creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

교육학박사 학위논문

도덕 판단에서 동기화된 추론의 문제 및 개선방안 연구

2023년 8월

서울대학교 대학원

윤리교육과

강 민 겸

도덕 판단에서 동기화된 추론의 문제 및 개선방안 연구

지도교수 정 창 우

이 논문을 교육학박사 학위논문으로 제출함
2023년 6월

서울대학교 대학원
윤리교육과
강 민 겸

강민겸의 박사 학위논문을 인준함
2023년 8월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

국문초록

본 논문은 도덕 판단에서 편향성이 나타나는 원인을 동기화된 추론 개념에 근거하여 분석하고, 이를 도덕교육적으로 개선하기 위한 방안을 모색하는 것을 목적으로 한다. 동기화된 추론이란 추론의 과정이 동기적 요인에 영향을 받는 현상을 지칭한다. 일반적으로 추론에 영향을 주는 동기는 자신의 입장을 정당화하려는 목표를 갖기 때문에 편향된 방식으로 작동하기 쉽다. 특히, 도덕 판단에서 동기화된 추론은 검토되지 않은 신념과 정체성을 보호하기 위해 증거를 왜곡하여 처리함으로써 도덕적 추론의 객관성과 공정성을 훼손하기도 한다.

콜버그의 연구로 대표되는 기존 도덕 심리학의 주류 논의는 도덕적 추론이 도덕 판단을 인과적으로 도출해내는 심리 기제라고 전제한다. 그러나 하이트의 연구는 도덕 판단에서 추론의 인과적 역할에 대해 의문을 제기한다. 하이트의 사회적 직관주의 모델에 따르면, 대부분의 도덕 판단은 도덕적 직관에 의해 즉각적으로 나타나며 도덕적 추론은 이미 내려진 결론에 부합하도록 사후적인 정당화 기능을 수행하기 때문이다. 하이트가 주목한 도덕적 추론은 일종의 동기화된 추론으로 해석할 수 있다.

도덕 판단에서 동기화된 추론은 비도덕적 행위를 정당화하며 사회 구성원들 사이의 도덕적·정치적 양극화를 심화시키는 문제를 유발한다. 먼저, 동기화된 추론은 도덕적이지 못한 행동을 하거나 판단을 내린 이후에도 여전히 스스로를 도덕적 주체로 생각하게 하는 그럴듯한 근거를 제공한다. 또한 우리는 동기화된 추론을 통해 독단적인 신념을 공고히 하여, 자신과 의견이 다른 타인에 대한 배타적인 성향을 강화하고 집단적 분열을 증폭시킨다.

도덕 판단에서 동기화된 추론으로 인한 문제는 신념을 보호하려는 마음의 경향과 편향의 사각지대 현상이라는 두 가지 주요한 요인에 의해 발생한다. 사람들은 검토되지 않은 자신의 신념과 상충하는 정보나 논증을 도덕적, 사회적 정체성에 대한 위협으로 받아들이고, 이를 본능적으로 거부하려는 모습을 보인다. 이 과정에서 자신의 신념과 정체성을 방어하기 위한 방향으로 추론은 동기화된다. 동기화된 추론은 우리의 비도덕적인 행위와 판단까지도 정당화하는 이유를 구성하고, 자신과 신념이 다른 타인을 배척하는 근거를 만들어내기도 한다. 나아가 우리는 종종 자신의 편향된 사고에 대해 객관적으로 인지하지 못한다. 편향의 사각지대에 빠진 사람은 자신의 도덕 판단과 이를 뒷받침하는 근거가 객관적이고 공정하다고 생각하는 한편, 자신의 의견과 상충하는 타인의 판단과 근거는 편향된 것으로 인식한다.

신념의 변화 가능성과 편향의 사각지대에서 벗어날 수 있다는 사실은 동기화된 추론 또한 교육을 통해 개선될 수 있음을 시사한다. 특히, 도덕적 메타인지의 함양은 검토되지 않은 신념에 대해 성찰의 기회를 제공하고 편향의 사각지대를 극복하는 계기를 마련해 준다. 나아가 추론의 기원과 본질적 기능에 대한 최근 인지과학 및 심리학의 학제적 접근은 동기화된 추론의 개선을 위해 추론이 갖는 사회적 기능을 고려해야 함을 시사한다. 이는 사람들이 윤리적 문제에 대한 공유된 이해를 구축하기 위해 서로 다른 관점에서 논증을 주고받는 협력적 논증 교환으로 구체화할 수 있다.

협력적 논증 교환은 특정 이슈에 대한 자신의 도덕 판단을 논증의 형태로 표현하고, 논증에 대한 평가를 주고받는 순환적인 과정으로 구성된다. 이러한 상호작용을 통해 논증 참가자들은 개인의 초기 신념을 객관적으로 파악하고 검토하여 수정할 수 있는 기회를 갖는다.

협력적 논증 교환 활동은 학교 도덕교육에서 수업의 형태로 활용될 수 있다. 교사는 학생들이 공개적으로 자신의 도덕 판단을 정당화하고, 동료 학생의 논증을 진정성 있게 평가하는 절차를 포함하여 수업을 설계해야 한다. 또한 학생들이 신뢰할 수 있는 소집단 내에서 자신과 의견이 다른 사람과 협력적 논증 교환 활동을 수행할 수 있도록 수업을 운영해야 한다.

협력적 논증 교환 수업을 효과적으로 적용하기 위해 교사는 기존에 개발된 도덕과 수업 모형을 활용할 수 있다. 반대 논증을 평가해 보는 과정이 명시적으로 포함된 콘스탄츠 딜레마 토론(KMDD)은 협력적 논증 교환 수업이 이루어지기 위한 필수적인 조건과 교사의 유의사항에 시사점을 제공한다. 논쟁적 주제를 수업에서 다루기 위해 개발된 건설적 논쟁 수업 모형에서는 협력적 논증 교환 수업의 절차와 세부 운영 지침을 참고할 수 있다. 하이트의 도덕심리학 이론을 직접적으로 적용해 보고자 한 하이트식 도덕적 통찰 모형은 시론적인 단계이지만, 협력적 논증 교환 수업과 상호보완적인 선순환의 관계를 통해 동기화된 추론의 개선에 기여할 것으로 기대할 수 있다.

주요어 : 도덕 판단, 동기화된 추론, 도덕적 추론, 인지 편향, 도덕적 메타인지, 협력적 논증 교환
학 번 : 2020-39021

목 차

제 1 장 서 론 ····· 1

제 1 절 연구의 배경 및 목적 ····· 1

제 2 절 연구의 방법 및 구성 ····· 9

제 2 장 동기화된 추론의 이해 ····· 12

제 1 절 동기화된 추론의 이론적 토대 ····· 12

1. 인식론적 기초 ····· 12

2. 경험과학적 근거 ····· 19

제 2 절 동기화된 추론의 의미와 유형 ····· 28

1. 동기화된 추론의 의미 ····· 28

2. 동기화된 추론의 유형 ····· 36

제 3 장 도덕 판단에서 동기화된 추론의 문제 49

제 1 절 동기화된 추론과 도덕 판단의 관계 ····· 49

1. 동기화된 추론에 관한 도덕 심리학적 접근 ····· 49

2. 도덕 판단에서 동기화된 추론의 특징 ····· 64

제 2 절 도덕 판단에서 동기화된 추론으로 인한 문제 ····· 76

1. 비도덕적 행위의 정당화 ····· 76

2. 도덕적, 정치적 불일치의 양극화 심화 ····· 91

제 3 절 도덕 판단에서 동기화된 추론 문제의 원인 ····· 100

1. 도덕적 신념의 보호 ····· 100

2. 편향의 사각지대 현상 ····· 111

제 4 장 도덕 판단에서 동기화된 추론의 개선방안 122
제 1 절 동기화된 추론의 개선 가능성 ····· 122
1. 도덕적 신념의 가변성 ····· 122
2. 편향의 사각지대 극복 ····· 130
제 2 절 동기화된 추론의 개선을 위한 방향 ····· 137
1. 도덕적 메타인지의 활용 ····· 137
2. 추론의 사회적 기능의 고려 ····· 143
제 3 절 동기화된 추론의 개선방안으로써
협력적 논증 교환 ····· 152
1. 협력적 논증 교환의 의미 ····· 152
2. 협력적 논증 교환의 교육적 타당성 ····· 159

제 5 장 협력적 논증 교환의 도덕교육적 활용 ·· 167
제 1 절 협력적 논증 교환 수업의 특징 ····· 167
1. 동기화된 추론 개선을 위한 수업 목표 설정 ····· 167
2. 인지적 작업의 대립적 분업으로서 수업 양상 ····· 180
3. 교사의 역할 및 실행 전략 ····· 188
제 2 절 협력적 논증 교환 수업의 사례 검토 ····· 203
1. 협력적 논증 교환 수업의 사례1: KMDD 수업 모형 ····· 203
2. 협력적 논증 교환 수업의 사례2: 건설적 논쟁 수업 모형 ····· 215
3. 협력적 논증 교환 수업의 사례3: 하이트식 도덕적 통찰 모형 226

제 6 장 결 론 ····· 238

참 고 문 헌 ····· 244
Abstract ····· 269

표 목 차

[표 2-1] 목표에 따른 동기화된 추론의 두 유형 ····· 37
[표 3-1] 도덕 발달 단계에 따른 도덕적 행위의 동기 ···· 54
[표 4-1] 대화와 갈등의 범위를 기준으로 구분되는 추론의 양상 155
[표 5-1] KMDD의 일반적 절차 ····· 208
[표 5-2] 논쟁 수업 모형의 구조 ····· 224
[표 5-3] 하이트식 도덕적 통찰 모형의 구조 ····· 229

그 림 목 차

[그림 1-1] 연구 구조 ····· 10
[그림 2-1] 동기화된 추론의 유형 ····· 44
[그림 3-1] 도덕 판단의 사회적 직관주의 모형 ····· 61
[그림 3-2] 도덕적 자기제재가 선택적으로 활성화되는 메커니즘 82
[그림 4-1] 도덕적 가치가 신념에 영향을 미치는 두 가지 방식 125

제 1 장 서 론

제 1 절 연구의 배경 및 목적

인간의 도덕성과 관련된 심리적 현상을 탐구하는 도덕 심리학 분야는 일반적으로 도덕적 추론을 도덕 판단에서 인과적인 역할을 수행하는 심리 기제로 이해해왔다(Bucciarelli et al., 2008; Kohlberg, 1981; Turiel, 1983). 도덕적 추론은 감정의 개입과는 독립적으로, 보편적 도덕 원칙에 입각하여 연역적인 방식으로 도덕 판단을 도출하는 과정이기 때문이다. 특히, 도덕 심리학에서 발달 심리학적 접근은 도덕적 추론을 인간의 고차원적 사고 능력을 바탕으로 정교한 언어를 사용하는 의식적인 과정으로 파악한다(Kohlberg, 1981; Turiel, 1983). 이때, 도덕 발달은 복잡한 윤리적 문제에 대해 자기중심적인 동기를 극복하고, 보다 공정한 정의추론(justice reasoning)의 방식을 취하는 것으로 이해된다(Kohlberg, 1981). 도덕적 추론을 사용하는 이성적 존재로서 인간의 모습은 인간과 인간 이외의 동물 종을 구분시켜주는 주요한 특징이기도 하다.

그러나 현실의 인간에게 이성과 추론의 사용은 도달하기 어려운 이상(理想)에 가까워 보인다. 일상에서 우리는 그리 합리적이지 않으며, 편향에 매우 취약하고 논리적 오류를 광범위하게 범하기 때문이다. 추론에 대한 심리학 연구들은 실험 참가자를 인지적 착각에 빠뜨리는 일상적 장치를 통해, 그들이 논리적 오류를 쉽게 범하는 모습을 포착하며(Kahneman et al., 1982/2001; Wason, 1969; Wason & Evans, 1975), 이를 인간의 이성과 추론 능력에 심각한 결함이 있다는 주장의 근거로 사용하기도 한다(Haselton et al., 2005; Nickerson, 1998; Evans, 2002). 인간 인지에서 나타나는 체계적인 오류 현상을 지칭하는 인지 편향은 심리학 연구에서 등장한 이래로(Kahneman et al., 1982/2001), 현재까지 약 180개 이상의 종류로 세분화되었다(하민수, 2016). 현대 사회를 이성의

위기(edge of reason) 시대로 규정한 철학자 바지니(J. Baggini)는 인간이 이성적으로 사고할 수 있다는 능력에 대한 과거의 지나친 과신을 경계하고 이성의 한계를 지적하는 여러 사실을 겸허히 인정하는 자세가 필요함을 강조한다(Baggini, 2016/2017). 이러한 사실은 인간의 도덕적 삶에도 적용될 수 있다. 도덕 행위자로서 인간은 이성적이고 합리적으로 행동하기보다는, 인지 편향에 취약하고 이로 인해 실제 삶에서 우리는 추론을 제한적으로 사용하는 것처럼 보이기 때문이다.

최근 도덕 심리학 연구는 도덕적 추론이 도덕 판단을 인과적으로 도출해내는 기능을 수행한다는 전제에 회의적인 입장을 취한다. 특히, '새로운 종합(new synthesis)[1]'의 관점에서 인간의 사회적 행동을 설명하고자 하는 심리학적 접근은 인간의 도덕 판단을 포함한 도덕적 삶에 대한 새로운 시각을 제공해준다. 이러한 관점에서 도덕 심리학 연구 분야는 다양한 실증적 연구 결과들을 바탕으로 사회 심리학 및 인지 과학 등의 접근을 통해 도덕적 추론에 대해 기술적(descriptive) 측면에서 더욱 설명력 있는 이론을 개발하려고 시도하고 있다. 대표적으로 하이트(J. Haidt)는 도덕 발달에서 정의추론의 역할에 주목한 기존의 콜버그식 도덕 판단 연구를 심리적 합리주의로 규정하고 이 모델이 일상의 도덕 판단을 설명하는 데 제한적이라고 비판한다(Haidt, 2001; Haidt et al., 2000). 하이트가 제시한 사회적 직관주의 모델에 따르면, 도덕 판단은 도덕적 추론에 의해 인과적으로 도출되기보다는 진화적 축적물인 도덕 기반(moral foundation)으로부터 촉발된 강렬한 정서적 반응으로 나타나는 도덕적 직관에 의해 도출된다. 이때, 추론은 판단 이후에 나타나는 사후적 정당

1) 새로운 종합(new synthesis)이란 사회생물학이라는 학문 영역을 개척한 에드워드 윌슨(E. O. Wilson)에 의해 처음으로 사용된 용어이다(Wilson, 1975). 그는 인간의 모든 사회적 행동은 적어도 부분적으로 생물학적 관점에 의해 설명될 수 있음을 주지한다. 이러한 관점은 인간 행동에 대한 진화 심리학적 논의를 촉발시켰으며, 최근의 뇌-신경과학적 연구 결과에도 잘 부합한다. 이에 하이트는 도덕 심리학에서 '새로운 종합'의 접근이 필요하다는 문제의식하에, 도덕 심리학이 발달 심리학적 접근에서 나아가 진정한 의미의 학제적 접근을 통해 인간의 도덕적 삶을 사실적으로 기술하는 역할을 수행해야 함을 강조한다(Haidt, 2007). 이러한 맥락에서 본 논문은 '새로운 종합'의 접근을 수용함으로써, 도덕 심리학의 주요 연구 주제인 추론에 대한 다양한 분야의 연구 성과들을 폭넓게 검토해보고자 한다.

화(post-hoc) 기능을 주로 수행한다. 그러나 도덕적 직관은 집단 내에서 사람들을 협력적이고 상호의존적인 관계로 묶는 동시에 이념, 종교, 정치적 신념의 양극화를 조장하고 집단 간 갈등을 부추기기도 한다(Haidt, 2012/2014; Haidt & Graham, 2007). 이 과정에서 도덕적 추론은 종종 도덕적으로 그럴싸한 이야기를 만들어 비도덕적인 판단을 정당화하고, 자신과 다른 신념을 가진 타인·타집단에 배타적인 태도를 견지하는 근거로 사용되기도 한다.

