로 숙고하는 정도
자기입법적 주체성	외부의 유혹이나 알고리즘적 편의에 의사결정권을 외주화하지 않고, 스스로 입법한 도덕적 원칙에 자율적으로 복종하는 능동적 행위자성	알고리즘 추천이나 외부 권위에 도덕적 판단을 위임하지 않으려는 경향의 정도, 자신의 도덕적 원칙을 스스로 수립하고 진술하는 능력의 정도, 즉각적 욕구와 자기 입법한 원칙이 충돌할 때 후자를 따르려는 자율적 자기조절의 정도
정서적 연대성	타인의 고통에 존재론적으로 공명하고, 공동체 전체를 자신의 확장된 자아로 받아들이는 심층적 유대감	친밀한 관계망을 넘어선 낯선 타자의 고통에 대한 정서적 공명의 정도, 공동체의 안녕을 자신의 안녕과 연결되어 있다고 인식하는 정도, 공동체 구성원과 정체성을 공유하는 심층적 유대 경험의 정도

에 이루어진 반면, 일반의지는 Rousseau의 사회계약론과 에밀이라는 명확한 철학적 원전과 정합하는 이론적 서사를 이미 갖추고 있다는 점이다. 즉, 일반의지는 Cleckley 단계에 해당하는 개념적 정초 작업을 이미 완료한 상태이며, 현재 남겨진 과제는 PCL-R에 해당하는 측정 도구의 개발이다. 본 연구에서는 이를 위해 Rousseau의 일반의지를 심리학적으로 측정 가능한 구성개념으로 전환한 일반의지 지향성(General Will Orientation: GWO)을 제안한다. GWO는 정치철학적 개념으로서 형이상학적 담론의 영역에 머물러 있던 일반의지를 경험과학적 연구가 가능한 심리학적 변인으로 실체화하기 위한 개념적 전환의 산물이라 할 수 있다. GWO를 구성하는 세 하위요인의 조작적 정의와 측정 영역은 표 2와 같다.

정신병질과 일반의지를 심리학적 구성개념의 차원에서 대비하는 작업은 단순한 수사적 대칭이 아니다. 이는 인간 도덕성의 지형도를 보다 완전하게 그려내기 위한 이론적 요청이다. PCL-R의 탄생 이후 정신병질은 사회적 유대의 해체와 자기중심성을 반영하는 심리학적 지표로서, 비윤리적 행동 및 반사회적 행동과 관련한 다양한 개인차 연구의 핵심 변인으로 기능해 왔다. 그러나 심리학은 오랫동안 이 부정적 극단에만 주목해 온 나머지, 그 반대편 극단에 해당하는 심리적 구성개념을 결여해 왔다. 인간이 얼마나 악할 수 있는가를 측정하는 도구는 정교하게 발달하였으나, 인간이 얼마나 선(善)을 지향하며 행동할 수 있는가를 하나의 통합된 심리학적 변인으로 포착하는 구성개념은 부재한 상태였다. 앞선 논의에서 확인한 바와 같이, 친사회적 행동, 이타성, 긍정심리학적 덕목들이 부분적으로 이 공백을 채워 왔으나(Penner et al., 2005), 이들은 미시도덕적 수준에 머물거나 공동체적 맥락과

주체적 입법 동기를 충분히 포함하지 못한다는 한계를 공유한다. GWO는 이 한계를 극복하기 위한 심리학적 구성개념으로서 제안된다.

정신병질의 두 요인이 드러내는 도덕성의 부정적 특성들은 GWO의 세 하위요인과 구조적으로 대응하며, 이 대응 관계를 통해 GWO 각 하위요인의 심리학적 실체성이 보다 선명하게 드러난다. 이를 정리해 보면 표 3과 같다(Krueger & Markon, 2006; Yun, 2018).

먼저, Factor 1이 반영하는 공감 능력의 결여와 타인의 도구적 취급은 공동체적 선(善)의 감각 자체의 부재를 의미한다. 정신병질적 개인은 타인의 안녕을 자신의 도덕적 고려 대상으로 포함시키지 못하며, 공동체의 이익보다 사적 이익의 극대화를 일관되게 선택한다. 이는 GWO의 공공선 지향성(Public Good Orientation)과 정면으로 대칭된다. 공공선 지향성은 구성원 모두에게 이익이 될 수 있는 방향을 스스로 숙고하고 그 방향으로 행동하려는 도덕적 지향성으로서, 타인의 안녕을 나의 도덕적 고려 안으로 통합하는 능동적 지향을 반영한다. Factor 1이 이 지향의 완전한 부재를 드러낸다면, 공공선 지향성은 그 지향의 능동적 발현을 보여준다.

다음으로, Factor 2가 반영하는 정서조절 실패에 따른 충동적 탈억제는 GWO의 자기입법적 주체성(Self-legislative Agency)과 대칭된다. 정신병질적 개인은 즉각적인 욕구와 충동에 의해 행동하며, 스스로 세운 원칙이나 공동체의 규범에 자율적으로 복종하는 내적 기제가 결여되어 있다. 반면 자기입법적 주체성은 외부의 유혹이나 알고리즘적 편의에 의사결정권을 외주화하지 않고 스스로 입법한 도덕적 원칙에 자율적으로 복종하는 내적 규율을 반영한다. Factor 2가 이 내적 규율의 붕괴를 드러낸다면, 자기입법적 주체성은 그 규율의 능동적 확립을 의미한다.