이와 같은 도덕적 추론의 비도덕적 면모는 동기화된 추론(motivated reasoning)이라는 개념으로 적절히 분석될 수 있다. 동기화된 추론이란 추론의 인지과정이 동기적 요인에 의해 영향을 받는 현상을 의미한다. 이 개념은 사람들의 동기(motive)[2]가 정보를 처리하고 판단하는 방식에 어떻게 영향을 미칠 수 있는지를 중심으로 탐구하는 사회 심리학의 연구 흐름에서 비롯되었다. 동기화된 추론은 '동기'의 유형에 따라 때로는 정확하고 합리적인 추론의 모습으로 나타날 수도 있지만 대부분의 경우 자신의 기존 신념에 부합하는 방향으로 정보와 근거를 편향되게 해석하는 양상으로 나타난다. 동기화된 추론 개념을 최초로 제시한 사회 심리학자 지바 쿤다(Ziva Kunda)는 이러한 두 가지 양상 중 후자에 초점을 두었다(Kunda, 1987; 1990). 그녀는 동기나 정서가 편향된 인지 과정, 즉 신념에 접근하고 구성하며 평가하는 전략에 의존함으로써 추론에 영향을 미치는 현상에 주목한다(Kunda, 1990: 480). 특히, 과학적 연구 방법론의

2) 동기(motive 또는 motivation)란 '움직이게 하다'의 의미를 지닌 라틴어 'movere'로부터 유래한 것으로, 개인의 행동을 어떤 목표를 향하여 일정한 방향으로 작동시키는 심리 상태를 의미하는 심리학 용어이다(Reeve, 2015/2018). 동기를 설명하기 위한 이론은 크게 내용 이론과 과정 이론으로 구분되어 이해된다. 내용 이론(content theory)이란 '무엇이' 사람들의 동기를 유발하는지에 관심을 갖는 이론으로 대표적으로는 매슬로우(A.H. Maslow, 1987)의 욕구 단계 이론과 허즈버그(F. Herzberg, 1987)의 동기-위생 이론, 그리고 알더퍼(C. P. Alderfer, 1972)의 ERG 이론, 맥클랜드(D. McClelland, 1973)의 표출 욕구 이론 등이 대표적이다. 과정 이론(process theory)이란 '어떠한 과정을 거쳐서' 사람들의 동기가 유발되는지에 관심을 갖는 이론으로, 대표적으로는 학습 이론, 공정성 이론, 목표 이론, 기대 이론 등이 있다. 내용 이론은 동기의 분류에 따른 동기화된 추론 유형을 세분화해 주는 데 기여하고, 과정 이론은 동기화된 추론이 도출되는 심리적 기제를 설명하는 데 도움을 준다.

결여로 비판을 받았던 사회 심리학의 초기 이론인 인지부조화(cognitive dissonance) 이론은 최근 동기화된 추론 연구를 통해 재해석되고 있다(Festinger, 1957/2016; Mooney, 2012). 쿤다에 따르면, 인지부조화는 사람이 자신의 신념이나 태도와 행동 사이에 갈등을 경험할 때 발생하는 심리적 현상이다. 이 갈등은 심리적 불쾌함과 긴장감을 만들어내는데, 부조화로 인한 불편함은 우리가 부조화 요인을 줄이거나 제거하도록 동기부여 한다. 그녀는 사람들이 인지부조화 해소를 위해 자신의 기존 신념과 가치를 확인하는 정보를 선택적으로 찾고 해석하는 반면, 모순되는 증거는 무시하거나 합리화하는 경향을 포착하며 이를 동기화된 추론이라는 개념 틀로 설명하고자 하였다[3](Kunda, 1987; 1990; 1999).

쿤다의 연구 이후 동기화된 추론이라는 심리 기제는 사회 심리학 연구뿐 아니라, 추론과 의사결정에 관한 관련 학문 영역 전체에서 상당한 주목과 지지를 받았다. 그의 대표적인 연구(Kunda, 1990)는 발표 후 30년이 넘는 기간 동안 꾸준히 인용되어 2023년 현재 12000회 이상 피인용되었다. 쿤다의 연구 등장 이후 연구자들은 정치적 신념(Kahan, 2013; 2015), 사회 집단에 대한 태도(Lodge & Taber, 2006), 도덕 판단 및 행동(Ditto et al., 2009)을 포함한 다양한 영역에서 동기화된 추론을 탐구했다. 특히 동기화된 추론 연구에 따르면, 우리의 도덕적 추론은 도덕적 원칙을 선택적으로 적용하는 양상으로 드러날 수 있다. 이러한 점에 주목하여 몇몇 연구자들은 '동기화된 도덕적 추론(motivated moral reasoning)'이라는 용어를 고안하여 사용하기도 한다(Ditto et al., 2009; Paharia et al., 2021). 이외에도 동기화된 도덕적 추론의 개념을 명시적으로 사용하진 않지만, 도덕적 원칙이 동기화된 방식으로 적용된다는 연구(Uhlmann et al., 2009), 도덕성이 동기화된 추론을 정당화해준다는 연구(Cusimano & Lombrozo, 2020; 2021), 동기화된 추론이 우리로 하여금 자기중심적으로 행동하면서도 도덕적 주체로서의 느낌을 유지하는데 기

3) 넓은 의미에서 동기화된 추론은 정확한 판단에 이르고자 하는 동기까지를 포함하는 개념이나, 일반적으로 동기화된 추론 연구에서는 편향된 판단과 관련된 심리기제로 자신의 기존 신념에 부합하고자 하는 동기를 전제하는 좁은 의미로 한정하여 사용한다. 동기화된 추론의 의미와 유형에 관한 논의는 본고 2장 2절을 참조.

여한다는 연구(Ellis, 2022; Gino et al., 2016) 등은 도덕 판단에서 도덕적 추론이 동기화된 추론의 한 유형으로 분석될 수 있음을 보여준다.

동기화된 추론 관련 국내 선행 연구는 주로 의사결정 심리학의 연구 결과를 활용하는 경영학 및 행동경제학(김창윤 외, 2017; 이현준, 2007; 방광수 외 2013), 소비자 심리학(류강석 외 2004; 박주영, 1998)등에서 다뤄지고 있으며, 해당 연구들은 주로 비합리적인 판단을 이끄는 내재적 요인으로 동기화된 추론을 하나의 변수로 적용하여 수행하였다. 판단 및 의사결정, 사고 과정 자체를 연구 대상으로 삼아 동기화된 추론의 심리 기제를 다룬 연구는 정치적 판단과 관련된 김미라, 민영(2014)의 연구와 교육과 관련하여 과학적 사고를 방해하는 주요 요인으로서 동기화된 추론의 존재를 지적한 신세인, 이준기, 하민수(2018)의 연구가 있다. 이들의 연구는 정치의 영역과 과학 교육의 영역에서 동기화된 추론으로 인해 정보를 왜곡하여 해석하여 편향된 판단에 도달하게 되는 과정에 주목했다.

대표적으로 신세인 등의 연구는 교육적 탐구와 관련하여 동기화된 추론이 과학적 사고 발달에 미치는 부정적인 영향에 대해 조사했다(신세인 등, 2018). 이들은 쿤(Kuhn, 2002)의 관점에 따라 과학적 추론의 핵심을 이론과 증거의 조정(theory-data coordination)으로 보았다. 즉, 과학적 추론의 과정이란 자신이 기존에 가지고 있던 인식의 틀인 '이론'과 실제 현상에서의 경험을 통해 새롭게 얻어진 정보인 '증거'들 사이에서 새롭게 지식을 구성해가는 과정이다. 그러나 동기화된 추론은 '조정'의 대상이 되어야 하는 '이론'이 하나의 고정된 해석의 틀로서 작용하기 때문에 증거들 사이에서 적절하게 조정하지 못하고, 추론의 과정을 고착시킬 위험이 있다. 동기화된 추론은 자신이 가지고 있던 기존의 '이론'을 기준으로 이에 부합하는 증거는 과대평가하여 수용하고, 그렇지 않은 증거는 왜곡하여 이해하거나 거부하는 기제로 작용하기 때문이다. 이로 인해 과학적 추론의 과정에서 나타날 수 있는 다양한 관점의 해석이나, 토론 및 논증과 같은 시도가 중단되며, 과학적 추론의 발달을 위한 교육에도 부정적인 영향을 보인다. 이처럼 학습자는 자신이 기존에 가지고 있던 개념 틀을 유지하고, 자신의 의견을 관철시키기 위해 이와 반대되는 정보를 무

시하거나 재해석한다. 해당 연구는 여기에서 나아가 개인적 사고의 수준뿐만 아니라 사회-문화적 수준에서도 과학의 영역과 관련된 동기화된 추론의 문제점을 지적한다. 역사적으로 동기화된 추론은 과학이 여러 사회-문화적 신념이나 특정한 이데올로기적 목적을 달성하기 위해 도구화되어 악용되는 사례[4]가 많이 발견된다는 것이다. 이러한 문제점은 도덕 판단의 문제와도 직결된다. 최근 윤리적 이슈로 대두되고 있는 기후 변화, 백신 접종, 인공 지능 기술 문제 등은 과학·기술과 도덕 판단이 긴밀하게 결합되어 있는 문제이기 때문이다[5].

국내 도덕 심리학 및 도덕·윤리 교육계에서 동기화된 추론의 개념을 직접적으로 다룬 연구는 없지만, 인지 편향 및 인지적 왜곡으로 인한 도덕 판단과 의사결정의 결함에 대한 분석이 시도되고 있다(양해성, 2016; 2020; 이인태, 2016; 2022; 정창우, 2022; 정창우, 김하연, 2020). 양해성(2016)은 하이트의 사회적 직관주의 모델의 도덕교육적 함의를 인지 편향의 맥락에서 제시하였다. 그는 하이트의 연구가 상황 해석 인지의 심리 기제를 적절히 설명해준다고 보았으며, 이를 바탕으로 상황 해석 인지가 편향되지 않도록 학생들을 이끄는 것이 도덕 발달에 있어 중요한 과제임을 강조했다. 또한 양해성(2020)은 다원화된 사회의 갈등을 해결하기 위해 필수적으로 요청되는 중요한 시민적 역량으로 메타 인지 능력으로서의 성찰 역량을 제안했다. 이 역량이 부족한 개인의 도덕 판단은 왜곡된 정보와 암묵적 요소 및 비도덕적 욕구 등에 취약할 수 있다. 특히, 성찰이 결여된 탐구 능력의 발달은 편향된 판단과 자기기만 등의 부작용으로 이어질 수 있으며 다원화된 가치를 추구하는 개인 및 집단 간

4) 대표적으로는 우생학, 골상학 등의 학문은 특정 인종, 성별의 우월함을 입증하고자 하는 목표에 의해 동기화된 추론의 문제를 발생시켰고(Gould, 1996/2003). 급성호흡기증후군 사스(SARS) 및 조류 독감과 신종 플루에 한국의 김치가 탁월한 예방 효과가 있다는 연구(Kim & Park, 2017) 등은 국가 이미지 및 경제적 이익이라는 목적에 의해 동기화된 사례로 해석될 여지가 있다(신세인, 이준기, 하민수, 2018: 639에서 재인용).

5) 이와 관련하여 영미권을 중심으로 동기화된 추론의 연구는 기후 변화(Kahan, 2012), 원자력 발전(Kraft, 2015), 동물 복지(João et al., 2016) 등의 다양한 윤리적 이슈에 있어, 사람들이 가지고 있는 기존의 도덕적, 정치적 신념 및 태도에 따라 관련 정보나 지식을 상이하게 왜곡하는 현상에 주목한다.

의 갈등을 심화시킬 위험성마저 내포하게 된다고 지적한다.

이인태(2022)는 인지 편향에 대한 기존의 관점과 달리, 비도덕적인 행위를 합리화하는 기제로서 인지 왜곡의 심리 기제 분석을 통해 합리적인 추론을 통해서도 편향된 도덕 판단과 의사결정이 내려질 수 있음을 보였다. 그는 발달 심리학자 깁스(J. Gibbs)의 논의를 바탕으로 자기 본위적 인지 왜곡(self-serving cognitive distortion) 개념을 통해 비도덕적인 행위를 저지르고도 그 행위에 대한 근거를 제시함으로써 정당화를 시도하는 심리적 기제를 분석하였다(이인태, 2022: 220). 또한 그에 앞서 이인태(2016)는 인지 왜곡에 미치는 요인으로 또래 압력을 제시하였으며, 이 영향력을 중재하는 조절 요인으로서 도덕적 정체성이 기여함을 밝혔다.

정창우, 김하연(2020)은 향후 도덕교육의 개선 방향에 있어 인지 편향을 극복하고 이를 위한 도덕적 메타인지의 중요성을 강조했다. 후속 연구에서 정창우(2022)는 검토되지 않는 신념으로 인한 인지 편향의 위험성에 주목하며 이 문제를 개선하기 위한 도덕교육적 처방으로 도덕적 메타인지의 개념을 체계적으로 제시하고 구체적인 교육 방안을 검토하였다. 특히, 해당 연구는 도덕적 추론의 한계를 지적하며 이로 인한 편향된 판단을 경계하고, 도덕적 메타인지의 중재를 통해 도덕적 추론 자체가 성찰의 대상이 되어야 함을 강조했다.

이처럼 최근 도덕교육 학문 공동체는 더 높은 수준의 도덕 발달에 대한 관심과 함께, 비도덕적 행위를 유발하는 인지적 요인에 초점을 두며 이를 개선하기 위한 다양한 교육적 조치에 주목하고 있다. 이와 같은 도덕교육 학계에서 인지 편향과 인지적 왜곡에 대한 관심은 상위의 범주로서 '비도덕적 행위와 정당화의 심리 기제'를 파악하고자 하는 도덕 심리학의 주제와 밀접한 관련을 지닌다. 대표적으로 도덕적 정당화의 심리 기제에 대한 다각적 분석을 수행한 연구(추병완, 2020a)나 사회 인지 이론가인 반두라의 도덕적 이탈 개념과 이를 개선하기 위한 교육적 노력에 관한 연구(정창우, 2020a; 추병완, 2020b; 신원동, 2022), 행동윤리학의 관점에서 제한된 윤리성을 바탕으로 비도덕적 행위의 심리 기제를 분석하고 이를 개선하기 위한 도덕교육적 활용 방안을 탐색한 연구(추병완,

2021; 2022a) 등이 이에 해당한다.

편향된 도덕 판단을 유발하는 심리적 기제로 동기화된 추론에 관한 연구가 영미권 심리학계를 중심으로 광범위하게 수행되고 있지만(Cusimano & Lombrozo, 2020; 2021; Ditto et al., 2009; Ellis, 2022; Gino et al., 2016, Paharia et al., 2021; Uhlmann et al., 2009), 이에 대해 아직 국내 도덕교육 학계는 본격적으로 접근하지 않고 있다. 이는 도덕 판단 및 추론에 대해 발달심리학적 관점으로 접근하는 도덕교육 학계의 연구 경향에서 찾을 수 있다. 또한 동기화된 추론 연구가 수행되는 사회 심리학 연구 분야는 주로 도덕적 행동과 판단에 대한 상황주의적 관점에 가깝다는 점도 고려해볼 수 있다(노영란, 2011). 상황주의 심리학은 도덕적 행동을 개인의 인격 특질보다는 상황적 압력, 집단에의 순응, 사회적 정체성의 유지 등의 동기에서 비롯된다고 본다(Aronson et al., 2010; Fiske, 2010). 이는 훌륭한 성품에 대한 관심을 바탕으로 도덕적 인격 특질에 주목한 기존 도덕교육의 성향주의적 인격교육 관점과 상반된 입장을 취하고 있다. 이로 인해 우리 도덕교육 학계가 사회 심리학 연구에 상대적으로 덜 주목한 것으로 보인다. 특히, 상황주의 심리학이 도덕적 관점에서 상황적 압력과 집단적 순응 성향을 강조함으로써 인간 의지를 상대적으로 간과하는 점은 인간의 내적인 변화의 힘에 대한 믿음을 바탕으로 교육을 통해 학생들의 안정적인 도덕적 인격을 함양하고자 하는 도덕교육의 목표와 대립되는 것으로 이해될 수 있다. 그러나 상황주의와 성향주의가 도덕교육적 차원에서 상호보완적으로 작용할 수 있다는 최근의 논의(이기연, 2023)는 도덕교육계에서 사회 심리학적 연구를 폭넓게 수용할 수 있는 가능성을 제공한다. 나아가 도덕적 추론의 인지 과정이 정서적 반응에 결합되어 있다는 도덕 심리학 연구를 참고하여 도덕 판단의 심리 기제를 규명하려는 시도(김동창, 2015; 문경호, 2017; 박장호, 2012; 양해성, 2016; 정창우, 2011; 2012; 추병완, 2020c; 2022b)는 매우 활발히 이루어지고 있는데, 이러한 학적인 지평은 도덕교육의 관점에서 동기화된 추론 논의를 본격적으로 전개하기 위한 토대로 작용한다. 이에 본 연구는 도덕 판단에서 동기화된 추론으로 인해 발생하는 문제를 종합

적으로 분석하여 이를 개선하기 위한 도덕교육적 방안을 제안해보고자 한다.

제 2 절 연구의 방법 및 구성

본 논문은 도덕 판단의 과정에 영향을 미치는 동기화된 추론의 심리 기제를 밝히고, 그것으로 인해 발생하는 여러 문제에 주목하며 이를 개선하기 위한 도덕교육적 방안을 모색해보고자 한다. 이러한 목적에 도달하기 위해 본 논문은 다음과 같은 연구 문제를 설정하였다.

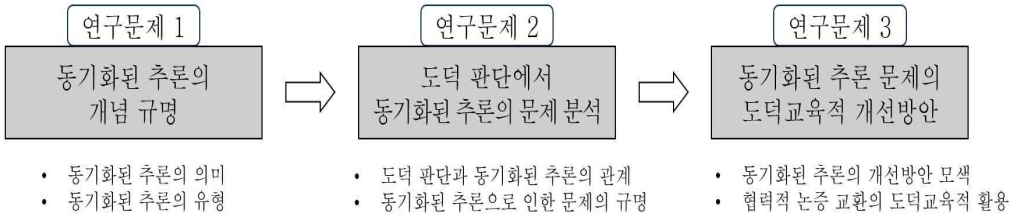
연구 문제: 도덕 판단에서 동기화된 추론으로 인해 발생하는 문제는 무엇이고 어떻게 개선할 수 있는가?

1) 동기화된 추론이란 무엇인가?
2) 도덕 판단에서 동기화된 추론은 무슨 문제를 유발하고, 그 원인은 무엇인가?
3) 도덕 판단에서 동기화된 추론의 문제를 개선하기 위한 도덕교육적 방안은 무엇인가?

이상의 연구 문제를 탐구하기 위해 본 연구는 탐색적 연구(exploratory research)를 위한 비판적 문헌 고찰(critical literature review)의 방법을 중심으로 탐구해갈 것이다. 탐색적 연구는 대상이 되는 이론 또는 해결해야 하는 문제의 특성을 명확히 하기 위한 연구 방법론이다(Stebbins, 2001: 3). 일반적으로 탐색적 연구는 새로운 문제나 쟁점을 제기하고 해답을 찾아갈 수 있는 올바른 방법이나 방향이 무엇인가를 제시할 때 사용한다(김구, 2011: 81-82). 이는 비판적 문헌 고찰을 통해 구체화 될 수 있다. 미국 심리학회(American Psychological Association, 이하 APA)에

따르면, 문헌 연구(literature review as a type of researches, literature-based research methodology)란, 연구를 종합하고 메타적인 분석을 포함함으로써 이미 출판된 자료를 비판적으로 평가하는 연구 방법론이다(APA, 2010: 37-38). 하트(Hart, 1998)에 따르면 문헌 고찰은 단순히 자료를 수집하거나 정리하는 차원이 아니라 연구자의 가치 지향이 드러나는 비판적인 학술 작업이다. 따라서 문헌 연구는 연구 문제와 관련된 광범위한 문헌들을 선택하고, 분류·분석하는 과정을 통해 연구 주제의 본질에 대해 이해하고, 그것의 타당성을 검토하여 효과적인 평가를 가능하게 한다(Hart, 1998: 27).

이와 같은 방법론을 활용한 연구의 구조를 도식화하면 다음과 같다.



[그림 1-1] 연구 구조

먼저 2장에서는 첫 번째 연구 문제를 탐구할 것이다. 도덕 판단에서 작동하는 동기화된 추론의 문제를 본격적으로 분석하기에 앞서, 해당 장에서는 동기화된 추론의 개념을 명료화하고자 한다. 사회 심리학자 쿤다(Kunda, 1990)에 의해 최초로 제안된 이 개념은 인식론적으로는 베이컨의 우상론과 흄의 회의주의라는 이론적 토대 위에서 논의된다. 또한 다양한 경험 과학적 연구들은 추론이 동기적 요인에 의해 영향을 받는다는 사실을 입증함으로써 동기화된 추론의 존재를 뒷받침한다. 이 장에서는 쿤다의 연구를 포함하여, 이후 동기화된 추론 개념을 경험적으로 입증해 주는 다양한 영역에서 수행된 관련 연구들을 살펴보며 동기화된 추론의 의미와 유형을 분석함으로써 도덕 판단에서 작동하는 동기화된 추론의 개념을 이해하기 위한 기초를 마련할 것이다.

3장에서는 두 번째 연구 문제를 탐구하기 위해 2장에서 살펴본 동기화

된 추론의 개념을 바탕으로 도덕 판단에서 동기화된 추론으로 인해 발생하는 문제를 규명하고자 한다. 이 장에서는 콜버그의 도덕 발달론을 중심으로 동기화된 추론에 대한 기존 도덕 심리학 연구의 이해를 살펴보고, 하이트의 연구를 통해 도덕 판단에서 작동하는 동기화된 추론의 존재를 확인한다. 이어 동기화된 추론으로 인해 발생하는 여러 문제를 주로 개인적 차원과 관련하여 비도덕적 행위의 정당화 문제와 사회적 차원과 관련하여 도덕적, 정치적 신념의 불일치를 심화시키는 문제를 검토할 것이다. 이러한 문제의 원인은 검토되지 않은 도덕적 신념을 보호하고자 하는 마음의 경향과 자신의 편향을 메타적으로 인식하지 못하는 편향의 사각지대 현상에 있다. 이는 도덕 판단에서 작동하는 동기화된 추론이 갖는 특징을 바탕으로 분석될 수 있다.

이어지는 4장에서는 3장에서 규명한 동기화된 추론 문제의 개선방안을 제안함으로써 세 번째 연구 문제를 해결하고자 한다. 구체적인 개선방안을 모색하기 위해 먼저, 동기화된 추론 문제의 원인이 변화 및 극복 가능하다는 사실을 보여주는 여러 이론과 연구를 바탕으로 개선 가능성을 점검할 것이다. 또한 구체적인 개선방안의 규준을 마련하기 위해 기존의 인지 편향 논의에서 개선의 아이디어로 제안된 도덕적 메타인지의 개념을 활용하되, 추론이 갖는 사회적 기능을 고려할 때 이를 효과적으로 촉진할 수 있다는 것을 논증할 것이다. 이를 바탕으로 도덕 판단에 관여하여 편향의 문제를 유발하는 동기화된 추론을 개선하기 위해 협력적 논증 교환이라는 구체적인 방안을 제안하고 그것의 교육적 타당성을 검토하고자 한다.

끝으로 5장에서는 4장에서 제안된 동기화된 추론의 개선방안인 협력적 논증 교환을 학교 도덕교육의 맥락에서 활용하기 위한 전략을 구체화한다. 이를 위해 도덕과 교수·학습 방법으로서 협력적 논증 교환을 활용한 수업이 갖추어야 할 조건을 목표, 양상, 교사의 역할 및 실행 전략 측면에서 제시할 것이다. 또한 도덕 교과에서 개발·적용되고 있는 기존의 교수·학습 모형들을 두루 검토함으로써 협력적 논증 교환 수업이 교육 현장에서 실질적으로 활용될 수 있는 방안을 제안할 것이다.

제 2 장 동기화된 추론의 이해

제 1 절 동기화된 추론의 이론적 토대

1. 인식론적 기초

1) 베이컨의 우상론

동기와 정서가 추론과 판단에 영향을 준다는 생각은 새로운 것이 아니다. 일찍이 영국의 철학자 베이컨(F. Bacon)은 인간 지성의 기능에 회의적인 시각을 견지하였다[6]. 『신기관(*Novum Organum*, 1620)』은 아리스토텔레스가 논리와 추론의 방법에 대해 저술한 『오르가논(*Organon*)』에서 영감을 받아, 근대 시기의 새로운 과학적 방법론을 제시한 인간의 지성(understanding)에 관한 베이컨의 저술이다. 베이컨은 연역적 추론과 철학적 지식에 의존하여 자연을 이해할 수 있다고 파악한 아리스토텔레스식 접근 방식에 대해 비판적이었다. 베이컨에 따르면 인간의 지

6) 서양철학사에서 베이컨의 인식론은 중세 스콜라 철학의 지배에 있던 서양의 지적 전통을 그로부터 해방시켜 근대 철학으로 넘어가는 과도기라는 지성사적 위치를 점하고 있는 것으로 평가된다. 베이컨은 신(神) 중심의 중세 철학의 인식론을 극복하고 인간이 지닌 지성이 비록 결함이 있지만, 이를 통해서도 신에 기대지 않고 자연의 진리를 탐구할 수 있음을 보이고자 하였다. 또한 그는 아리스토텔레스로 대표되는 전통적인 연역적 탐구 방법을 비판하며 새로운 진리 탐구의 방법으로서 귀납법을 제시하면서 근대 서양 철학에서 경험주의 전통의 초석을 쌓았다. 이러한 맥락에서 베이컨은 인간 지성의 결함을 지적하며, 추론이 마음의 특정한 편견과 선입견에 의해 제한된다는 생각에 중심을 두고 있으며, 이는 자연의 진리를 탐구하기 위해 극복되어야 한다고 믿었다. 동기화된 추론 연구가 베이컨의 인식론을 이론적 기반으로 하여 시작된 것은 아니지만, 이와 같은 베이컨의 인식론적 문제의식을 경험적으로 확인시켜주는 결과라고 볼만하다. 따라서 본고에서 주된 관심인 동기화된 추론의 문제와 개선 가능성이 충실히 검토되기 위해서는 베이컨의 인식론에 대한 이해를 기반으로 논의될 필요가 있다. 베이컨의 회의론적 인식론에 대한 평가와 관련한 논의는 폽킨(R. H. Popkin, 2003)을 참고할 수 있다.

성은 단순히 잘못된 신념을 잘못된 것으로 인식하지 못할 뿐 아니라, 심지어 모든 것을 자신의 잘못된 신념에 억지로 꿰어맞추고자 한다. 또한 그것이 잘못임을 알려주는 사례를 굳이 피하려는 경향이 있으며(박은진, 2006: 45), 따라서 우리는 이를 경계해야 한다. 이러한 경계의 대상으로 베이컨이 제시한 것이 바로 네 가지 우상론(Bacon, 1620/2001: 41-56)이다. 우상(idola, 偶像)은 인간 추론이 갖는 결함으로, 편향에 근거하여 잘못된 판단을 범하게 만들고 진리 습득을 방해하는 인간 마음의 모든 경향을 의미한다. 이는 인간의 인지가 편향에 취약함을 보여주는 인식론적 근거로 이해된다. 특히, 종족의 우상은 베이컨이 제시한 네 가지 우상 중 인간의 추론이 자기 중심적인 방향으로 동기부여 되어 이루어지는 양상을 지칭한다[7]. 다음은 베이컨의 『신기관』 중 일부이다.

> *인간의 정신을 사로잡고 있는 우상에는 네 종류가 있다. (편의상) 이름을 짓자면 첫째는 '종족의 우상' (Idola Tribus)이요, 둘째는 '동굴의 우상' (Idola Specus)이요, 셋째는 '시장의 우상' (Idola Fori)이요, 넷째는 '극장의 우상' (Idola Theatri)이다(Bacon, 1620/2001: 48).*

> *인간의 지성(understanding)은 사물의 빛을 있는 그대로 반영할 수 없는 울퉁불퉁한 거울과 같은 것이며, 그래서 자신의 본성과 자연의 본성을 섞어버림으로써 사물들의 본성을 왜곡하고 훼손한다(같은 책: 49).*

베이컨에 따르면, 종족의 우상이란 모든 것을 인간 중심적으로 해석하려는 마음의 경향을 의미한다. 인간의 지성은 의지와 감정의 영향을 받기 때문에 자칫 잘못된 판단을 범하게 된다. 가령, 사람들은 자신의 행동이 일정한 목적을 추구하고 있다는 사실로 미루어, 자연현상에도 그와 유사한 목적이 있을 것으로 미루어 생각한다. 종족의 우상이 비유하고자

---

7) 종족의 우상 이외에도 베이컨은 개인의 제한된 경험과 환경에서 비롯된 편견으로 세계를 인식하려는 종족의 우상, 언어의 오남용으로 인해 진정한 의미를 이해하지 못하는 시장의 우상, 무비판적으로 잘못된 원칙이나 학설, 전통과 권위에 기대어 판단하고자 하는 극장의 우상을 편향된 마음의 경향으로 지적하였다(Bacon, 1620/2001). 본고에서는 동기화된 추론 개념의 인식론적 기초가 되는 종족의 우상을 중심으로 검토하고자 한다.

하는 것은 인간 지성이 본성적으로 갖는 두 가지 오류이다. 하나는 인간 지성이 잘못된 신념을 가지면서도 인지하지 못한다는 것이고, 다른 하나는 잘못된 신념을 기준으로 모든 것을 해석하고자 한다는 것이다. 첫 번째 오류와 관련하여 베이컨은 과학의 구체적인 사례들을 제시한다.

> *인간의 지성은 그 고유한 본성으로 인해 (자신의 관점에서) 실제로 보이는 것 이상의 질서와 동등성이 존재한다고 생각하는 경향이 있다. 자연계의 많은 사물들은 본질적으로 속성이 다르고, 같은 것이 전혀 없음에도 불구하고 (인간의 지성은) 병행, 대응, 관계 따위를 찾아내는데, 그런 것은 실재하지 않는다. '천체들은 모두 완전한 원운동을 하고 있다' 는 환상 때문에, 나선이나 사행선운동(이런 명칭을 설명의 편의상 쓰고는 있지만)의 가능성은 아예 거들떠보지도 않는 것과 같다(같은 책 51-52).*

인간은 자연에 대해 갖는 자신의 신념이 잘못된 것인 경우에도 잘못되었음을 인식하지 못하기 때문에 오류를 범할 수 있다. 인간의 판단은 잘못된 신념에 기반하는 주관성에 따르는 것이기 때문에 병행, 대응, 관계 등과 같은 논리들을 근거로서 제시하지만, 이는 결함을 갖는다. 그럼에도 자신의 잘못된 신념을 인식하지 못하기 때문에 스스로 그것을 참이라고 받아들이는 것이다.

두 번째 오류는 더욱 심각한데, 인간의 지성이 기존에 가지고 있는 잘못된 신념을 바꾸려고 하기보다는 오히려 잘못된 신념을 유지하는 방향으로 세계를 이해하기 때문이다. 자신의 신념을 강화하는 것은 아전인수격으로 수용하고, 이에 반하는 것은 회피하거나 거짓으로 파악한다. 베이컨은 이러한 것들을 억지로 수용하거나 회피해야 할 대상이 아니라 극복의 대상으로 봐야 함을 강조한다. 특히, 그는 인간 본성상 자신에게 부정적인 것은 피하려는 경향이 강하기 때문에 오히려 인간은 자신의 신념에 반하는 부정적인 사례를 더욱 적극적으로 살펴봐야 한다고 주장한다.

> *인간의 지성은 (널리 승인되고 있거나 많은 사람들이 그렇게 믿고 있기 때문이든 아니면 자기 마음에 들어서이든) 한번 '이것이다' 하고 생각하고 나면, 다른 모든 것을 그것을 뒷받침하거나 그에 합치되도록 만든다. 아무리 유력한*

반증사례들이 있다 해도 무시하거나 경멸하거나 그것만 예외로 치부해 제외하거나 배척하고 만다. 이것은 순전히 처음에 내세운 주장의 권위가 손상될까 두려워하기 때문이다(같은 책: 53).

베이컨은 인간 지성이 갖는 이러한 결함의 원인을 감정의 영향과 지각 경험의 불완전성에서 찾는다.

또한 인간의 지성은 '마른 빛(dry light)'과 같은 것이 아니라 의지와 감정의 영향을 강하게 받기 때문에 이로 인해 인간은 자연과 세계를 있는 그대로 보지 못한다. 따라서 인간은 누구나 자기가 진실이기를 바라는 것을 더 쉽게 믿는다(같은 책: 55).

베이컨에 따르면 인간의 지성은 감정이 결여된 메마른 것이 아니다. 오히려 인간의 지성은 본성상 의지와 감정의 영향에서 자유로울 수 없기 때문에 올바른 탐구가 매우 어려운 것이다. 인간의 감정은 여러 방법으로, 특히 때로는 무의식적으로 지성을 오염시킨다. 그뿐만 아니라 인간 지성은 지각 경험의 불완전성으로 인해 결함을 갖는다.

인간 지성의 가장 큰 장애와 착오는 감각의 우둔과 무력과 기만에서 발생한다. 감각은 아무리 중요한 것이라 하더라도 직접적인 자극이 없는 것에 대해서는 무심히 대하는 반면, 직접적인 자극이 있는 것은 매우 중요하게 여기게 된다(같은 책: 56).

인간은 대체로 자신의 관점에서 자연을 해석하고 이해하려는 경향을 보인다. 이처럼 인간의 지성은 불완전함에도 우리는 이를 완전한 것으로 착각하며 추론의 결과를 인식론적으로 옹호한다. 이처럼 베이컨은 종족의 우상이라는 비유를 통해 인간 지성의 불완전함에 대한 무지 내지는 착각이 갖는 문제들을 지적하고자 하였다. 베이컨이 지적하고 있는 종족의 우상은 인간 중심적인 인지 편향의 현상만을 지적하는 것이 아니라, 그 배후에 있는 원인에도 주목한다. 인간의 지성은 대상을 관찰할 때 감각으로 인식되는 것만을 그대로 수용하지 않고, 자신의 기존 관점이나

선지식에 근거하여 해석하려는 경향을 지닌다. 그리고 그 해석은 주로 자기중심적인 방향으로 이루어진다. 같은 사물에 대해서도 다르게 반응하는 것은 각자가 가지고 있는 기존의 정보와 신념, 태도가 다르기 때문이며, 이는 인간의 판단과 추론의 과정에서 의식적으로도, 무의식적으로도 작용한다.