나아가 정신병질의 두 요인을 관통하는 근본적 특성은 사회적 유대의 해체이다. Factor 1의 자기중심성과 Factor 2의 정서조절 실패에

표 3. 정신병질 요인 구조와GWO 하위요인의 대응 관계

정신병질 Factor	핵심 행동지표	GWO 대응요인	대응 행동지표
Factor 1 (정신병질 요인) : 자기중심성, 도구적 위반행동	공감 결여, 냉담함, 후회·죄책감 결여, 타인의 도구적 취급, 기생적 생활양식, 얕은 정서	공공선 지향성	타인의 안녕에 대한 정서적 고려, 공동체적 결과에 대한 책임감, 공공선을 향한 숙고
Factor 2 (사회병질 요인) : 정서조절 실패, 반응적 위반행동	충동성, 행동통제 결여, 무책임성, 자극추구, 장기 계획 결여	자기입법적 주체성	자율적 자기조절, 외부유혹에 대한 내적 통제, 스스로 입법한 원칙에 대한 자발적 복종
두 요인 관통 (사회적 유대 해체)	사회적 일탈, 공동체로부터의 단절	정서적 연대성	타인의 고통에 대한 정서적 공명, 공동체 구성원과의 심층적 유대, 확장된 자아로서의 공동체 인식

따른 충동성은 모두 타인과의 사회적, 정서적 유대를 단절시키는 방향으로 수렴된다. GWO의 정서적 연대성(Emotional Solidarity)은 이 단절의 정확한 대척점에 위치한다. 정서적 연대성은 타인의 고통에 존재론적으로 공명하고 공동체 전체를 자신의 확장된 자아로 받아들이는 심층적 유대감으로서(Batson, 2011), 정신병질이 반영하는 사회적 유대의 해체에 맞서 그 재구성을 가능하게 하는 심리적 기제이다.

주목할 점은 앞서 일반의지와 철학적 유사개념 및 심리학적 유사 변인들을 비교하는 과정에서 기술한 세 가지 차원의 논거 - 인지적·의지적 차원, 정서적 차원, 행동적 차원 - 가 GWO의 세 하위요인인 자기입법적 주체성, 정서적 연대성, 공공선 지향성에 각각 대응하고 있다는 사실이다. 이는 GWO의 하위요인 구성이 임의적인 것이 아니라 본 연구의 변별 논리 전반을 관통하는 일관된 이론적 서사 위에서 도출된 것임을 보여준다. 그리고 이러한 이론적 정합성 위에서 살펴볼 때, 정신병질과 GWO의 이상과 같은 대응은 단순한 수사적 대칭이 아니라 동일한 행동 영역 위에서 반대 방향으로 작동하는 측정학적 대칭의 구조를 갖는다. 즉, PCL-R이 측정해 온 정신병질의 행동지표와 GWO가 측정하고자 하는 대응 행동지표는 공통된 행동 차원 위에서 그 발현 방향이 반대일 뿐인 바, 이는 두 구성개념이 인간 도덕성이라는 동일 측정 공간의 양극을 이루고 있음을 의미한다.

정신병질이 악(惡)을 반영하는 심리적 구성개념으로 자리 잡은 것처럼, GWO는 선(善)을 반영하는 심리적 구성개념으로서 정립될 필요가 있다. 인간의 도덕적 행동을 온전히 이해하기 위해서는 도덕성의 부정적 극단만이 아니라 긍정적 극단을 포착하는 구성개념이 함께 요청된다. 정신병질이 악의 심리학적 실체를 규명해 왔다면, GWO는 선의 심리학적 실체를 규명하기 위한 이론적 토대를 제공한다. 여기서 선은 형이상학적 관념으로서의 선이 아니라, 공동체 구성원이 자율적 숙의를 통해 합의할 수 있는 공공선-즉 앞서 논의한 Rousseau의 일반의지가 지향하는 바로 그 선-을 의미한다. GWO는 이처럼 합의 가능한 선을 향한 심리적 지향을 측정 가능한 변인으로 포착한다는 점에서, 형이상학적 당위가 아닌 경험과학적 실체로서 자리매김한다. 그리고 정신병질과 GWO가 인간 도덕성의 양쪽 극단을 함께 다뤄나갈 때, 비로소 도덕적 행동에 대한 보다 균형 있고 완전한 심리학적 이해가 가능해진다.

규범윤리와 덕윤리 간의 통합적 대안으로서의 일반의지

당신은 지병으로 병원에서 오랫동안 투병생활을 이어오고 있는 중이다. 몹시 지루하고 불안해하고 있던 차에 스미스가 다시 병문안을 왔다. 당신은 그가 당신의 병문안을 위해 멀리서 오는 수고를 번번이 마다하지 않는 것을 보면서 그가 정말 좋은 사람이고 진정한 친구라고 굳게 믿는다. 당신이 너무 감격해서 스미스에게 칭찬과 감사를 쏟아내자 그는 손사래를 치면서 그저 자신의 의무라고 생각하는 일, 가장 최선이라 생각하는 일을 했을 뿐이라고 말한다. 처음엔 그가 겸손한 사람이라 당신의 마음을 덜어주려 한다고 생각했으나, 대화를 나눌수록 그가 진실을 말하고 있음이 점점 더 분명해진다. 스미스가 당신을

> 만나러 온 것은 당신을 좋아해서도 당신과 친구이기 때문도 아니다. 단지 옳은 일을 하는 것이 자신의 의무라고 생각했기 때문인데, 어쩌면 그는 당신보다 더 위로가 필요한 사람이나 더 쉽게 찾아갈 수 있는 사람을 알지 못했기 때문에 당신을 찾아 왔을 수 있다(Stocker, 1976).

이상은 Michael Stocker(1976)의 논문 『The schizophrenia of modern ethical theories』에서 소개되는 유명한 사례이다. 이 장면에서 스미스의 행위엔 잘못된 것이 없다. 문제는 그의 동기이다. 우리는 타인의 곤란에 공명을 느끼고, 곤란에 처한 타인과 기꺼이 함께 하고자 하는 사람을 원한다. Stocker는 '조현병'이라는 비유를 통해 현대윤리학 이론엔 도덕적 행위를 정당화하는 이유와 그 행위를 실제로 추동하는 동기 사이에 깊은 균열이 있음을 지적하고 있다. 규범윤리학은 무엇이 옳은지를 설명하지만, 왜 그 옳음이 행위자의 내면에서 자발적으로 흘러나와야 하는가에 대해서는 침묵한다. 이하에서는 이 균열을 중심으로 규범윤리와 덕윤리를 각각 검토하고, 일반의지가 두 입장의 통합적 대안으로서 어떻게 기능할 수 있는가를 논증한다.