그러나 인간의 지성은 자연을 이해하고 진리를 탐구하는 데 유용하다. 베이컨은 이러한 추론을 적절히 해석하기 위해 추론에 따른 예단을 경계하되, 경험과 이성의 결합이 중요함을 강조하였다. 적절한 해석은 인간의 추론에 대한 의지와 감정의 영향과 경험의 불완전함을 인정하는 것에서 시작되는 것이다. 즉, 베이컨의 인식론은 인간의 판단에 있어 추론이 결함을 갖지만, 독단에서 벗어날 수 있다면, 추론을 통해 우리는 진리를 탐구하고 세상을 이해할 수 있다는 점을 시사한다.

2) 흄의 회의주의적 인식론

베이컨의 『신기관』이 출간되고 약 100년 후, 이성에 대한 회의주의적 시각은 스코틀랜드의 경험론자인 흄(D. Hume)에 의해 한 번 더 엄밀히 검토되었다. 흄의 회의론적 인식론은 인간의 모든 지식은 경험에서 파생되는 것이며, 어떤 것도 확실하게 알 방법은 없다는 생각에 기초하고 있다. 그는 또한 원인과 결과에 대한 우리의 이해가 이성이나 직관에 기초하는 것이 아니라 관습과 습관에 기초한다고 믿었다. 그는 원인과 결과에 대한 우리의 믿음은 세상의 규칙성을 관찰한 경험에서 비롯되며, 우리는 과거의 경험을 바탕으로 인과관계에 대해 추론한다고 주장했다.

원인과 결과에 관한 추론은 모두 오직 습관에서 유래하며, 믿음은 우리 본성에 있는 사유라는 부분의 활동이라기보다는 오히려 감수성이라는 부분의 활동이라고 하는 것이 더욱 타당하다(『논고(Treatise)』[8], 1권: 196).

---

8) 이하에서 『논고(*Treatise*)』는 흄(Hume)의 주저인 *A Treatise of Human Nature*(전 3권)을 지칭하며, 1권은 *Of the Understanding*(「오성에 관하여」, 이준호 역, 1994), 2권은 *Of the Passions*(「정념에 관하여」, 이준호 역, 1996), 3권

흄의 인식론은 모든 지식이 경험에서 파생된다는 생각에 기반을 두고 있다. 세상에 대한 우리의 신념은 불확실하며, 우리의 이해는 감각을 통해 우리가 이용할 수 있는 정보에 의해 제한된다. 또한 그는 인간의 추론과 신념이 정념에 의해 형성된다고 보았다. 예를 들어, 어떤 사람이 특정 결과를 믿고 싶은 욕망에 의해 동기 부여를 받는다면, 그는 무의식적으로 자신의 신념을 뒷받침하는 방식으로 정보를 찾고 해석하는 반면, 그것과 모순되는 정보는 무시할 수 있다.

우리의 정념과 상반되는 원리는 이성과 같은 것일 수 없고, 부적절한 의미에서만 이성이라고 일컬어지는 것으로 생각된다. 우리가 이 정념과 이성의 싸움을 말할 때, 우리가 말하는 것은 엄밀하지도 않고, 철학적이지도 않다. 이성은 정념의 노예이고 또 노예일 뿐이어야 하며, 정념에게 봉사하고 복종하는 것 외에 결코 어떤 직무도 탐낼 수 없다 (『논고(Treatise)』, 2권: 160).

흄은 우리의 신념과 행동을 형성하는 데 있어 정념의 역할에 주목한다. 그에 따르면 이성의 힘만으로는 신념을 형성할 수도, 어떠한 행동이나 욕구를 도출할 수도 없다. 인간을 움직이게 하는 동기부여의 힘은 감정 즉, 정념에서 비롯되며 이성은 특정 욕구를 거절하거나 선택할 수 없다(장동익, 2004: 193). 인간의 이성은 오직 정념에 의해 야기된 행동과 욕구를 충족시키는 수단을 탐색하고, 정념에 기여하는 방향으로 작동할 뿐이다. 이처럼 지식의 본질에 대한 흄의 회의주의적 인식론은 인간 이성의 한계를 강조한다. 우리의 신념은 객관적인 증거나 합리적 논증에 근거하기보다는 감정과 열정에 의해 영향을 받는 경우가 많기 때문이다.

이러한 견해는 도덕의 영역에서도 적용된다. 도덕적 구별은 이성이나 추론을 통해 도출될 수 없다. 특히, 도덕은 실천에 관한 것이며 이는 인간 지성의 판단을 넘어서는 것이기 때문이다. 이성이나 추론은 어떤 행동이나 욕구를 일으키거나 억제하는 힘이 없는 것처럼, 도덕을 나타내는 규칙 또한 이성이나 추론의 산물이라 할 수 없다. 즉, 이성만으로는 도덕

은 *Of Morals*(「도덕에 관하여」, 이준호 역, 1998)를 지칭한다.

적인 선과 악을 구별할 수 없는 것이다.

> *도덕성은 언제나 실천 철학에 포함되므로, 우리는 도덕성이 우리의 정념과 행동에 영향을 미치며 오성[9]의 차분하고 냉담한 판단을 넘어서는 것으로 가정한다. 그리고 우리는 일상 경험을 통해 이 사실을 확인한다. 우리가 일상 경험을 통해 인지하기로는, 인간은 흔히 자신들의 의무[감]에 지배되고, 불의라는 의견 때문에 어떤 행동을 단념하며, 책임이라는 의견 때문에 어쩔 수 없이 다른 행동을 추진한다(『논고(Treatise)』, 3권: 27).*

> *행동의 가치는 이성과 부합되는 속에서 유래되는 것이 아니고 그 허물도 이성과 상반되는 속에서 유래되지 않는다. […] 이성은 어떤 행동에 대해 부인하거나 찬동함으로써 직접 그 행동을 중단시키거나 유발할 수 없으므로, 그 행동을 중단시키거나 유발하는 영향력을 가진 것으로 밝혀진 도덕적 선, 악을 구별하는 원천일 수 없다. […] 행동의 잘잘못은 흔히 우리의 자연적 성향과 상충되며, 우리의 자연적 성향을 억제할 때도 있다. 그러나 이성에는 이런 영향력이 전혀 없다. 그러므로 도덕적 구별은 이성의 산물이 아니다(『논고(Treatise)』, 3권: 28-29).*

이처럼 흄의 인식론에 따르면 도덕적 신념과 태도의 형성은 객관적인 사실이나 원칙에 근거한 것이 아니라 다양한 상황과 행동에 대한 우리의 감정적 반응에 근거한다. 이성은 도덕 판단에서 부차적인 역할을 하며, 도덕적 신념은 감정적 동기부여에 의해 형성된다. 이성은 우리의 도덕적 신념을 지지하거나 도전하기 위해 사용될 수 있지만, 무엇이 옳고 그른지를 결정할 수는 없으며 이성과 추론만으로는 이러한 도덕적 신념을 만들어낼 수는 없는 것이다. 우리의 도덕적 신념은 관습과 습관에 의해 형성되고, 우리의 욕망과 감정에 의해 영향을 받는다. 나아가 그는 우리의 도덕적 신념이 특정한 행동에 대한 우리의 감정적 반응에 기초하고 있으며, 이는 다른 사람들의 행복이나 고통에 대한 우리의 시인과 부인의 감

9) 흄의 철학에서 오성(understanding)은 감각 기능과 대비되는 정신의 기능으로서 넓은 의미의 이성을 뜻한다(Bailey & O' Brien, 2006/2015). 오성은 논리적 사고, 개연성의 판단, 경험적으로 알게 된 관계를 파악한다. 즉, 이성은 판단과 분석 등을 수행하는 종합적인 추론 능력이다. 그러나 흄은 지식 탐구에서와 마찬가지로 인간 행동과 의지의 실천과 관련하여 이성의 역할에 회의적이다.

정에 의해 주도된다고 보았다. 따라서 우리의 도덕적 신념은 연역적 추론의 결과가 아니라 감정적 반응의 결과이다. 개인이 느끼는 도덕적 감정은 우리가 도덕적 신념에 따라 행동하고 도덕적 목표와 욕구를 추구하도록 동기를 부여할 수 있다.

또한 그것은 우리가 거짓되거나 비합리적인 도덕적 신념을 고수하고 그러한 신념을 바로잡을 증거에 저항하도록 이끌 수도 있다. 이처럼 감정에 근거한 신념과 판단은 편향될 수 있으며 감정이 향하는 방향에 의해 좌우될 수 있는 것이다. 예를 들어, 우리가 특정한 도덕적 입장에 대해 강한 감정적 애착을 가지고 있다면, 우리는 그 입장과 모순되는 증거를 무시하고 그것을 뒷받침하는 증거에만 집중할 가능성이 높다. 이에 따라 우리의 도덕적 신념은 불완전하거나 부정확한 정보에 기반할 수 있고, 기존의 신념을 반박하는 새로운 증거가 제시되더라도 이를 변경하는 데 저항하는 상황이 발생할 수 있다. 따라서 우리는 도덕적 신념과 판단을 이끌어내는 감정이 향하는 방향에 주목할 필요가 있다.

2. 경험과학적 근거

베이컨의 우상론과 흄의 회의주의적 인식론은 최근 추론에 관한 실험심리학 연구들에 의해 인지 편향이라는 현상으로 설명된다. 인지 편향은 그릇된 판단과 의사결정을 유발하는 체계적인 사고의 오류이다(Haselton et al., 2005/2019: 565-566). 특히, 인간의 판단과 의사결정에 대한 사회심리학 연구는 인지 편향을 주된 연구 주제로 삼아 이를 경험 과학적으로 입증하고자 한다. 사회 심리학이란 개인이 타인과 어떻게 상호작용하고 이러한 상호작용이 우리의 생각, 정서, 행동에 어떻게 영향을 미치는지를 연구하는 심리학의 한 분야이다(Aronson et al., 2010; Fiske, 2010). 사회 심리학적 관점에서 인지 편향은 우리의 인식, 기억, 그리고 추론에 영향을 미치면서 널리 퍼져있고, 주로 무의식 수준에서 처리된다. 인지 편향의 개념은 사회 심리학 및 사회인지이론 분야에서 광범위하게 연구되어 왔으며, 우리의 태도와 신념을 형성하는 데 중요한 역할을 한다.

사회 심리학의 중심 주제 중 하나는 우리의 사고, 정서, 행동이 우리 주변의 사람들과 상황에 의해 영향을 받는다는 생각이다. 이것은 사회적 영향으로 알려져 있으며 동조, 복종, 설득, 그리고 사회적 규범과 같은 많은 형태를 취할 수 있다. 예를 들어, 동조의 개념은 개인이 집단의 태도, 신념, 행동을 채택하는 경향을 의미한다. 사회 심리학자들은 사람들이 종종 다른 사람들의 의견에 흔들린다는 것을 보여주었다(Cialdini & Goldstein, 2004; Aronson et al., 2010). 이는 개인이 타인의 의견을 정보의 원천으로 보는 경우가 많고, 집단에 의해 수용되기를 원하기 때문이다. 반면에 복종은 개인이 권위 있는 인물의 명령을 따르려는 의지를 의미하며, 심지어 이러한 명령이 그들 자신의 도덕적 신념과 충돌할 때도 마찬가지이다. 사회 심리학자들은 스탠리 밀그램의 유명한 복종 연구(Milgram, 1974)를 포함하여, 많은 다른 맥락에서 복종을 연구해왔다. 설득은 의사소통을 통해 다른 사람의 태도나 신념을 바꾸는 과정을 말한다(Petty & Cacioppo, 1986; Chaike et al., 1989). 사회 심리학자들은 메시지의 출처, 메시지 자체, 그리고 수신자의 특성을 포함하여 설득에 영향을 미치는 많은 요소를 연구해왔다. 마지막으로, 사회 규범은 특정 사회 집단이나 문화에서 수용 가능할 수 있거나 적절한 행동으로 간주하는 것에 대한 암묵적이거나 명시적인 규칙을 가리킨다(Cialdini & Goldstein, 2004). 사회적 규범은 우리의 의사결정과 행동에 강력한 영향을 미칠 수 있으며, 이러한 규범을 위반하는 개인은 배제나 처벌과 같은 사회적 제재에 직면할 수 있다.

사회인지이론(social cognition theory)은 개인이 다른 사람과 사회적 상황에 대한 정보를 어떻게 처리하는지를 설명하는 사회 심리학의 하위 분야이다. 이 이론은 인간의 사회적 행동, 태도, 정서를 이해하고 예측하는 능력의 기초가 되는 정신적 과정을 탐구한다(Kunda, 1999; Zimmerman, 2000; Bandura, 2001). 의사결정과 추론은 사회인지의 핵심 요소이다. 의사결정을 할 때, 사람들은 이용 가능한 정보를 저울질하고, 여러 대안의 잠재적인 결과를 평가하고, 선택을 해야 한다. 반두라는 이 과정을 행동적 요인, 개인적 요인(믿음, 인지, 기술, 영향), 사회/환경 영

역 내의 요인들 사이에 상호작용으로 설명한다. 자기효능감은 행동과 성취와 관련된 중요한 개인적 요인이다.

> *자신이 자기 효능감을 가지고 있다는 신념은 환경에 대한 적응과 변화에 영향을 미칠 뿐만 아니라 (판단과 행동의) 결정 요인에 영향을 미친다. 그러한 신념은 사람들이 비관적으로 생각하는지 낙관적으로 생각하는지 그리고 자기를 향상시키는 방식으로 생각하는지 또는 자기를 방해하는 방식으로 생각하는지에 영향을 미친다(Bandura, 2001: 10).*

따라서 동기가 유지될 때 인지 주체가 수행하는 행동은 자신의 노력 또는 과제 선택에 대한 지속성과 관련된 자기효능감의 영향을 받는다. 사회적 환경에서 누군가가 "나는 네가 할 수 있다는 것을 안다"라고 말하는 긍정적인 강화를 경험한다면, 이는 자기 효능감에 대한 신념으로 이어진다. 그러한 신념은 인지 주체에게 자신의 행동, 정서, 생각에 대한 많은 자기 조절 능력으로 이어진다. 자기 조절은 개인적인 목표를 달성하기 위해 계획되고 주기적으로 적응되는 생각, 정서, 행동을 의미한다 (Zimmerman, 2000: 14). 우리는 특정 행동이나 학습 목표를 달성하기 위해 과거의 경험을 참고하여 전략을 개발한다. 또한 인지 주체는 자신의 내부 기준과 목표에 따라 행동을 조절한다. 이는 다시 자신이 어떠한 결과를 산출해낼 수 있고 행동을 수행할 수 있다는 신념, 즉 자기 효능감을 강화하고 동기화를 촉진한다. 동기화는 자신의 행동이 긍정적인 결과를 가져올 것으로 기대하는 정도에 비례하여 지속된다. 자신의 행동으로부터 성공적이고 긍정적인 결과를 얻을 것이라는 믿음이 인지 주체를 움직이게 하는 강한 동기가 되는 것이다.

사회인지이론이 전제하는 인간은 자신의 경험과 이로부터 접하는 정보를 바탕으로 세상에 대한 이해를 끊임없이 구성하고 재구성하는 적극적인 정보 처리자이다. 그러나 이 과정이 항상 완전히 정확한 것은 아니며 이전 신념, 정서 및 사회적 맥락과 같은 다양한 요인의 영향을 받을 수 있다. 특히, 사회인지이론은 인지 편향이 긍정적 자아상을 유지하도록 동기화된 개인은 정체성을 보호하려는 방향으로 정보를 해석하는 과정에서

발생할 수 있음을 강조한다. 사회인지이론에서는 이를 정체성 보호 인지(Identity protective cognition)로 개념화한다(Kahan, 2017; Stanovich; 2021/2022). 이는 문화적으로 다양한 개인들이 그들의 집단에서 지배적인 신념을 반영하는 패턴으로 증거를 선택적으로 수용하고 무시하는 경향을 의미한다. 이와 같은 개념은 우리의 인지 편향이 단지 비합리적인 정서적 개입으로 방해되는 것이 아니라, 인지적 작동에 의해 합리적인 추론의 결과로도 도출될 수 있음을 시사한다.