규범윤리의 두 축: 의무론과 공리주의

앞서 논한 바와 같이, 규범윤리의 한 축인 Kant의 의무론은 경향성을 도덕의 토대에서 배제함으로써 도덕적 순수성을 확보한다. 그러나 '해야 한다'는 의무가 '하고 싶다'는 경향성을 극복할 것을 요구한다는 점에서, 행위자에게 지속적인 내적 긴장을 부과하는 심리학적 한계를 남긴다. Stocker의 병원 사례는 이 한계의 핵심을 드러낸다. 친구를 의무감에서 돕는 것이 친구를 사랑해서 돕는 것보다 도덕적으로 더 순수하다는 논리는, 도덕적 행위의 관계적 맥락과 그것을 의미 있게 만드는 정서적 동기를 이론의 외곽으로 밀어내기 때문이다.

Bentham(1789/1996)과 Mill(1863/2002)의 공리주의는 의무론의 반대 방향에서 도덕적 기준을 제시한다. 도덕적으로 옳은 행위란 결과로서 산출되는 행복의 총량을 극대화하는 행위이며, 의도나 동기가 아닌 결과가 도덕 판단의 준거가 된다. 정책 결정이나 의료 윤리의 영역에서 공리주의적 추론은 여전히 유효한 규범적 도구로 기능한다. 그러나 투입과 산출을 계량화하고 최적의 결과를 산출하는 공리주의의 논리는 사용자 만족을 극대화하기 위해 설계된 추천 알고리즘의 논리와 구조적으로 유사하다. Harari (2017/2023)가 경고한 것처럼 데이터리즘의 핵심은 인간 경험을 데이터 흐름으로 환원하고 처리 효율을 극대화하는 것인 바, 행복을 데이터화할 수 있다면 공리주의의 행복 극대화 원리는 알고리즘에 의한 도덕적 최적화와 원리적으로 구별되지 않는다.

따라서 의무론과 공리주의는 도덕을 외재적 기준에 의한 행위의 평가 체계로 구성한다는 공통점을 지닌다. 그 결과 두 이론 모두 도덕적 행위를 추동하는 내면의 동기 - 특수한 타자에 대한 관심, 공동체에 대한 헌신, 선을 향한 자발적 지향 - 를 도덕 이론의 주변부로 밀어낸다. Ryan과 Deci(2000)의 자기결정성 이론(SDT)은 이 한계를 심리학적으로 포착한다. 의무나 결과 때문에 하는 도덕적 행위는 외재적 규제 수준에 머물며, 심리적 건강과 지속적 친사회적 행동에 미치는 영향이 제한된다. 행위 자체가 자신의 가치와 통합되어 자발적으

로 흘러나오는 동기만이 공동체적 연대와 장기적 헌신을 가능하게 한다는 것이다.

덕윤리와 배려의 윤리: 도덕적 동기의 복원

Aristotle는 『니코마코스 윤리학』에서 도덕의 핵심 물음을 행위의 규칙이 아닌 행위자의 존재 방식으로 전환한다. "무엇을 해야 하는가?"가 아니라 "어떤 사람이 되어야 하는가?"가 윤리학의 근본 질문이 된다. 덕은 올바른 상황에서, 올바른 방식으로, 올바른 정도로 느끼고 행동하는 성향이다(Aristotle, 연도미상/2013). 도덕적으로 탁월한 사람(phronimos)은 의무나 결과의 계산을 통해 도덕적 행위에 도달하는 것이 아니라, 그것이 자신의 존재 방식과 일치하기 때문에 그렇게 행동한다. 진정한 친구(philia)는 우정의 의무를 계산하는 자가 아니라, 타자의 선을 자신의 선과 동일시하는 자이다(Aristotle, 연도미상/2013). 이는 도덕적 행동이 규칙의 내면화가 아닌 자아 개념과의 통합 여부에 따라 결정됨을 보여주는 Bandura(1986)의 사회인지이론과도 부합한다.

Gilligan(1982/1997)의 배려의 윤리는 덕윤리가 갖는 관계적 차원을 심화한다. Gilligan은 Kohlberg의 도덕발달 이론이 정의와 권리의 언어에 편향되어 관계와 반응성이 담당하는 도덕적 역할을 체계적으로 배제한다고 비판한다. 배려의 윤리는 세 가지 전제에서 출발한다. 첫째, 자아는 관계 속에서 구성된다. 둘째, 도덕적 판단은 추상적 원리의 연역이 아닌 구체적 맥락에 대한 민감한 반응이다. 셋째, 타자의 필요를 인식하고 반응하는 실천으로서의 배려가 도덕의 핵심 덕목이다(Gilligan, 1982/1997). 이는 Stocker의 병원 사례에 직접적인 응답을 제공한다. 그 친구의 방문이 도덕적으로 불완전하게 느껴지는 이유는, 아픔 속에서 연결을 필요로 하는 타자에 대한 감응으로부터 비롯된 것이 아니기 때문이다. 타자의 고통에 대한 감응적 반응이 도덕적 행위의 핵심 동기임은 Hoffman(2000)과 Eisenberg(2000)의 공감 발달 연구들을 통해서도 확인된다.

그러나 심리학 이론들과의 이론적 부합에도 불구하고, 덕윤리와 배려의 윤리는 또 다른 한계를 갖는다. 덕윤리는 어떤 것이 진정한 덕인지를 규정할 보편적 기준이 불분명하여 공동체마다 다른 덕의 목록을 제시한다면 결국 윤리적 상대주의로 귀결될 수밖에 없다(Hursthouse, 1999; MacIntyre, 1981). 배려의 윤리는 구체적 관계에 대한 반응성을 강조하지만, 특수한 관계(가족, 친구, 내집단)를 넘어 낯선 타자와 더 넓은 공동체로 배려를 확장하는 논리가 취약하다. 요컨대, 규범윤리는 보편성을 확보하지만 도덕적 동기의 내면성을 결여하고, 덕윤리는 내면성을 복원하지만 보편성의 기준이 불분명하다. 이 두 한계를 동시에 극복하는 것이 일반의지의 과제이다.

유덕한 시민과 GWO: 규범윤리와 덕윤리의 통합적 대안

앞서 일반의지의 이론적 서사에서 검토한 바와 같이, Rousseau(1762/2011)는 "일반의지는 항상 옳으며 공공의 유익을 지향한다"고 선언한다. 일반의지가 갖는 이러한 무오류성은 종종 형이상학적 과잉이나 전체주의적 함의의 근거로 비판받아왔으나, 심리학적 구성개념의 맥락에서는 다른 방식으로 재해석될 필요가 있다.

Rousseau의 무오류성은 '집합적 의사결정이 항상 옳은 결과를 산출한다'는 경험적 주장이

아니라 구성적(constitutive) 주장이다(Searle, 1995). 일반의지란 정의상 공공선을 지향하는 의지이다. 구성원 각자가 사적 이익을 넘어 공동체 전체에 무엇이 참된 선인가를 숙고할 때, 그 숙고의 성격상 표출되는 의지는 공공선을 향할 수밖에 없다. 만약 의지가 공공선을 향해 있지 않다면 그것은 일반의지가 아니라 특수의지이다(임의영, 2020). 즉, 무오류성은 일반의지가 항상 공공선을 가리키는 나침반임을 의미하며, 나침반 자체가 항상 북쪽을 향하듯 일반의지는 그 구성적 정의상 공공선의 방향을 향해 있다. 이 구성적 무오류성은 규범윤리와 덕윤리 간의 통합 원리로서 다음과 같은 함의를 가진다. 규범윤리와의 관계에서, 무오류성은 공공선이라는 보편적 방향성이 일반의지의 내재적 조건이라는 점에서 규범윤리의 보편성 요구를 충족한다. 또한 이 무오류성은 공공선을 향한 지향이 외부에서 부과된 법칙이 아니라 자기입법적 주체에 의한 것이라는 점에서, 덕윤리의 통찰('선한 사람은 선한 것을 하고 싶어한다')과 공명한다.

Rousseau에게 있어서 덕은 단순한 선함이 아니라, 자신의 사적인 욕망(개별의지)을 억제하고 일반의지에 일치시키는 의지의 힘이다(Rousseau, 1755/2003a, 1762/2011). 그가 일컫는 '유덕한 시민'은 개별의지와 일반의지가 일치하는 인간상으로서 유덕한 시민만이 진정으로 자유롭다. 이 개념은 규범윤리와 덕윤리의 긴장을 '내가 하고 싶은 일이 곧 선한 일일 수 있는가?'라는 하나의 질문으로 압축한다. 이에 대한 규범윤리의 답은 부정적이다. 욕구와 의무의 분리가 도덕의 조건이기 때문이다. 덕윤리의 답은 조건적 긍정이다. 올바른 것을 하고 싶어 하도록 성품이 형성된 사람에게는 가능하기 때문이다. 그러나 이 답은 어떤 성품이 올바른 성품인지를 규정하는 보편적 기준의 문제를 미해결로 남긴다. Rousseau의 유덕한 시민 개념은 이 물음에 대한 제3의 답을 제시한다. 그것은 올바른 성품이 갖는 주관성과 상대성의 문제를 공공선이라는 보편적 기준을 통해 해결한다. 개별의지와 일반의지가 일치하는 사람이 하고 싶어 하는 일은 공공선을 벗어나지 않거나 이미 공공선을 전제로 하고 있다(Annas, 2011). 따라서 유덕한 시민은 규범윤리의 보편성(일반의지라는 공공선의 방향)과 덕윤리의 내면성(개별의지와의 자발적 일치)을 동시에 구현한다(Blasi, 1984).

GWO는 Rousseau의 유덕한 시민 개념을 심리학적 구성개념으로 실체화한 것이다. GWO는 공동체의 보편적 가치를 지향하며 스스로 입법한 도덕적 명령에 자율적으로 복종하려는 동기적 상태로 정의될 수 있다. 이 정의에서 보편적 가치 지향은 규범윤리로부터, 자율적 복종의 내면성은 덕윤리로부터 각각 계승된다.

끝으로 유덕한 시민 개념은 GWO의 양적 측정 가능성에 대한 이론적 근거를 제공한다. Rousseau가 말하는 개별의지와 일반의지의 일치는 이분법적 상태가 아니라 양적 개념으로서의 연속적 스펙트럼이다. 어떤 사람은 사적 욕망이 공공선의 방향과 높은 수준으로 일치하고, 어떤 사람은 그 간극이 클 수 있다. 이 일치의 정도, 즉 '얼마나 유덕한 시민에 가까운가?'는 원리적으로 양적으로 측정 가능하며, 이것이 GWO 척도 개발에 있어서 심리학적 타당성의 근거가 된다(Aquino & Reed, 2002; Blasi, 1984; Lapsley & Narvaez, 2004).

결론 및 논의

본 연구는 Rousseau의 일반의지를 18세기 정치철학의 담론으로부터 소환하여 21세기 심리학적 구성개념으로 재정립하고자 하는 시도였다. 이를 위해 Rousseau의 자연상태 이론, 자기사랑과 이기심의 분기, 탈자연화 개념, 사회계약론의 핵심인 공동자아 개념을 심리학적 언어로 번역하는 작업을 수행하였다. 그리고 이러한 개념적 전이를 통해 심리학적 구성개념으로서 일반의지 지향성(General Will Orientation: GWO)을 제안하고, 이것이 Kant의 정언명령, Rawls의 공정으로서의 정의, Kohlberg의 도덕발달 이론, 긍정심리학의 VIA 체계와 어떻게 구별되는 독자적 구성개념인지를 논증하였다. 아울러 인간 도덕성의 부정적 극단을 반영하는 정신병질과의 대칭적 구조를 통해 GWO가 측정 가능한 심리학적 실체임을 논증하였다. 마지막으로 규범윤리와 덕윤리의 이분법을 넘어 이 둘을 통합할 수 있는 대안으로서 일반의지의 이론적 위상을 제시하였다. 본 절에서는 이상의 논의를 종합하고, 데이터리즘이라는 21세기적 위기의 맥락에서 GWO가 갖는 이론적 함의와 한계를 검토한다.