사회 심리학과 사회인지이론의 초기 연구를 촉발시킨 페스팅거(Leon Festinger, 1957/2016)는 인간 심리 내부의 인지요소[10]들 간의 충돌로 인한 부조화[11] 상태와 이로부터 내적인 일관성을 유지하고자 하는 심리적 기제로 인한 다양한 인지적 전략의 사용을 인지부조화 이론을 통해 설명하고자 한다. 인지부조화 이론은 특정한 대상에 대해 인간이 가지고 있는 인지적 요소(지식, 의견, 신념 등)들이 내적 일관성 유지의 동기에서 기인한다고 본다. 인지부조화 이론은 사람들이 자신과 다른 의견이나 신념, 판단의 결과에 따른 심리적인 불편함을 줄이기 위해 자신의 입장을 지지하는 정보만을 수용하고, 그렇지 않은 정보는 거부하거나 왜곡하는

10) 페스팅거는 인지(cognition)의 개념을 주위 환경이나 자기 자신, 또는 자신의 행동에 관한 지식이나 의견 또는 신념이라는 의미로 사용한다(Festinger, 1957/2016). 또한 내적인 일관성을 유지하려는 마음의 심리적 기제를 조화(consonance)의 개념을 통해 설명하고자 한다. 즉, 인지부조화란 자신의 의견이나 신념을 구성하는 지식(knowledge) 등의 인지요소들 사이의 조화롭지 않은 상태로 인해 발생하는 불쾌한 상태로, 인간은 인지부조화의 감소를 지향하며 이는 강력한 행위의 동기가 된다(Festinger, 1957/2016: 20).

11) 페스팅거는 '부조화'의 개념을 여러 인지 내용들 사이에 서로 부합하지 않는 관계가 존재하는 상태로 개념 정의하며, 이를 엄밀한 논리학적 개념인 '비일관성(inconsistency)'을 대신하여 사용하였다(Festinger, 1957/2016: 19). 페스팅거는 '생각 사이의 불일치'와 '불일치로 인한 심리적 불편함'을 모두 부조화의 의미로 사용하였는데, 인지 심리학자 다니엘 록스턴은 이 둘이 별개의 개념임을 지적한다(Loxton, 2021: 68). 실제 불일치가 있다는 사실을 알아차리지 못하면 불편함도 느끼지 않을 것이기 때문이다. 인지부조화가 발생하기 위해서는 충돌하는 두 가지 인지 요소를 동시에 품고 있어야 하는데, 실제로 부조화를 유발할 요소가 공존하고 있음에도 그것을 알아차리고 있지 않다면 이는 충돌을 일으키지 않을 것이다. 따라서 이러한 맥락에서 '인지부조화'를 엄밀하게 정의한다면, 자신이 인지한 '생각 사이의 불일치'로 인해 발생하는 심리적 불편함이라 할 수 있으며, 이는 일치시키고자 하는 심리적 경향으로 인해 일정한 행위를 유발하는 동기가 된다.

현상을 설명해준다.

인지부조화 이론이 등장할 당시의 기존 주류 심리학 이론은 인간이 내면의 '일관성'을 추구한다고 암묵적으로 전제하였다. 인지부조화 이론은 이 전제에 대한 인간의 비일관적인 태도를 설명하는 새로운 해석 틀을 제공하려는 시도였다(Festinger, 1957/2016: 17). 인지부조화 이론은 다음과 같은 기본 가설을 설정한다.

> *1. 부조화의 존재는 심리적 불편함을 느끼게 하므로 부조화를 감소시켜 조화를 달성하려는 동기를 유발할 것이다.*
>
> *2. 부조화가 발생하면 그것을 감소시키려 할 뿐만 아니라, 부조화를 증가시킬 가능성이 있는 상황이나 정보를 적극적으로 회피하고자 할 것이다 (Festinger, 1957/2016: 19).*

페스팅거는 여러 인지 요소들 사이에서 서로 부합하지 않는 관계가 존재하는 상태인 인지부조화 상태 자체가 인간 행동의 동기 요인이라고 보았다. 인지 요소들 사이에 부조화가 발생하게 되면 불쾌감을 느끼며, 이때 자신의 자존감을 지키고, 자아 개념을 유지하기 위해 불쾌감을 해소하고자 하기 때문이다. 불쾌감의 해소, 즉 인지 요소들을 조화시키는 방법[12]은 크게 세 가지가 있다.

첫째, 우리의 행동과 부조화를 유발한 인지 요소가 일치하도록 변화시키는 방법이다. 부조화가 주위 환경에 관한 특정 지식에 대응하는 인지요소(환경적 요소)와 행동적 인지 요소 사이에 발생하는 것이라면, 이 부조화는 환경적 요소에 부합하도록 행동적 인지요소를 수정함으로써 제거될 수 있다(Festinger, 1957/2016: 41). 즉, 인지부조화를 일으키는 행동이나 태도를 바꾸는 것으로, 인지부조화를 해결하는 가장 간단한 방법에

12) 피아제와 콜버그 등의 인지 발달론자들은 인지부조화의 발생과 해소를 도식의 평형화 기제로 이해한다. 자신이 가지고 있던 기존의 도식으로 동화가 되지 않는 경우, 조절을 통해 도식을 질적으로 확장시킬 것이다. 그러나 인지부조화 이론은 부조화 상태로 인해 '자기 개념(self-concept)'의 중요한 요소가 위협 당할 때에는 도식을 조절하지 않는 경향을 발견하였다. 이 경우에도 여전히 심리적 불편함을 느끼기 때문에, 이를 해소하기 위해 정보와 상황을 왜곡함으로써 기존의 도식을 유지시킨다.

해당한다. 예를 들어, 계속 담배를 피우고자 하는 욕구를 지니면서 건강하게 살고자 하는 흡연자가 흡연이 건강에 해롭다는 정보를 접하게 된다면, 그는 인지부조화를 줄이기 위해 담배를 끊을 수 있는 것이다. 그러나 이러한 방법은 강한 저항에 직면한다. 우리는 행동의 변화가 수반하는 고통이나 손실이 매우 크거나, 해당 행동이 다른 측면에서는 만족스러운 경우, 또는 행동의 변화 자체가 단순히 불가능한 경우에는 행동을 변화시키기 어렵기 때문이다. 즉 금연으로 인해 발생하는 불안 증세와 같은 금단 현상이나 스트레스가 심한 경우, 흡연이 건강을 해치기는 하지만 순간의 긴장감을 해소하거나 사회적 상호작용을 촉진한다는 점에서 만족을 주는 경우 등으로 인해 부조화를 해소하는 데 실패할 수 있다.

둘째, 부조화 인지 요소 하나를 변화시켜 우리의 행동을 정당화하는 방법이다. 때로는 인지부조화 상황에서 상황 자체를 변화시킴으로써 환경적 인지 요소를 바꾸는 것도 가능하다. 주로 물리적인 환경보다는 사회적 환경에 해당한다. 가령, 흡연이 즐거움을 제공하고, 긴장을 늦추게 하거나 정신적 스트레스를 낮추기 때문에 건강에 해로운 것이 아니라고 생각하며 스스로를 설득하는 것이다(Gibbons et al., 1997). 이러한 방식을 통해 사람들은 자신이 가지고 있는 상충되는 신념, 태도 또는 행동을 정당화하거나 합리화함으로써 인지부조화를 줄일 수 있다. 그러나 이 경우에도 저항이 있다. 특정한 인지 요소에 대응하는 환경적 요소가 실제 존재하는 것이 확실한 경우에는 정당화가 어려우며, 그 요소가 모호한 경우에도 사회적인 지지가 없다면 정당화되기 어렵기 때문이다. 담배를 태우면서 발생하는 화학 성분이나, 담배 자체에 포함된 건강 악화 성분의 효과가 과학적 사실로 드러난 경우에는 정당화를 통해 인지부조화 상태가 해소되기 어렵다. 나아가, 인지 요소에 대응하는 실재가 사회적인 환경과 관련 경우, 즉 타인과의 상호작용에 의해 형성된 경우에는 자신의 정당화를 지지하는 환경이 부조화 해소 여부를 결정한다.

셋째, 새로운 인지 요소를 추가하여 우리의 행동을 정당화하는 방법이다. 이 방법은 부조화를 완전히 제거하진 못하지만, 부조화의 총량을 감소시킬 수 있다. 건강상의 위험을 알면서도 흡연을 계속하는 사람은 자

동차 사고로 인한 사망률을 조사하여 비교할 수 있다. 흡연으로 인한 사망보다 자동차로 인한 사망의 가능성이 더 높다는 요소를 추가함으로써 부조화 상황을 일정 부분 해소할 수 있는 것이다. 사람들은 인지부조화를 해소하기 위해 자신에게 필요하다고 생각하는 새로운 인지 요소를 지지하는 사람과는 상호작용을 유지하고자 하지만, 그 반대의 경우, 즉 부조화를 유발한 인지 요소를 지지하는 사람과 의견을 나누는 것은 피하고자 한다. 사람들은 또한 조화로운 상태를 증가시키는 새로운 요소를 추가할 것으로 기대되는 정보의 출처는 살펴보려고 하겠지만, 부조화를 증가시키는 정보의 출처는 피하려고 한다(Festinger, 1957/2016: 55).

페스팅거는 일단 확신이 정해지면 그것을 바꾸는 것이 매우 어렵다는 사실을 관찰했다(Festinger, 1957/2016: 56). 사람들은 그 확신을 지키기 위해 현실을 부정하거나, 변명을 둘러대거나, 고의적으로 오해하거나, 객관적 증거를 무시하는 경향을 보인다. 이처럼 다양한 방법을 사용하여 자신의 신념을 보호하고 인지부조화가 유발한 불쾌감을 해소하는 것이다. 심지어 그는 자신의 신념이 명백히 틀린 것임을 입증하는 증거가 제시되는 경우에도 믿음을 쉽게 바꾸지 않을 것이라고 예측하였다. 페스팅거는 이를 검증하기 위해 관찰과 실험 결과를 제시한다[13].

페스팅거의 인지부조화 이론 이래로 동기가 태도(Festinger, 1957/2016) 뿐만 아니라, 지각(Erdelyi, 1974), 귀인(Heider, 1958) 등에도 영향을 미칠 수 있다는 명제는 1960년대까지 널리 수용되었다(Kunda, 1990: 480에서 재인용). 초기 연구자들은 특정 동기가 이기적이고 자기중심적인 귀인을 만들고 자신이 보고 싶은 것만 보고 믿고 싶은 것을 믿

13) 대표적으로 사이비 종교 집단 관찰 연구가 있다. 지구 종말 예언을 믿는 사이비 종교 집단에 참여한 페스팅거 연구진은 예언이 틀렸다는 것이 드러난 상황에서 집단 구성원이 두 부류로 구분되는 것을 확인했다. 단체를 떠난 사람들과 여전히 예언을 믿는 사람들이었다. 예언을 마지막까지 믿은 사람들은 결코 자신들의 신념이 틀렸다고 생각하지 않았다. 예언에 착오가 있다는 식의 변명을 통해 다른 날짜에 멸망할 것이라고 예언을 변경하거나, 다른 방식으로 예언이 실현되었다고 주장하였다. 그들은 서로를 위로하며, 새로운 주장들을 만들어내는 모습을 보였다. 페스팅거는 집단을 떠난 신도 집단과 마지막까지 믿었던 집단 사이의 차이에 주목했다. 예언을 믿기 위해 상대적으로 정신적, 물질적 투자를 많이 한 사람들은 자신의 신념을 지키기 위해 마지막까지 새로운 주장을 만들어냈던 것이다. 이와 관련한 자세한 연구 설계 및 결과는 페스팅거의 저서(1956/2020)를 참고할 것.

게 할 수 있다는 것을 당연하게 여겼지만, 이 견해와 그것을 뒷받침하기 위해 사용된 연구는 1970년대에 집중적인 비판을 받았다(Kunda, 1990: 480). 인지부조화 이론에 대한 가장 강력한 비판은 연구 방법론에 대한 비판이었다. 사람들이 충돌하는 인지 요소가 동시에 나타날 때, 이러한 마음속의 충돌을 규명하는 객관적인 방법이 부재하다는 것이다. 당시의 인지부조화를 다룬 연구에서, 인간의 인지가 특정한 동기에 의해 도출되었다는 것을 증명한 연구들은 비동기적 용어로 재해석될 수 있었기 때문이다(Bem, 1972; Nisbett & Ross, 1980; Tetlock & Levi, 1982, Kunda, 1990: 480에서 재인용). 따라서 사람들이 자기중심적인 결론을 도출할 수 있었던 것은 그들이 원했기 때문이 아니라 그들의 이전 신념과 기대를 고려할 때 이러한 결론들이 더 그럴듯해 보였기 때문이라는 것이다(Kunda, 1990: 480). 즉, 인지부조화 해소의 메커니즘으로 설명되었던 인지적 왜곡 현상은 정서적 반응이라기보다는 합리적인 인지 처리 과정의 일환으로 볼 수 있다는 지적이다.

한편, 90년대 이후 신경과학 연구의 발전을 바탕으로 인간의 인지 과정에 대해 보다 객관적인 방법으로 접근하게 되면서, 인지부조화의 존재가 사실적 차원에서 규명되고 있으며(van Veen et al., 2009), 특히, 추론 과정 자체에 동기적 요인들이 강력한 역할을 수행한다는 것이 밝혀지고 있다(Kunda & Sinclair, 1999). 또한 fMRI 등을 활용한 뇌-신경과학적 연구는 인지의 과정과 정서적 경험을 분리하여 이해하기 어렵다고 지적한다(May, 2021).

일부 신경과학자들은 사람들이 인지부조화 상태를 감소시키려고 동기화될 때 뇌의 어떤 부위가 활성화되는지 추적하였다(Harmon-Jones, 2010; Westen et al., 2006). 그들은 fMRI를 활용하여 사람들이 부조화를 경험하는 동안 신경 활동을 관찰하고, 부조화가 발생하는 시간 동안에 활성화되는 뇌의 부위를 발견했다(Harmon-Jones, 2010). 해당 부위는 선조체(corpus striatum)와 전전두피질(prefrontal cortex; PFC)인데, 선조체는 대뇌피질의 정보를 받아서 보상, 집행, 자기 조절 및 운동 처리에 관여하는 중요한 뇌 영역이고, 전전두피질은 전두엽의 앞부분을 덮고 있

는 대뇌 피질로, 충동을 억제하거나 계획 및 판단, 결정하는데 주요한 역할을 수행하는 부위이다. 또한 뇌에서 추론을 담당하는 영역이 부조화 정보를 접하면 실제로 활동을 정지하며, 부조화가 해소되었을 때에는 뇌의 정서회로가 활성화된다는 사실을 발견하였다(Westen et al., 2006). 웨스턴 연구진은 추론 과정에서 정서를 조절하는 데 관여하는 부분이 활성화되는 것에 주목한다. 이들에 따르면 자신이 가진 기존 정보에 부합하지 않는 새로운 정보에 직면했을 때, 뇌는 그 정보를 원하는 결론에 도달할 때까지 계속해서 인지적 만화경(cognitive kaleidoscope)을 작동시킨다. 마치 만화경을 돌려 원하는 그림 모양으로 맞춰 보듯, 우리의 뇌는 추론을 통해 인지부조화로 인한 불편한 감정 상태를 제거하고, 긍정적인 감정 상태를 활성화하면서 기존의 신념을 강화하는 경향을 보인다.

인간의 뇌가 작동하는 방식과 유사한 영장류의 뇌를 촬영한 실험은 인지부조화가 적응적 기능을 수행함을 보여준다. 폴 블룸 등이 수행한 인지부조화의 근원을 조사한 한 연구(Egan et al., 2007)는 미취학 아동과 카푸친 원숭이를 대상으로 '선택하지 않은 선택지'에 대한 태도를 바꿀 것인지 결정하도록 요구했다. 실험에 참가한 아동과 원숭이는 모두 똑같이 매력적인 두 가지 선택지(색깔이 다른 초콜릿)를 제시받았고 그중 하나를 선택한 후 나머지 선택지에 대해 가치를 평가하였다. 이러한 요구는 참가자들에게 두 선택지가 동등한 가치를 지닌다는 기존의 신념과 상충되기 때문에 인지부조화를 유발했다. 실험 참가자들은 선택하지 않은 선택지에 대한 태도를 바꾸고(가치를 더 낮게 매김) 그것이 덜 매력적이고 바람직하지 않은 선택지라고 답했다. 블룸은 이러한 연구 결과를 두고, 인지부조화가 청소년기 이후에 사회화 과정을 통해 획득된 특성이라기보다는 발달적으로 더 이른 시기에 그리고 진화적으로 오랜 적응의 결과로 해석될 여지가 있다고 제안하였다. 이와 같은 연구들은 우리의 추론이 정서와 긴밀하게 결합이 되어있으며, 추론의 과정에 동기적 요인이 영향을 준다는 사실을 밝혀냄으로써 동기화된 추론 개념의 경험과학적 기초를 제공한다.