일반의지의 심리학적 소환이 갖는 이론적 함의

본 연구에서 가장 핵심적인 작업은 일반의지의 심리학적 전이였다. 정치철학의 맥락에서 일반의지는 주권자로서의 시민이 사적 이해관계를 초월하여 공동체 전체에 유익한 방향을 자발적으로 입법하는 의지를 가리킨다. 이를 심리학적 구성개념으로 전이하는 과정에서 본 연구는 두 가지 핵심 문제를 해결해야 했다. 첫째는 형이상학적 당위의 문제이다. 일반의지는 Rousseau의 체계 내에서 항상 옳다는 무오류성을 전제하는 바, 이것이 경험과학의 언어로 번역될 때 형이상학적 과잉이 되지 않는가 하는 질문이다. 본 연구는 이 무오류성을 경험적 주장이 아닌 구성적 주장으로 재해석함으로써 이 문제를 해결하였다. 일반의지란 정의상 공공선을 지향하는 의지이며, 이 구성적 정의에 의해 일반의지는 항상 공공선의 방향을 향한다. 이 구성적 무오류성은 일반의지를 형이상학적 당위의 영역으로부터 측정 가능한 심리학적 변인의 영역으로 이동시키는 데 결정적인 역할을 한다.

둘째는 측정 가능성의 문제이다. 어떤 개념이 심리학적 구성개념으로 인정받기 위해서는 조작적 정의를 통해 측정 가능한 변인으로 실체화되어야 한다. 본 연구는 이를 위해 정신병질의 심리학적 실체화 과정을 선례로 삼았다. Cleckley(1941)가 임상 관찰을 통해 개념을 기술하고, Hare(1991)의 PCL-R이 이를 측정 가능한 구성개념으로 완성한 것처럼, GWO는 Rousseau의 원전이라는 철학적 기초 위에서 측정 도구 개발을 통해 심리학적 실체성을 획득하게 될 것이다. 특히 정신병질이 반영하는 사회적 유대의 해체라는 부정적 극단에 대응하여 사회적 유대의 재구성과 공공선 지향을 반영하는 긍정적 극단으로서 GWO를 위치시킴으로써, 인간 도덕성의 지형도를 보다 균형있게 완성하고자 하였다. 이처럼 일반의지의 심리학적 소환은 단순한 개념 차용이 아니라, 도덕심리학의 외연을 철학적 깊이로 확장하는 이론적 시도라 할 수 있다.

데이터리즘 비판과 GWO의 시의성

본 연구가 일반의지를 21세기적 맥락에서

소환하는 이유는 데이터리즘이 초래한 철학적, 심리학적 위기에 있다. Harari(2017/2023)가 명명한 데이터리즘은 인간의 경험과 의사결정을 정보의 흐름으로 환원하고, 알고리즘을 통해 최적화된 선택지를 인간에게 제시하는 세계관이다. 알고리즘이 건네주는 선택지 안에서 인간의 도식이 조형되는 환경에서, 자유의지는 알고리즘적 반응으로 치환되고 주체적 숙의의 공간은 급격히 협소해진다. 이러한 상황은 Hobbes, Locke, Rousseau가 가정한 자연상태를 디지털 공간에서 재연하는 이른바 디지털 신자연상태라 할 수 있다.

이 신자연상태에서 가장 두드러지는 심리학적 현상은 두 가지이다. 하나는 의지의 외주화이다. 개인이 자신의 도덕적 판단과 선택의 권한을 알고리즘 시스템에 위임하는 과정에서 주체적 입법 능력은 점차 약화된다(Ryan & Deci, 2000). 다른 하나는 파편화된 개인주의이다. 초연결 사회의 역설적 이면에서 공동체적 선(善)에 대한 감각이 마비되고, 개인은 데이터 생태계 내의 수동적 개체로 전락한다(Harari, 2017/2023). 한국 사회에서도 이러한 파편화는 전통적 공동체 문화의 급속한 해체와 맞물리며 심화되고 있다(박준성, 2024). GWO가 측정하고자 하는 공공선 지향성, 자기입법적 주체성, 정서적 연대성은 이 두 현상에 대한 심리학적 대항 기제이다(Bandura, 2001). 특히 GWO가 단순한 친사회적 태도나 기존 규범에 대한 순응과 구별되는 지점은, 기존 맥락 자체를 비판적으로 성찰하고 보편적 가치에 부합하는 방향으로 재구성하려는 능동적 도덕 지향을 포함한다는 데 있다. 이는 알고리즘이 제공하는 최적화된 선택지에 순응하는 착한 부품으로의 전락을 거부하고, 스스로 입법자가 되어 결단하는 주체적 시민성의 심리학적 기반이 된다.

유사 개념들과의 비교를 통한 GWO의 독자성

본 연구의 이론적 논의는 GWO가 철학적 유사 개념들 및 심리학적 유사 변인들과 어떻게 구별되는지를 체계적으로 밝히는 데 상당한 지면을 할애하였다. 이 비교 논의를 통해 드러난 GWO의 독자성은 크게 세 가지 층위로 요약된다.

첫째, 기제의 차원에서 GWO는 Kant의 정언명령과 구별된다. Kant의 도덕 모델이 자연적 경향성을 억제하는 자기극복의 윤리라면, Rousseau의 일반의지는 자기사랑을 이성과 양심을 통해 사회적 수준으로 확장하는 통합의 윤리이다. 해야 한다는 의무가 하고 싶다는 경향성을 극복할 것을 요구하는 Kant의 모델은 행위자에게 지속적인 내적 긴장을 부과하는 데 반해, 일반의지에 기반한 도덕적 행동은 확장된 자기사랑이 공동체의 안녕과 일치하는 상태에서 자발적으로 흘러나온다. 이 차이는 데이터리즘 시대에 더욱 선명한 함의를 가진다. Kant의 정언명령은 입력된 준칙이 보편화 가능한지를 검증하는 알고리즘적 절차로 환원될 수 있지만, Rousseau의 일반의지는 타자의 고통에 정서적으로 공명하고 공동체의 운명에 주체적으로 헌신하는 뜨거운 마음의 영역에 속하며, 이는 어떤 고도화된 인공지능도 대체할 수 없는 인간 고유의 실존적 역량이다.

둘째, 동기의 차원에서 GWO는 Rawls의 공정으로서의 정의와 구별된다. Rawls의 원초적 입장에 놓인 합의 당사자들은 상호 무관심한 합리성에 기초하여 이타적 원칙에 합의하지만, 그 동기는 타인에 대한 연대가 아니라 불확실

한 미래에 대한 불안과 회피에 있다. Rawls의 시민은 서로 사랑하거나 연대하지 않아도 제도가 보장하는 안전 안에서 정의롭게 공존할 수 있다. 반면 GWO가 반영하는 일반의지의 동기는 계산된 합리성이 아니라 타인의 아픔을 나의 아픔으로 느끼는 연민과 공동체를 향한 정서적 유대이다. 이 차이는 단순히 철학적 입장의 차이를 넘어, 알고리즘이 합리적 선택을 대신해주는 시대에 제도적 안전망만으로는 주체적 시민성이 담보되지 않는다는 심리학적 통찰로 이어진다.

셋째, 범위의 차원에서 GWO는 긍정심리학의 VIA 체계와 구별된다. VIA 체계는 덕이론을 심리학적 변인으로 전환하는 과정에서 덕이 존재하는 조건인 폴리스적 공동체성을 제거하고, 덕의 궁극적 지향점을 공공선의 실현에서 개인의 주관적 안녕감과 현실 적응으로 치환하였다. 그 결과 VIA 체계의 자기조절, 친절, 사랑, 시민정신, 공정성은 미시도덕적 수준에 머물게 되었다. 그러나 GWO는 VIA 체계가 담아내지 못한 거시적 차원을 미시적 동기 에너지와 함께 통합하는 구성개념이다. VIA 체계의 미시적 덕성이 알고리즘 시스템에 순응하는 개인의 적응을 강화하는 데 그칠 수 있다면, GWO는 그 시스템 자체를 비판적으로 성찰하고 공동체적 입법자로서 재구성하려는 능동적 도덕 지향을 포함한다.

GWO와 정신병질: 도덕성 지형도의 완성

정신병질과 GWO의 대칭적 구조는 본 연구에서 제안하는 이론적 기여의 핵심 중 하나이다. PCL-R(Hare, 1991)의 등장 이후 정신병질은 사회적 유대의 해체와 자기중심성을 반영하는 심리학적 지표로서, 비윤리적 행동 및 반사회적 행동과 관련한 다양한 개인차 연구의 핵심 변인으로 기능해 왔다. 그러나 도덕적 행동에 대한 심리학적 이해는 얼마나 악(惡)할 수 있는가라는 물음만으로는 완성되지 않는다. 인간이 얼마나 선(善)을 지향하며 행동할 수 있는가를 하나의 통합된 심리학적 변인으로 포착하는 구성개념이 함께 요청된다. GWO는 이 요청에 응답하기 위한 심리학적 구성개념으로 제안된다.

정신병질의 Factor 1이 반영하는 공감 결여와 타인의 도구적 취급에 대응하여 GWO의 공공선 지향성이 위치하며, Factor 2가 반영하는 정서조절 실패에 따른 충동적 탈억제에 대응하여 GWO의 자기입법적 주체성이 위치하고, 두 요인을 관통하는 사회적 유대의 해체에 대응하여 GWO의 정서적 연대성이 위치한다. 이 대응 관계는 수사적 대칭이 아니라, 동일한 행동 영역 위에서 반대 방향으로 작동하는 측정학적 대칭이며, 인간 도덕성의 지형도를 보다 완전하게 그려내기 위한 이론적 요청이다. 정신병질이 악의 심리학적 실체를 규명해 왔다면, GWO는 선의 심리학적 실체를 규명하기 위한 이론적 토대를 제공한다. 그리고 정신병질과 GWO가 인간 도덕성의 양쪽 극단을 함께 다뤄나갈 때, 비로소 도덕적 행동에 대한 보다 균형 있고 완전한 심리학적 이해가 가능해진다. GWO 점수가 높을수록 개별의지와 일반의지의 일치 정도가 높은, Rousseau의 유덕한 시민에 가까운 심리 상태를 반영하며, 이 연속적 스펙트럼 위의 분포가 인간 도덕성의 보다 입체적인 지형도를 가능하게 한다.

본 연구의 한계와 후속 과제

본 연구는 이론적 논의에 집중한 문헌 연구

로서, 몇 가지 한계를 내포하고 있다. 우선, GWO의 이론적 타당성은 논증되었으나, 이를 실증적으로 검증하는 측정 도구의 개발과 경험적 연구가 후속 과제로 남아 있다. Rousseau의 원전이라는 철학적 기초 위에서 GWO의 개념적 토대는 구축되었지만, PCL-R이 정신병질을 완성한 것처럼 GWO 척도의 개발과 타당화 작업이 수행될 때 비로소 심리학적 구성개념으로서의 실체성이 온전히 확보된다. 이는 본 연구의 후속 과제로서 전문가 델파이 조사, 탐색적 및 확인적 요인분석, 정신병질 척도(LSRP: Levenson et al., 1995) 및 긍정심리학 척도들과의 변별타당도 및 수렴타당도 검증을 통해 완성될 것이다(DeVellis, 2017).

또한 Rousseau의 일반의지를 심리학적 구성개념으로 전이하는 과정에서 철학적 원개념의 일부 층위가 불가피하게 단순화되었을 가능성이 있다. Rousseau의 체계 내에서 일반의지는 고도로 정합적인 정치철학적 맥락 안에서 의미를 가지는 바, 이를 측정 가능한 심리학적 변인으로 조작화하는 작업은 원개념의 풍부함을 일부 희생하는 것을 전제한다. 이 점에서 본 연구는 정치철학적 엄밀성과 심리학적 조작화 가능성 사이의 긴장을 완전히 해소하지 못했다는 한계를 인정한다.

나아가 본 연구에서 제시한 GWO의 세 하위요인, 즉 공공선 지향성, 자기입법적 주체성, 정서적 연대성이 단일 구성개념의 하위요인으로 수렴되는지, 아니면 상호 독립적인 구성개념들의 집합인지는 향후 요인분석을 통해 경험적으로 검증되어야 한다. 또한 문화적 맥락의 문제도 남아 있다. Rousseau의 일반의지가 서구 근대 정치철학의 산물인 만큼, 집단주의 문화권에서의 GWO 발현 양상이 서구 개인주의 문화권과 어떻게 다른지에 대한 비교문화적 검토가 요청된다(Triandis, 1995).