제 2 절 동기화된 추론의 의미와 유형

1. 동기화된 추론의 의미

앞서 살펴본 것처럼, 우리의 추론은 순수하게 추상적으로 이루어지기보다는 다양한 동기적 요인에 의해 영향을 받는다. 베이컨과 흄으로 대표되는 회의주의적 인식론과 경험과학적으로 입증되고 있는 인지부조화 이론의 핵심적인 통찰들은 최근 동기화된 추론(motivated reasoning) 개념으로 통합되는 추세이다(Mooney, 2012: 50).

먼저, 동기화된 추론의 개념을 분석하기에 앞서 상위 개념인 추론(reasoning)과 추리(inference)의 인지과정에 대한 심리학적 접근을 살펴볼 필요가 있다[14]. 일반적으로 심리학에서 추론은 추리(inference)의 일종으로 여겨진다. 추리(inference)의 과정은 인간을 포함한 유기체 전반에서 확인되는 의사결정의 과정으로서, 사용 가능한 정보를 기반으로 특정 결론에 도달하는 인지 과정을 의미한다(May, 2021; Mercier & Sperber, 2017; Moshman, 2004). 추리는 유기체가 가진 제한된 정보를 바탕으로 예측과 판단을 할 수 있도록 하며, 이는 주로 무의식적으로 일어나지만, 우리는 추리의 과정을 의식하기도 한다. 추리는 종종 세계에 대한 우리의 이해의 격차를 좁히는 방법으로 여겨지는데, 우리가 모르는 것에 대해 예측하기 위해 우리가 알고 있는 것을 사용하는 과정이기 때문이다. 추론(reasoning)이란 기존에 가지고 있던 지식과 정보로부터 논증을 통해 새로운 신념을 도출하거나 판단을 이끌어내는 인지 과정을 의미한다. 추론의 과정은 '이유를 대는' 모습으로 표출된다(May, 2021; Mercier & Sperber, 2017; Moshman, 2004). 추리와 마찬가지로 추론은 기존 지식과 정보를 사용하여 새로운 결론에 도달하거나 문제를 해결하

14) reasoning과 inference의 번역어는 학계 및 학자에 따라 추론(推論)과 추리(推理)를 호용하여 사용한다. 본고에서는 도덕 심리학 및 사회 심리학계의 통상적인 번역을 따라 reasoning을 추론으로 inference를 추리로 번역하여 사용하고자 한다.

는 과정으로 정의할 수 있는데, 여기에는 정보에 입각한 판단을 내리기 위해 정보를 논리적으로 분석하고 비판적으로 평가하는 인지 능력을 사용하는 것이 포함된다. 추론은 연역적이거나 귀납적일 수 있다. 연역적 추론은 일반적인 원리에서 시작하여 특정 결론을 도출하는 데 사용하는 것을 포함하는 반면 귀납적 추론은 특정 관찰에서 시작하여 일반적인 결론을 도출하는 데 사용한다.

현대 심리학의 초석을 마련한 윌리엄 제임스는 이와 같은 추론에 대한 이해를 다음과 같이 명료화하고 있다(James, 1968/2014: 466-468).

> *새로운 정보(data)를 다루는 능력이 추론(reasoning)의 본질적 특성이다. (중략) 추론은 분석(analysis)과 추상(abstraction)을 수반한다. 여기에 유사한 것이라곤 아무것도 암시하지 않는 어떤 한 가지 사실이 있다고 가정해보자. 이런 경우에 단지 경험만을 근거로 하는 사상가라면 그 사실을 있는 모습 그대로 응시할 것이지만, 추론에 탁월한 사람은 그 사실을 잘게 쪼갠 다음에, 그 사실이 지닌 별도의 속성 중 하나에 주목할 것이다. 이 속성을 그 사람은 자신의 앞에 놓인 사실의 근본적인 부분으로 받아들일 것이다. 이 속성은 그때까지 그 사실이 갖고 있는 것으로 알려지지 않은 특성 혹은 중요성을 갖고 있다.*

제임스에 따르면 추론의 본질은 '새로운 정보'를 해석해내는 것에 있다. 이성의 고차원적인 사용으로서 추론은 하나의 사실을 분석하고, 그로부터 추상적인 속성을 파악하는 과정을 포함한다. 이를 통해 개인은 신념을 형성하고 수정하며, 결정을 내리게 된다. 제임스는 추론의 본질을 토대로 추론의 두 단계를 학습(learning)과 통찰(혹은 예지[15], sagacity)

15) 아리스토텔레스에 따르면 실천적 지혜(phronesis)는 phronein에서 유래한 것인데, 이는 'have understanding', 'be wise', 'think', 'mean', 'feel by experience', 'be well aware of' 등을 의미한다. 또한 실천적 지혜에는 bouleusis(deliberation), deinotes(cleverness), aithesis(perception), synesis(sagacity of understanding), gnome(judgment), architektonike(architecture)과 같은 능력이 포함되는데, 이때, synesis가 곧 sagacity를 지칭한다. synesis란 이해력, 양찰, 분별지 등으로 번역될 수 있는데, 이는 구체적이고 개연적인 현실 상황을 목적 실현의 관점에서 총체적으로 직관하는 능력으로, 구체적 상황에서 문제되는 것을 분석하고 찾아내어 즉시 착수할 수 있는 능력을 의미한다. 또한 목적을 위한 수단을 최종적으로 점검하는 역할을 한다(박전규, 1995). 한편, 베이컨은 경험 세계 내에서 우선 접하는 개별자로부터 새로운 개별자나 알려져 있지 않은 추상적 보편성을 발견할 수 있

로 구분하여 접근한다. 학습이란 정보가 가지고 있는 본질 M이 수반하는 것과 암시하는 것들이 무엇인지를 즉시 생각해내는 능력이다. 통찰은 자신에게 주어진 정보 S를 보고, 그 정보의 본질 M을 발견해내는 능력이며, 제임스는 학습과 통찰의 개념을 바탕으로 추론의 과정을 다음의 삼단논법으로 정식화한다.

> *M은 P이다. (M이 수반하는 것은 P이다.)*
> *그리고 S는 M이다. (S로부터 M을 발견한다.)*
> *따라서 S는 P이다. (S로부터 우리가 알 수 있는 새로운 사실은 P이다).*

이 삼단 논법에서 대전제(M은 P이다)가 학습의 단계이고, 소전제(S는 M이다)는 통찰의 단계임을 알 수 있다. 추론의 과정은 좀 더 용이한 학습으로부터 보다 어려운 통찰의 과정을 통해 새로운 정보를 이해하는 방향으로 나아가는 과정이다. 이때, 추론 과정의 기본 형태는 동기화된 추론의 존재 가능성을 내포한다.

> *S에 대해 단순히 M이라고 생각할 때, 우리는 S가 가지고 있을 수 있는 다른 모든 속성들을 무시하고 이 한 가지 속성에만 주의를 쏟는다. (중략) 그 사물에 대해 그런 식으로 생각하는 사람은 그 사물의 나머지 속성을 다 무시하게 될 것이다. (중략) 자연의 요소들이 존재하는 그 사물의 시시각각 변화할 것이고, 우리의 관심도 마찬가지로 변할 것이다. 이런 식이라면, 지금 이 순간의 편파성이 다음 순간의 다른 종류의 편파성에 의해 부분적으로 상쇄될 것이다. (중략) 이처럼 인간들은 편파적인 성향이 아주 깊다(James, 1968/2014 : 468-470).*

새로운 정보를 논리적으로 다루는 추론은 항상 주관적인 관심, 특히 특정한 결론을 얻거나 특별한 호기심을 만족시키기 위한 것이다. 따라서

는 능력으로 창의력(ingenuity)과 명민함(sagacity)이라는 특별한 재능을 제시하며 탐구자가 그 재능을 이용해 사물들의 숨겨진 속성을 발견함으로써 획기적으로 새로운 것을 발견하거나 발명할 수 있도록 이끈다(김시형, 2021: 100-103). 이러한 맥락에서 제임스는 인간의 고유한 이성의 사용 방식으로서 추론의 과정은 다른 보편적인 연합적 사고와 구분되는 본질적인 특징으로서 '새로운 정보'를 해석하는 것으로 보았고, 그 과정 중 하나로 통찰(sagacity)을 제시하였다.

추론은 우리 앞에 놓인 새로운 정보들을 분석할 뿐만 아니라 추상적으로 종합하기도 한다. 이때, 좋은 결론에 도달하기 위해서는 정보를 정확하게 이해해야 하는데, 우리가 가지고 있는 기존의 지식이나 관심과 호기심은 새로운 정보를 이해하는 인지적 틀을 제공하게 된다.

이처럼 어떠한 결론에 도달하기 이전에, 판단과 의사결정의 주체는 그러한 결론을 내리고자 하는 동기를 지닌다. 이때, 동기는 인간이 한 대상에 가질 수 있는 신념이나 태도를 형성하는 데 영향을 미칠 수 있으며, 판단과 의사결정에서 나타나는 추론의 과정에 큰 영향을 미친다. 동기는 행동을 활성화하고, 지시하고, 유지하도록 하는 내적 과정이다(Reeve, 2015/2018: 11). 이러한 동기는 외적 사건이나 사회적 맥락과 같은 선행 요인 조건에 의해 추동되며, 욕구, 인지, 정서 등으로 구체화 된다. 판단 및 의사결정에 관한 기존 심리학 연구들은 인지와 정서적 동기를 독립적인 영역으로 다루었으나, 동기가 인지 과정에 강력한 영향을 미치는 본질적인 요인으로 대두되면서 동기화된 추론 개념이 주목받게 되었다. 추론(reasoning)이 '동기화되었다(motivated)'는 것은, 이와 같은 동기 요인들에 의해 추론 과정이 활성화되고 목표 지향적이며 지속적인 상태에 도달했다는 의미이다.

동기화된 추론의 개념은 사회 심리학자 지바 쿤다(Ziva Kunda)의 연구에서 비롯되었다(Kunda, 1987; 1990; Klein & Kunda 1992). 쿤다는 페스팅거 이후 활발하게 이루어진 많은 사회 심리학 연구들을 종합하여 목표나 동기[16]가 추론에 영향을 미친다는 사실을 밝히기 위해 동기화된 추론의 기초가 되는 심리적 메커니즘을 연구했다. 그는 정서적 동기부여가 우리의 추론을 주도할 수 있다는 개념이 페스팅거의 인지부조화 이론에서 기인했지만, 인지부조화 연구는 판단을 생성하고, 평가하는 과정과 요인을 구체적으로 연구하지 못했다고 지적한다(Kunda, 1987: 637). 이와 달리 동기화된 추론 연구는 우리가 신념을 구축하고 접근하여 평가하는

16) 동기는 판단의 주체가 도달하기를 바라는 최종적인 상태로, 목표(goal)의 의미를 포함한다. 이에 일반적으로 동기화된 추론 연구에서는 주로 동기의 구체적인 내용을 지칭할 때는 목표라는 용어로 사용한다(Kunda, 1990; Leeper & Slothuus, 2014; Taber & Lodge, 2006).

과정에서, 추론이 동기에 의해 영향을 받을 수 있는 가능성을 탐구한다. 동기화된 추론은 정보를 처리하여 결론을 생성하고, 평가하는 추론의 과정이 그 사람의 목표와 욕망에 어떻게 영향을 받을 수 있는지를 설명하기 위해 고안된 개념이다. 따라서 동기화된 추론 그 자체가 단순히 편향된 추론을 지칭하는 것은 아니며(Leeper & Slothuus, 2014), 그 동기 즉, 추론의 목표가 무엇인지에 따라 다른 형태로 나타날 수 있다[17].

쿤다의 연구에 앞서, 크루글란스키 연구진은 특정한 신념이나 사전 지식에 의해 추론이 동기화된다고 보았다. 즉, 자신이 가지고 있는 기존의 신념이나 지식의 구조가 정보를 편향적으로 수용하도록 하는 동기로 작동하며, 이것이 추론에 영향을 미친다는 것이다(Kruglanski, 1980; Kruglanski & Klar, 1987). 또한 자존감(self-esteem)을 유지하고자 하는 동기가 자기본위적(self-serving)인 추론에 영향을 미친다는 것을 발견한 학자들도 있다(Pyszczynski & Greenberg, 1987). 이들의 연구에 따르면 추론의 전체 과정 -분석과 추상을 거쳐 가설을 세우고, 이를 평가하고 추론 규칙을 적용하고 증거를 수용하는 과정- 전반에 걸쳐 이러한 이기적인 동기가 영향을 미친다. 이와 같은 연구들에서 동기화된 추론이라는 개념을 직접적으로 사용하지 않았지만[18], 우리의 추론이 감정에 의해 강하게 동기화된다는 측면과 그 요인이 주로 기존의 신념과 자존감을 유지하기 위한 것임을 밝혀냈다.

동기화된 추론의 개념을 최초로 제시한 연구(Kunda, 1987; 1990)에서 쿤다는 사람들이 정확한 결론에 도달하고자 하는 동기에도 영향을 받지만, 주로 자신이 원하는 결론에 도달하기 위해 인지적 과정과 표상에 의존한다는 사실에 주목했다. 특히, 그는 추론이 긍정적인 자아상을 유지하고자 하는 목표에 의해 동기 부여되며 나아가 이러한 동기가 인과적 판단을 생성하는 과정에 영향을 주어 우리의 추론이 편향된 판단을 도출한

17) 이에 추론의 주체가 도달하고자 하는 목표(동기)의 종류에 따른 동기화된 추론의 유형은 본 절 다음 항에서 상세히 검토된다.

18) 가령 '무효에 대한 두려움(a fear of invalidity)'과 같은 감정 상태는 추론에 영향을 미치는 주요한 동기이며(Kruglanski & Klar, 1987), 이는 다음 항에서 논의될 쿤다의 동기화된 추론의 유형 중 '정확성 목표' 에 의한 동기화된 추론과 상응한다.

다고 주장한다. 자기에 대한 긍정적인 이미지를 보호하기 위해 사람들은 자신의 행동에 대해 이기적인 설명을 생성하고 부정적인 행동을 외부 원인에 귀인 할 가능성이 더 높기 때문이다.

쿤다는 이러한 이기적인 편향이 새로운 주장을 생성할 때뿐 아니라, 새로운 정보를 해석하고 평가하는 과정에도 영향을 준다는 사실을 확인했다. 즉, 자신의 정체성을 긍정하고자 하는 결론을 위협하고 인지부조화를 유발하는 부정적인 원인에 대해 이기적인 방식으로 해석하고 평가함으로써 자신을 보호하고, 타인과의 관계를 유지한다는 것이다(Kunda, 1987: 636). 이러한 자기본위적 증거 평가는 각자 신념이나 기준이 다른 경우에 서로 다른 추론 규칙을 적용할 수 있기 때문에 가능하다(Kunda & Nisbett, 1986; Kunda, 1987). 이기적인 방식으로 자신의 신념과 판단을 생성하고, 이것을 증거의 평가에 선택적으로 적용하기 때문에 사람들은 자신을 편향되게 이해한다. 그리고 이 과정은 단지 강한 정서적 반응으로만 해석되거나, 순수한 추론의 작용만으로 이해되기 어렵다. 따라서 쿤다는 추론의 심리적 기제를 이해하기 위해 동기부여의 요인과 인지적 요인을 모두 고려해야 함을 강조했다. 이 관점에서, 인지적 요인은 동기부여의 목적을 지지하는데 활용된다. 사람들은 그들이 원하는 결론에 도달하기 위해 인지 추론 메커니즘과 과정을 사용하지만, 동기 부여의 힘은 주어진 상황에서 어떤 과정이 사용될 것이고 어떤 증거가 고려될 것인지를 결정한다(Kunda, 1987: 637). 따라서 증거에 기반한 논증 및 판단은 합리적인 것으로 보일 수 있지만, 증거 자체가 특정 동기에 의해 해석될 여지가 있기 때문에 그것의 객관성에 대해서는 재론의 여지가 있다.

이후 연구[19]에서 쿤다는 동기화된 추론이 발생하는 심리적인 기제를

19) 동기화된 추론 개념을 제기한 최초 연구를 제외하고, 이후의 모든 연구에서 쿤다는 동기화된 추론이라는 표현을 motivated inference에서 motivated reasoning으로 수정한다(Kunda, 1990). 그에 대해 쿤다는 명시적으로 언급하지 않지만, 이전 연구에 비해 동기화된 추론이 다양한 영역에서 발견되며, 이것이 보다 '의식적' 차원에서 체계적으로 이루어진다는 점에 초점을 둔다는 점에 그 의도를 미루어 짐작해볼 수 있다. 동료 사회 심리학자인 폴 사가드(Paul Thagard)는 최근 한 인터뷰에서 동기화된 추론의 양상이 무의식적 차원 뿐 아니라, 언어를 명시적으로 사용하는 의식적 차원에서도 활발히 나타난다는 점을 드러내기 위해 그녀가

모델링했다(Kunda, 1990). 이 모델에 따르면, 원하는 결론에 도달하려는 동기는 인지 과정의 작동을 위한 초기의 방아쇠(trigger) 역할을 한다. 즉, 인간의 추론 과정은 의식적으로든 무의식적으로든 정서적인 동기부여에 의해 촉발된다는 것이다. 이 경우, 동기는 접근하려는 지식 및 정보와 사용되는 인지 과정에 영향을 주어 추론에 영향을 준다. 이때 동기는 정서적 요소(즉, 정서와 욕구가 정보 해석에 미치는 영향), 이기적 편향(즉, 자신의 이익에 유리한 방식으로 정보를 보는 경향), 자존감(즉, 사람들이 자신의 자존감을 보호하기 위해 동기 부여된 추론을 사용하는 경향)으로 구체화 될 수 있다.