끝으로, 본 연구에서 GWO의 심리학적 동력으로 예비적으로 제시한 성숙한 자기애 개념과의 관계는 이론적 수준의 논의에 그쳤으며, 성숙한 자기애가 GWO를 매개하거나 고양하는 인과적 경로에 대한 실증적 규명이 후속 연구의 과제로 남는다. Kohut(2009)의 자기심리학에서 도출된 성숙한 자기애 개념과 Rousseau의 자기사랑 개념 간의 유비는 이론적으로 설득력이 있지만, 이 유비가 경험적 검증을 통해 실제로 확인될 때 GWO 연구의 이론적 완성도는 한층 높아질 것이다.

결론: 데이터리즘 시대의 유덕한 시민을 향하여

Rousseau는 법은 대리석이나 청동이 아니라 시민들의 가슴 속에 새겨져야 한다고 역설하였다(Rousseau, 1997). 이 선언은 알고리즘이 인간의 선택을 선점하고 데이터가 의지의 방향을 결정하는 오늘날, 그 어느 때보다 강한 울림을 가진다. 아무리 정교한 제도와 공정한 절차가 설계되더라도, 그것을 작동시키는 시민들의 마음에 공동체를 향한 덕이 없다면 제도는 공허해진다. 본 연구가 GWO를 통해 규명하고자 한 것은 바로 이 마음의 습관, 즉 데이터리즘의 유혹 속에서도 스스로 입법하고 공동체를 향해 주체적으로 결단하는 심리적 역량이다.

18세기 Rousseau의 통찰을 21세기 심리학으로 소환하는 이 작업은, 파편화된 개인들이 알고리즘의 최적화된 선택지에 순응하는 착한 부품으로 전락하지 않고 스스로를 공동체의 주권자로 재발견하는 과정을 심리학적으로 뒷받침하기 위한 시도이다. GWO는 이 과정의 측정 가능한 심리학적 지표로서, 인간이 얼마

나 유덕한 시민에 가까운가를 포착하는 도구가 될 것이다. 일반의지는 뜨거운 마음의 영역에 속한다. 그것은 어떤 알고리즘도, 어떤 고도화된 인공지능도 대체할 수 없는 인간 고유의 실존적 역량이다. 바로 그렇기 때문에, 데이터리즘의 시대야말로 일반의지의 심리학적 소환이 가장 절실한 시대이다.

참고문헌

김시형 (2015). 루소 비판을 통한 칸트의 새로운 철학이념의 확립: 학문과 지혜의 통일. **철학연구**, 135, 179-207.

김용민 (2016). 루소와 공화주의. **한국정치연구**, 25(1), 167-192.

김은주 (2023). 루소의 자연주의 교육론 고찰: 『에밀』을 중심으로. **학습자중심교과교육연구**, 23(11), 1-13.

박준성 (2024). 한국 사회에서 공존을 위한 문화심리학적 이해: 수직적 관계주의와 수평적 관계주의를 중심으로. **한국심리학회지: 문화 및 사회문제**, 30(4), 635-649.

박찬영 (2022). 루소의 감정론에 대한 고찰: amour de soi와 amour-propre를 중심으로. **교육사상연구**, 36(2), 67-89.

박호성 (1993). 루소의 자연개념: '비판적' 자연과 '창조적' 자연. **한국정치학회보**, 27(2), 37-58.

오근창 (2013). 일반의지의 두 조건은 상충하는가?: 루소와 '자유롭도록 강제됨'의 역설. **철학사상**, 47, 67-98.

오수웅 (2017). 루소의 일반의지와 공동심의: 의회 심의의 기준과 원리. **한국정당학회보**, 16(1). 137-164.

윤 황 (2022). Piaget의 자기중심성은 전조작기적 특성인가?: 세 산 모형 실험에서 심리화까지. **차세대융합기술학회논문지**, 6(10), 1971-1978.

윤 황 (2024a). 콜버그 이론은 피아제 이론을 대체했는가?: 피아제와 콜버그의 도덕발달론 비교 고찰. **인간연구**, 53, 263-293.

윤 황 (2024b). 신콜버그 학파의 피아제 이론으로의 회귀현상. **차세대융합기술학회논문지**, 8(10), 2236-2248.

임의영 (2020). 공공성의 정치철학적 기초: J. Rousseau의 문명관과 일반의지를 중심으로. **정부학연구**, 26(1), 37-73.

장경원, 전성주, 김근영 (2022). 문화성향과 SNS 사용 간의 관계: 오프라인 문화성향의 효과를 통제한 위계적 회귀분석. **한국심리학회지: 문화 및 사회문제**, 28(3), 393-417.

Annas, J. (2011). *Intelligent virtue.* Oxford University Press.

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423-1440.

Aristotle. (2013). **니코마코스 윤리학** (천병희 역). 숲. (원저 출판연도 미상).

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1-26.

Batson, C. D. (2011). *Altruism in humans.* Oxford University Press.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control.

Current directions in psychological science, 16(6), 351-355.

Bentham, J. (1996). *An introduction to the principles of morals and legislation.* Oxford University Press. (원저 1789년 출판).

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin, 88*(1), 1-45.

Blasi, A. (1984). Moral identity: Its role in moral functioning. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp. 128-139). Wiley.

Cleckley, H. (1941). *The mask of sanity: An attempt to clarify some issues about the so-called psychopathic personality.* Mosby.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.

Edwards, C. P. (1986). Cross-cultural research on Kohlberg's stages: The basis for consensus. *Lawrence Kohlberg: Consensus and controversy*, 419-430.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51,* 665-697.

Fowers, B. J. (2005). *Virtue and psychology: Pursuing excellence in ordinary practices.* American Psychological Association.

Gabbard, G. O. (2014). *Psychodynamic Psychiatry in Clinical Practice (5th ed.).* American Psychiatric Publishing.

Gilligan, C. (1997). 다른 목소리로: 심리 이론과 여성의 발달 (허라금 역). 동녘. (원저 1982 출판).

Harari, Y. N. (2023). 호모데우스: 미래의 역사 (김명주 역). 김영사. (원저 2017년 출판).