동기가 신념의 형성과 발달에 영향을 미친다는 사실은, 우리의 추론이 순수 이성적 작용이 아니라 때로는 강한 정서적 영향에 좌우됨을 시사한다(Klein & Kunda, 1992: 146-147). 동기화된 추론의 개념을 제시한 쿤다의 연구는 발행 직후 해당 학계에서 주목을 받았고, 약 30년 동안 12000회 이상 피인용되었다. 이후 동기화된 추론에 관한 연구는 인간의 판단과 의사결정과 관련된 다양한 영역에서 광범위하게 이루어졌다. 로지와 테이버(Lodge & Taber, 2000)는 쿤다의 연구에서 나아가, 동기화된 추론의 구체적인 심리적 메커니즘을 규명하고자 하였다. 이들의 연구는 동기화된 추론과 관련된 인지와 정서의 결합 과정을 세 가지 요소로 구분하여 동기화된 추론의 삼자 이론(the tripartite theory of motivated reasoning)을 정립하였다. 동기화된 추론 과정의 각 요소는 '뜨거운 인지

표현을 바꾸었다고 추측한다(Thargard, P., 'Motivated reasoning: a brief history' at Psychology Today, May 25, 2022). 추론(reasoning)을 언급할 때 보통 의도적이고, 느리고, 의식적이고, 언어적이고, 정서적인 영향에서 자유로운 것을 나타내기 때문이다. 따라서 쿤다가 수정하여 사용한 동기화된 추론(motivated reasoning)이라는 용어는 강한 정서적 영향인 '동기화'라는 표현을 정서의 영향에서 벗어난 것처럼 보이는 '추론'을 구체화해줌으로써 역설적으로 보이는 우리의 인지 과정을 포착하기 위한 것이라고 해석할 수 있다. 쿤다의 연구 등장 이후, 동기화된 추론을 다루는 대부분의 연구자들 또한 쿤다의 개념 정의에 따라 motivated reasoning라는 표현을 사용하고 있다. 참고로, 사가드는 동기화된 추론이 비언어적, 무의식 차원에서 활성화된다는 점을 강조하기 위해서 의식과 무의식적 차원을 모두 포괄할 수 있는 의미를 담기 위해 motivated inference 표현이 더 적절하다고 지적한다. 동기화된 추론이 이중 처리 이론의 관점에서 양분되기 어렵고, 두 인지 과정 모두에서 나타난다는 점에 미루어 볼 때, 사가드의 지적 또한 일면 타당하다.

(hot cognition)', '정보처리의 실시간 처리(on-line processing)', '감정 휴리스틱(How-do-I-feel heuristic)'으로 구성된다.

첫 번째 요소는 모든 사회적 평가와 관련된 정보가 정서의 영향을 받는다는 뜨거운 인지 가설이다(Brand, 1985). 뜨거운 인지는 사람이 환경 요인에 반응하여, 정서적 개입과 분리되지 않는 인지의 과정을 의미한다. 이는 자동적이고, 빠르고, 정서적 반응에 의해 유발되기 때문에 편향된 의사결정을 유발할 가능성을 내포한다. 뜨거운 인지 가설은 주로 정치적 판단 및 도덕 판단과 같은 정서적 반응과 밀접한 관련이 있는 판단의 영역을 설명하는 개념으로 제안된다. 나아가 하나의 정보와 결합된 정서적 반응은 그 개념과 함께 장기기억에 직접 저장되며(Lodge & Taber, 2000: 196), 따라서 유사한 사건이나 정보가 출현할 때 해당 정서적 반응이 즉각적으로 유발된다.

두 번째 요소는 이러한 정서와 연결된 인지 노드의 자동 활성화 과정으로 정보처리의 실시간 모델 가설이다[20]. 정보 처리의 실시간 모델에 따르면, 사람들은 그들이 정치적 대상에 대한 새로운 정보를 접할 때마다, 그들의 태도를 실시간으로 갱신한다. 즉 개인은 사안에 대한 새로운 정보를 접할 때, 장기기억의 형태로 저장되어 있던 관련된 자신의 정서를 의식적인 작업 기억으로 가져와 새로운 정보를 해석하고 이를 최신화한 다음 다시 기억으로 저장한다는 것이다(Lodge & Taber, 2000: 208). 즉, 뜨거운 인지의 처리 과정을 통해 형성된 정서는 실시간 처리 가설을 통해 새로운 정보를 해석하고 수정하여 수용하는 것이다. 이러한 점에서 동기화된 추론의 과정은 사전 확률과 추가적인 정보를 통해 해당 대상의 사후 확률을 유추해내는 베이지안(Bayesian)식 처리 과정과도 유사하다.

세 번째 요소는 감정 휴리스틱[21]이다. 감정 휴리스틱은 감정이 판단에

---

20) 정보처리의 실시간 모델은 정치적 판단과 관련한 심리적 정보 처리 이론에서 논의 되어왔다. 정치 판단 관련 연구에서 정보 처리 모델은 크게 '기억 처리 가설(memory based processing)'과 '실시간 처리 가설(on-line processing)'로 구분된다. 이전에는 이 두 가설은 경쟁적인 모델로 이해되어 왔으나, 최근에는 동기화된 추론의 개념적 틀 내에서 통합적으로 이해되고 있는 추세이다(Sung-youn Kim, 2009; Lodge & Taber, 2000).

21) 원어로는 How-do-I-feel (about it) heuristic이며, 국내 선행 연구에서 이를 그대로 번역하여 '내가 그것에 대해 어떻게 느끼는지'라는 휴리스틱으로 제시하기도

영향을 미치는 데에 있어 특정 정보가 개입할 수 있다는 것이다(Lodge & Taber, 2000: 210). 즉, 정보의 추가로 인해 처음에 가지고 있던 감정과 다른 감정이 새롭게 나타나면 의사결정이 변화한다. 기존의 정보와 관련된 감정은 장기기억의 형태로 저장되어 있고(뜨거운 인지 가설), 이는 새로운 정보를 처리하기 위해 작업기억의 형태로 표현되며(실시간 정보 처리 가설), 이때 새로운 정보로 인해 발생한 새로운 감정은 정보에 대한 기존의 감정과 함께 작용하여(감정적 휴리스틱) 의사결정에 영향을 미친다는 것이다.

로지와 테이버의 동기화된 추론의 삼자 이론은 인지와 정서가 통합되는 메커니즘, 즉 추론이 어떻게 동기 부여되는지에 대한 심리적 메커니즘을 설명해준다. 뜨거운 인지는 모든 정치적, 도덕적 판단의 대상은 정서와 결합 되어 있음을 시사하고, 정보의 실시간 처리 모델은 이러한 정서적 수치를 최신화하기 위한 메커니즘을 제공한다. 나아가 감정 휴리스틱은 자동적으로 이러한 감정 요인이 판단에 영향을 미치는 과정을 설명해준다. 즉, 이 세 하위 이론을 통해 장기기억의 정서와 실시간으로 처리되는 인지 요소가 함께 작용하고 판단 과정에서 이들이 동시에 나타나 동기화된 추론을 형성하는 것이다(Lodge & Taber, 2000: 186).

로지와 테이버는 이 세 가지 하위 과정이 선행연구에서 밝혀낸 동기적 요인들 즉, 자존감을 유지하려는 욕구, 다른 사람과 사회적으로 연결되어 있다고 느끼고 싶은 욕구, 특정 결과를 얻고자 하는 동기 등 다양한 요인에 의해 동기 부여된다고 제안한다. 이 개념 틀은 주로 도덕과 정치의 영역에서 동기화된 추론이 발생하는 심리적 기제를 설명하며 그 근본적인 동기에 대한 포괄적인 이해를 제공한다.

2. 동기화된 추론의 유형

동기화된 추론 개념을 정립한 쿤다는 기존의 선행연구를 토대로 추론

한다(양정애, 2009). 본고에서는 휴리스틱의 여러 하위 유형 중 감정적 반응을 기반으로 어림짐작을 한다는 점에서 '감정 휴리스틱'으로 번역하여 사용하고자 한다.

을 동기화하는 두 가지 목표를 제시한다(Kunda, 1987; 1990). 하나는 '정확한 결론'에 도달하고자 하는 정확성(accuracy) 목표(Freund et al., 1985; Payne et al., 1988; Tetlock, 1983)이며, 다른 하나는 정확성과 상관없이 '특정한 결론'에 도달하고자 하는 방향성(directional) 목표이다(Kruglanski, 1980; Pyszczynski & Greenberg, 1987). 정확성 목표와 방향성 목표로 구분되는 각각의 동기화된 추론의 유형은 각기 다른 인지 전략을 사용하며, 따라서 이 구분은 유의미하다. 정확성 목표의 경우 추론의 과정에서 가장 적절하다고 여겨지는 신념과 전략의 사용을 유도하는 반면, 방향성 목표는 원하는 결론을 도출할 가능성이 높다고 여겨지는 신념과 전략을 사용하도록 동기부여 한다. 특히, 쿤다는 일반적으로 동기화된 추론의 양상은 후자의 모습을 보이기 때문에 편향된 판단을 유발한다고 주장한다. 쿤다가 분류한 정확성과 방향성이라는 두 가지 목표는 이후 연구자들에 따라 [표 2-1]과 같이 다양한 용어를 사용하기도 하지만 그 의미는 쿤다의 구분과 일맥상통한다.

[표 2-1] 목표에 따른 동기화된 추론의 두 유형

	정확성 목표	방향성 목표
Kunda(1990)	accuracy	directional(ity)
Molden & Higgins(2005)	non-directional	directional
Taber & Lodge(2006)	accuracy	partisan
Hart et al.(2009)	accuracy	defensive

1) 정확성 목표

정확성(accuracy) 목표에 의해 동기화된 추론에 관한 연구는 사람들이 최적의(optimal)의 결론에 도달하고자 할 때, 해당 이슈와 관련된 추론에 더 많은 인지적 노력을 투여하고, 더 세심하게 관련 정보에 주의를 기울이며, 더 복잡한 추론의 규칙과 기술을 사용하여 더 깊이 있게 정보를 처리하는 모습을 시사한다(Kunda, 1990: 481). 정확성 목표는 의사 결정자들이 좋은 대안을 최대한 많이 탐색하도록 동기유발하고, 그 수준에 맞는 대안을 찾을 때 비로소 탐색을 종료하고 의사결정을 내리게 한다.

이 과정에서 우리의 추론 과정은 더 많은 인지적 자원을 투입하게 되고, 그 결과 더 복잡하고 시간이 많이 소요되는 의사결정 전략을 선택하게 된다. 이러한 의사결정 전략의 핵심은 특정한 결론을 더 매력적으로 평가하거나, 잘못된 결론을 도출하는 데 위험을 증가시키는 요인들을 제거함으로써 인지 편향을 최소화해준다.

정확성 목표에 의해 동기화된 추론 연구에서 실험 참가자들의 경우 정보나 지식을 수용함에 있어 인지 편향이 거의 나타나지 않았다(Freund et al., 1985; Payne et al., 1988; Tetlock, 1983). 이들은 첫인상을 형성함에 있어 정서적 편향을 야기하는 초두효과(Primacy Effect)를 덜 보였고(Freund et al., 1985: 482), 고정관념을 사용하여 이슈를 평가하는 정도가 상대적으로 낮았다(Payne et al., 1988: 549). 또한 정치 및 법적 판단에서 동기화된 추론의 효과를 분석한 연구에서는 실험 참가자들이 모의 살인 재판의 유죄 판결에서 초두효과를 보였던 이전의 연구와 달리, 실험 참가자들에게 자신의 판단에 대한 공개적 정당화를 통해 책임 있는 판단을 요구한 경우보다 객관적으로 판단을 내렸다(Tetlock, 1983: 288). 정확성에 의해 동기화된 추론을 보인 실험 참가자들은 자신의 신념을 다른 사람들에게 정당화하고 설득하는 과정에서 다른 실험 참가자들에 비해 기본적 귀인 오류[22]에 덜 취약했으며, 이들은 더 깊고 신중한 정보처리 과정을 통해 결론에 도달했다. 이와 같은 간접적인 증거 이외에도 정확성 목표가 더 복잡하고 정교한 추론을 이끄는 직접적인 증거도 있다(Tetlock & Kim, 1987). 정확성에 의해 동기화된 실험 참가자들은 다른 사람에게 자신의 입장을 정당화하고 설득하는 과정에서 인지적으로 더

22) 기본적 귀인 오류란 개인이 자신의 행동의 이유를 평가함에 있어 환경적 요인이나 특수한 외부적 요인과 같은 상황적 귀인(situational attribution) 보다 개인의 성향이나 성격과 같은 내적 요인과 같은 기질적 귀인(dispositional attribution)을 과장하는 경향을 의미한다(Heider, 1958). 상황적 귀인이란 행동의 원인을 사회규범이나 외부 환경 혹은 우연의 일치 등으로 돌리는 경향이다. 즉 한 사람의 행동의 원인을 평가할 때 그것이 상황적인 요인으로 인해 일시적인 것이라고 판단하는 경우를 의미한다. 반면, 기질적 귀인은 행동의 원인은 개인의 성격, 동기, 태도 등에서 찾는 것으로 한 사람의 행동을 이해하고 평가함에 있어 특정한 방식으로 느끼는 개인의 기질적 특성에 초점을 맞추는 경우를 의미한다. 정확성 목표에 의해 동기화된 추론은 정보를 처리할 때, 극단적인 기질적 귀인을 덜 만들어냈으며, 기본적 귀인 오차의 크기를 효과적으로 줄였다(Pittman & Dagostino, 1985).

복잡한 설명 방식을 사용했다. 이들은 더 많은 대안을 고려했고, 논증을 구성하고 이를 뒷받침하는 객관적인 정보를 최대한 많이 수집했으며, 다양한 관점에서 증거들을 검토하였으며, 상대를 설득하기 위한 토론의 기술을 사용하기도 했다.

정확성에 의해 동기화된 추론을 보이는 사람들의 공통된 특징은 특정 결론을 다른 결론보다 더 선호할 유인이 없었다는 점이다(Kunda, 1990; 482). 그들의 유일한 목표는 정확하고, 보다 적합한 결론에 도달하는 것에 있었다. 따라서 정확성에 의해 동기화된 추론은 편향적인 전략이나 휴리스틱과 같은 인지적 지름길을 최소한으로 사용하고자 하였으며, 사실적 증거와 논리적 근거에 기반하여 결론을 도출한다.

한편, 쿤다는 정확하고자 하는 목표만으로는 인지 편향을 줄이지는 못한다는 연구 결과(Lord et al., 1984; Tversky & Kahneman, 1973)도 함께 검토한다. 해당 연구들은 정확성 목표에 동기부여 되었는지보다 실험 참가자들이 적절한 추론 전략을 가지고 있고, 이를 다른 전략보다 우월하다고 판단하고, 이를 바탕으로 최적의 결과를 도출하기 위한 인지적 능력을 갖추었는지에 주목한다. 이러한 능력이 결여되어 있는 사람들은 잘못된 추론 절차가 최선이라고 믿고, 이를 바탕으로 성찰하거나 오랜 시간 숙고하더라도 잘못된 절차에 접근할 가능성이 높기 때문이다. 또한 이들은 자신의 잘못된 결론이 다른 사람을 설득하고, 정당화할 수 있는 것으로 기대하는 경향이 있다. 따라서 이러한 능력이 결여된 상태에서 정확성 목표에 의해 동기화된 실험 참가자들은 다른 참가자들에 비해 편향에 취약한 결과가 나오기도 했다. 따라서 정확성 목표에 의한 동기화가 더 복잡한 처리로 이어지기 때문에, 일정한 수준 이상의 인지 능력이나 훈련되지 않는 사람의 경우에는 오히려 잘못된 결론을 내리기 쉬운 것이다. 이는 동기화된 추론이 무의식적 과정뿐 아니라 의식적 과정으로도 처리됨을 시사한다.