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised.* Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised (2nd ed.).* Multi-Health Systems.

Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology, 56*(5), 741-747.

Hobbes, T. (2018). 리바이어던 (진석용 역). 나남. (원저 1651년 출판)

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice.* Cambridge University Press.

Hursthouse, R. (1999). *On virtue ethics.* Oxford University Press.

Kant, I. (2009). 실천이성비판 (백종현 역). 아카넷. (원저 1788 출판).

Kant, I. (2018). 윤리형이상학 정초 (백종현 역). 아카넷. (원저 1785 출판).

Kierstead, F. D. (1974). *Education for a Transitional Democracy: A Comparison of Jean Jacques Rousseau's Concept of General Will to John Dewey's Concept of Collective Intelligence.* Doctoral Dissertation of The University of Oklahoma.

Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice* (Essays on moral development, Vol. 1). Harper & Row.

Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages* (Essays on moral development, Vol. 2). Harper & Row.

Kohut, H. (2009). *The analysis of the self: A systematic approach to the psychoanalytic treatment*

of narcissistic personality disorders. The University of Chicago Press.

Kristjánsson, K. (2013). *Virtues and vices in positive psychology: A philosophical critique.* Cambridge University Press.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2*, 111-133.

Lapsley, D. K., & Narvaez, D. (2004). A social-cognitive approach to the moral personality. In D. K. Lapsley & D. Narvaez (Eds.), *Moral development, self, and identity* (pp. 189-212). Lawrence Erlbaum.

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*(1), 151-158.

Locke, J. (2022). 통치론 (강정인, 문지영 역). 까치. (원저 1689 출판).

MacIntyre, A. (1981). *After virtue: A study in moral theory.* University of Notre Dame Press.

Mill, J. S. (2002). 공리주의 (이을상 역). 지식을만드는지식. (원저 1863년 출판).

Narvaez, D. (2005). The neo-Kohlbergian tradition and beyond: Schemas, expertise, and character. In C. Pope-Edwards & G. Carlo (Eds.), *Nebraska Symposium on Motivation: Vol. 51. Moral motivation through the lifespan* (pp. 119-163). University of Nebraska Press.

Niemiec, R. M. (2018). *Character strengths interventions: A field guide for practitioners.* Hogrefe.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you.* Penguin Press.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, 56,* 365-392.

Peterson, C., & Seligman, *M. E. P. (2004). Character strengths and virtues: A handbook and classification.* Oxford University Press.

Piaget, J. (1981). *Intelligence and affectivity: Their relationship during child development.* (T. A. Brown & C. E. Kaegi, Trans.). Annual Reviews, Inc.

Rawls, J. (1995). Political liberalism: Reply to Habermas. *The Journal of Philosophy, 92*(3), 132-180.

Rawls, J. (2003). 정의론 (황경식 역). 이학사. (원저 1999 출판).

Rest, J. R. (1979). *Development in judging moral issues.* University of Minnesota Press.

Rest, J., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach.* Lawrence Erlbaum Associates.

Rousseau, J. J. (1997). Considerations on the government of Poland. In V. Gourevitch (Ed. & Trans.), *The social contract and other later political writings* (pp. 177-260). Cambridge University Press. (Original work published 1782)

Rousseau, J. J. (2003a). 인간불평등기원론 (주경복, 고봉만 역). 책세상. (원저 1755 출판)

Rousseau, J. J. (2003b). *에밀* (김중현 역). 한길사. (원저 1762 출판).

Rousseau, J. J. (2011). 사회계약론 (김중현 역). 웅진씽크빅. (원저 1762 출판).

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Searle, J. R. (1995). *The construction of social reality.* Free Press.

Snarey, J. R. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin, 97*(2), 202-232.

Sreenivasan, G. (2000). What is the general will?. *The Philosophical Review, 109*(4), 545-581.

Stocker, M. (1976). The schizophrenia of modern ethical theories. *Journal of Philosophy, 73*(14), 453-466.

Triandis, H. C. (1995). *Individualism and collectivism.* Westview Press.

Yun, H. (2018). Instrumental and reactive transgressions of primary and secondary psychopathic traits: Focus on moral emotion, moral disengagement, and unethical decision making. *Korean Journal of Clinical Psychology, 37*(3), 323-338.

논문 투고일 : 2026. 03. 12
1 차 심사일 : 2026. 05. 03
게재 확정일 : 2026. 05. 20

한국심리학회지 : 문화 및 사회문제
Korean Journal of Culture and Social Issues
2026, Vol. 32, No. 2, 295～327.

General Will Orientation (GWO) as a Psychological Construct: Evoking General Will in the 21st-Century

Hwang Yun

Department of Psychology & Counseling, Pai Chai University

This study theoretically reconceptualizes Rousseau's General Will—a cornerstone of 18th-century political philosophy—as a psychological construct for the 21st century. Amid an existential crisis in which Dataism supplants individual autonomy and dissolves communal bonds, this study translates Rousseau's theory of the state of nature, the bifurcation of amour de soi and amour-propre, the concept of denaturalization, and the common self (moi commun) into psychological language. Through this transposition, General Will Orientation (GWO) is proposed as a novel psychological construct comprising three subfactors: public good orientation, self-legislative agency, and emotional solidarity. This study argues that GWO is distinct from Kant's categorical imperative, Rawls's justice as fairness, Kohlberg's Stage 5 of moral development, and the VIA Classification of Character Strengths. Each framework illuminates morality from a different vantage point—autonomous legislation, procedural fairness, cognitive moral reasoning, and individual virtue—yet shares a common limitation in failing to integrate affective-motivational mechanisms with communal context. The psychological reality and measurability of GWO are demonstrated through its structural symmetry with psychopathy, representing the negative end of human morality. GWO is further positioned as an integrative alternative transcending the dichotomy between normative and virtue ethics. The infallibility of the General Will is reinterpreted not as an empirical claim but as a constitutive one, providing theoretical basis for translating a metaphysical imperative into a measurable psychological variable. GWO is anticipated to expand the scope of moral psychology with philosophical depth, serving as a psychological counterforce that safeguards human moral agency in the age of algorithmic governance.

Key words : Rousseau, General Will, GWO, Dataism, psychopathy