2) 방향성 동기

동기화된 추론의 또 다른 유형은 방향성 목표(directional goal)에 의한 것이다. 정확성 목표와 달리 방향성 목표는 추론의 결과로 도출된 판단의 정확성의 여부보다는 추론이 특정 결론이라는 일정한 방향에 의해 동기화되는 경우이다. 정확성 목표는 가장 적절하다고 여겨지는 신념과 전략의 사용을 유도하는 반면, 방향성 목표는 원하는 결론을 도출할 가능성이 가장 높다고 여겨지는 신념과 전략의 사용을 유도한다. 또한 특정한 결론에 도달하기 위해 동기화된 추론가들은 스스로를 '이성적'인 주체로 여기고, 감정에 좌우되지 않는 공정한 관찰자를 설득할 수 있는 정당한 논증을 구성한다고 생각한다(Kunda, 1990: 483). 즉, 추론이 정확성 목표에 따라 유발되지 않는다는 것이 곧 동기화된 추론이 정확성을 전적으로 무시한다는 것을 의미하는 것은 아니다.

동기화된 추론 연구자들은 이와 관련하여 '객관성 환상(illusion of objectivity)'을 지적한다. 객관성 환상이란, 자신을 실제보다 더 객관적이고, 더 공평하고, 더 통찰력 있고, 덜 편향된 것으로 보는 경향을 의미한다(Armor, 1999). 인지부조화의 상황에서 자존감 유지를 위해 상황을 왜곡하는 것과 같이 특정한 결론에 도달하려는 목표에 따라 동기화된 추론을 사용하는 경우, '객관성 환상'을 유지하기 위해 추론에 영향을 미치는 정보를 통제한다(Pyszczynski & Greenberg, 1987: 302). 객관성 환상은 방향성 목표에 동기화된 추론가들 스스로가 논리적으로 타당하다는 생각을 유지하게 해준다. 이들은 자신이 원하는 결론을 뒷받침할 수 있는 믿음과 규칙들을 자신의 기억 속에서 찾고, 때로는 창의적인 방식으로 기존의 신념과 새로운 정보를 결합하는 방식을 취하기도 한다(Greenwald, 1980: 610-612). 이처럼 인지 편향과 인지왜곡의 심리적 기제로 지적되고 있는 동기화된 추론은 주로 좁은 의미로서 방향성 목표에 의해 동기화된 추론을 의미한다[23].

23) 대다수의 동기화된 추론 연구는 방향성 목표의 경우에만 한정하여 동기화된 추론 개념을 사용한다(Ellis, 2022). 앞서 살펴본 것과 같이, 정확하거나 진정한 결론

동기화된 추론 연구의 모태가 되는 인지부조화 현상은 방향성에 의해 동기화된 추론이 나타난다는 강력한 증거이다[24]. 인지부조화 이론을 고안한 페스팅거는 자신이 원하는 결론, 즉 자신의 신념을 지키고자 하는 목표에 의해 동기화된 추론의 편향성을 검증하기 위해 다음의 실험을 설계했다(Festinger & Carlsmith, 1959). 페스팅거와 연구진은 실험 참가자들에게 쟁반에 숟가락을 올렸다 내리는 반복적이고 지루한 작업을 지시했다. 연구자들은 초시계를 들고 실험 참가자들의 모습을 기록하였다. '실험이 끝났다'라는 말과 함께 실제 이 실험은 시작된다. 연구자들은 실험 도우미를 통해 실험 참가자들 사이를 돌아다니면서 '시간 가는 줄 모르는 재미있는 연구였다'고 말하도록 지시했다. 실험 참가자는 크게 세 집단으로 구분된다. 대조군 집단에서는 과제가 끝난 후, 그 과제의 흥미도를 진실되게 평가하도록 했다. 나머지 두 집단은 과제를 평가하기 전에 대기 중인 실험 참가자들에게 이 과제를 추천해달라는 요청을 받았다. 연구진들은 과제에 대한 거짓말을 대가로 한 집단에게는 1달러를, 다른 집단에게는 20달러를 제안했다. 기존의 행동주의적 관점에 따르면 보상을 더 많이 받은 집단은 해당 요청을 더 잘 수행해야 할 것이다. 그러나 결과는 달랐다. 통제 집단의 경우 -5점에서 5점 사이 범위에서 과제의 흥미도를 -0.5점으로 평가했다. 1달러를 받은 집단은 +1.35점을, 20달러를 받은 집단은 -0.05점을 주었다. 1달러를 받은 집단이 20달러를 받은 집단보다 과제가 재밌었다는 거짓말을 한 것이다. 페스팅거는 이러한

에 도달하려는 동기에 의한 추론은 인지 왜곡이나 편향을 유발할 가능성이 매우 낮기 때문이다. 본고에서는 다양한 동기를 고려하여 넓은 의미의 동기화된 추론 개념을 함께 제시하면서, 방향성 목표에 의한 동기화된 추론을 좁은 의미의 동기화된 추론 개념으로 사용하고자 한다.

24) 방향성 목표에 의해 동기화된 추론이 편향적이라는 증거는 1950년대 등장한 인지부조화 이론을 시작으로 심리학, 교육학, 정치학, 경영학을 포함한 인간의 인지과정과 판단 및 의사결정에 주목하는 학문 분야에서 광범위한 주제(도덕적 위선, 정치적 양극화, 인종 차별, 사형제도 존폐, 기후 변화, 원자력 발전의 영향, 육류 소비, 코로나19 관련 가짜뉴스 등)로 연구되고 있다. 이러한 연구는 인지적 왜곡과 편향으로 인해 발생하는 문제의 원인을 동기화된 추론의 기제로 분석하는 데 초점을 맞춘다. 이 논의는 3장에서 자세하게 논의될 것이다. 본 절에서는 방향성 목표에 의한 동기화된 추론이 갖는 인지 편향의 메커니즘을 설명하고 있는 초기 연구들을 중심으로 살펴볼 것이다.

결과를 인지부조화 이론의 틀로 설명한다. 즉, 누군가 자신의 의견과 반대되는 행동을 하거나 말하도록 동기화된다면, 그는 자신의 행동과 자신의 생각을 일치시키기 위해 의견을 바꾸는 경향이 있는 것으로 보인다는 것이다. 사람들은 자신의 태도와 불일치 하는 행동(과제가 재미있었다는 거짓말)을 했음에도, 그렇게 한 행동에 대해 강력한 정당화(20달러)가 되기 때문에 인지부조화가 덜 발생하였다. 반면, 자신의 태도와 불일치 하는 행동에 대한 정당화가 불충분(1달러)한 경우 더 큰 수준의 인지부조화를 느끼게 되고 그로 인한 불쾌감을 줄이기 위해 자신의 의견을 행동에 맞추어 변화시킨 것이다. 이와 같은 유도된 순응(induced compliance) 실험은 여러 다른 실험들에서도 입증되었다(Aronson & Calrsmith, 1962; Masataka & Perlovsky, 2012).

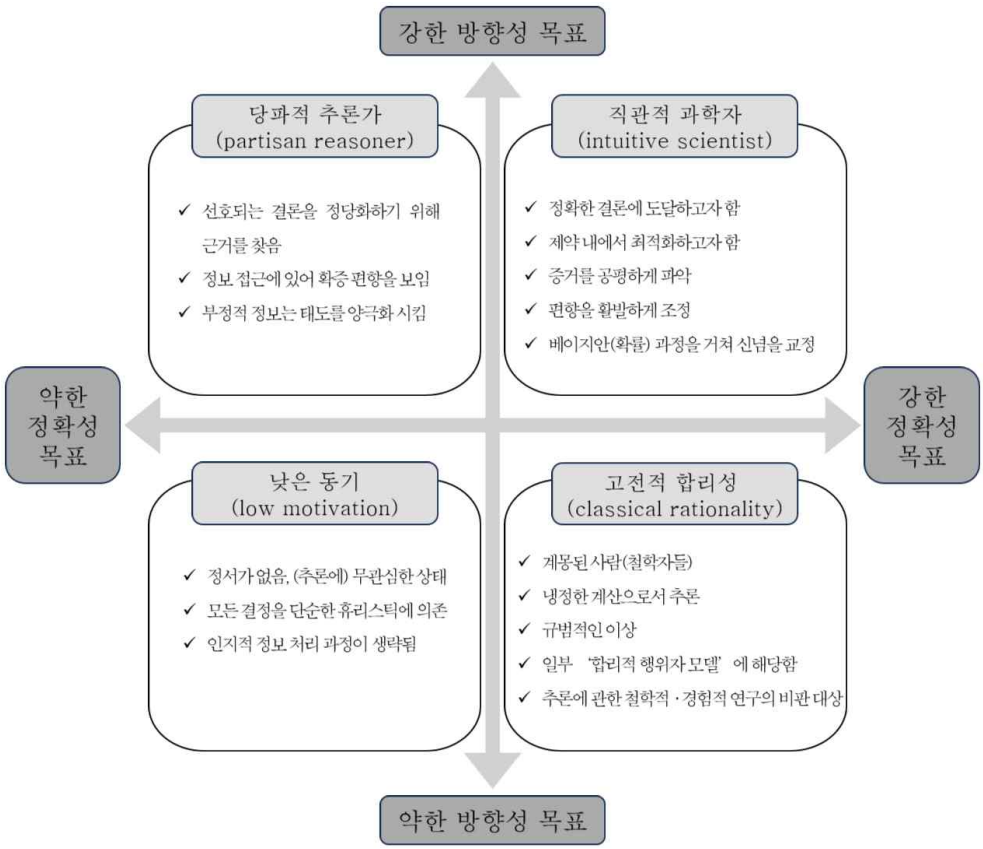
자신의 신념을 지키기 위한 목표라는 일정한 방향을 바탕으로 동기화된 추론은 역효과 현상(backfire effect)의 원인이 되기도 한다. 역효과 현상이란 명백히 잘못된 신념을 유지할 경우, 과학적이고 객관적인 반론에 직면하더라도 오히려 더 집요하게 자신의 틀린 관점을 고집하고자 하는 현상이다(Monney, 2012: 71). 사람들은 자신의 신념과 모순되는 사실적 근거를 부정하고, 때로는 정교한 반론을 구성하기도 한다. 사회 심리학자 무니는 동기화된 추론으로 인해 발생하는 이와 같은 현상이 비전문가나 교육 수준이 낮은 사람보다는 오히려 전문적 과학 지식을 신뢰하며 교육 수준이 높은 사람들에게 더욱 두드러진다는 점에 주목하여 이들을 '똑똑한 바보'라고 지칭하기도 한다. 가령, 낙태 문제에 대해 반감을 가진 사람들이 교육 수준이 낮은 경우에는 이유 없이 그러한 정책을 거부할 것이다. 그러나 교육을 받은 사람들은 그럴듯한 반론을 제시한다. 자신이 내놓은 반론은 더 정교하게 스스로를 설득하게 되고 심지어 처음의 확신보다 더욱 강하게 믿게 되는 것이다. 앞서 간단히 살펴본 동기화된 추론의 신경과학적 근거를 제시한 웨스턴 등의 연구(Westen et al., 2006)는 무니가 주장하는 똑똑한 바보 효과를 지지해준다. 해당 연구는 2004년 미국 대선에서 당파적인 정치적 판단의 과정을 fMRI를 사용하여 확인하였다. 이들은 대선 기간 동안 자신이 지지하는 정당의 입장을 직접적으

로 반박하는 객관적인 정보가 나타났을 때, 정당의 열성 지지자들이 어떠한 추론 양상을 보이는지 연구했다. 이들은 지지자들의 뇌를 스캔하여 동기화된 추론이 이루어지는 동안 활성화되는 뇌의 부위를 발견하였다. 실험 참가자들은 공화당과 민주당 지지자들 각각에게 자신이 지지하는 후보자가 했던 말과 행동이 다른 사례를 제시받았다. 모순적인 근거를 제시하고 나서 연구진들은 실험 참가자들에게 그들의 말과 행동이 서로 일관되지 않은 정도를 평가하도록 요구했다. 자신의 반대 정당 후보자의 말과 행동은 위선으로 보는 경향이 강한 반면, 자신의 후보자의 경우에는 그렇지 않다고 보고했다.

신경과학적 관점에서 동기화된 추론은 뇌가 부정적인 판단을 최소화하고 긍정적인 영향을 최대화하는 암묵적인 정서 조절의 한 형태로 이해된다(Westen et al., 2006: 1947). 동기화된 추론은 복내측 전전두엽피질(vm-PFC), 복외측 전전두피질(VLPFC) 전측 대상회 피질(ACC), 후두엽피질(PCC), 뇌섬엽피질(insular cortex), 안와전두피질(ORC)의 활성화와 관련이 있었다. 웨스턴은 동기화된 추론이 기능적으로 통합된 뇌 영역의 네트워크의 역할에 의해 뒷받침된다고 보았다. 가령, 정서적으로 위협적인 정보를 평가함에 있어 vm-PFC가 활성화되는 것은 인지와 정서가 상호작용하고 있음을 간접적으로 드러낸다. 즉, 교육 수준이 높은 집단은 자신의 신념을 훼손시키는 객관적 근거에 직면할 때 감정적인 영향을 느끼고 이를 회복하기 위해 인지적 자원을 활용하는데, 이 과정에서 근거를 재평가하여 자신의 입장을 정당화하는 것이다. 이때, 인지와 의사소통을 담당하는 VLPFC에서 감정적인 반응을 조절한다. 웨스턴은 감정 조절에 대한 이전의 연구들(Anderson et al., 2004; Hariri et al., 2003; Ochsner et al., 2002, Westen et al., 2006: 1955에서 재인용)을 참고하여 실험 참가자가 부정적인 감정 자극에 대한 반응을 인지적으로 억제했을 때, vm-PFC가 활성화됨을 관찰했으며, 이는 불안 등의 정서적 반응과 관련된 편도체 반응을 감소시킨다. 즉, 우리의 뇌는 자신의 신념을 보호하고자 하는 방향성 목표에 도달하기 위해 동기화된 추론을 사용하여 인지부조화를 해소하는 것이다.

3) 두 동기의 결합

이처럼 동기화된 추론을 결정하는 동기는 크게 정확성과 방향성으로 구분된다. 바우마이스터와 뉴만은 동기화된 추론의 두 유형을 '직관적인 과학자(최적의, 정확한 결과를 목표하는 과학자의 모습)'와 '직관적인 변호사(법률대리인으로서 의뢰인의 이익 보호라는 특정한 방향을 목표하는 변호사의 모습)'로 비유하기도 한다(Baumeister & Newman 1994: 6-14). 또한 로지와 테이버는 정확성과 방향성이라는 두 가지 동기를 결합하여 동기화된 추론을 다음과 같이 네 가지로 범주화하여 접근하였다.



[그림 2-1] 동기화된 추론의 유형(Lodge & Taber, 2000: 187-191 재구성)

우리의 추론은 정확한 결론에 도달하고자 하는 목표와 자신이 원하는 결론에 도달하고자 하는 목표 사이에서 수행된다. 따라서 동기화된 추론은 최적의 정확성과 자신의 신념을 위한 추진력 사이의 긴장 속에서 유형화될 수 있다. 로지와 테이버는 이처럼 서로 다른 정확성과 방향성 정도의 차이가 결합되어 동기화된 추론이 각기 다른 양상으로 나타난다고 확인하였다. 먼저, 정확성과 방향성 측면이 모두 약한 경우는 의사결정 상황에 몰입되지 않은 상태로 정서적 개입이 없고 무관심한 상태이다. 이 경우는 낮은 동기부여의 상태로 결정을 할 때 추론의 과정이 개입되기보다는 단순한 어림짐작에 의존하게 된다. 이처럼 동기부여가 거의 없는 상황에서 사람들은 상대적으로 무관심하며, 정보를 처리하는 데 많은 노력을 기울이지 않는다. 이처럼 인지적 정보 처리 과정이 생략된 '낮은 동기부여' 유형은 특히, 도덕적·정치적 판단과 같이 자신의 신념에 의존하여 추론을 사용하는 경우에는 거의 발생하지 않기 때문에 동기화된 추론 관련 논의에서 제외된다.

직관적 과학자 유형은 앞서 정확성 목표에 의한 동기화된 추론 관련하여 논의되었다(Baumeister & Newman, 1994). 이 유형의 추론은 인지와 정서가 결합된 동기화된 추론의 가장 이상적인 형태이다. 쿤다의 유형 분류는 단지 정확성 목표에 의해서 동기화된 경우를 모두 직관주의적 과학자 유형으로 보았지만, 로지와 테이버는 이 유형은 방향성의 측면에서도 특징지어진다. 즉, 자신의 신념에 기반하여 원하는 목표에도 동기부여되지만, 최적의 목표에 도달하고자 하는 강한 동기부여와 함께 결합되기 때문에 편향되거나 왜곡하여 정보를 인식하지 않는다. 이들은 원하는 결론에 도달하고자 하는 방향성 목표에 의해 동기부여되지만, 원하는 결론이 곧 더 나은 신념의 형성이기 때문에 정확성 목표를 동시에 추구한다. 따라서 정확한 결론에 도달하고자 하므로, 철저한 탐색과 증거에 대한 공정한 평가를 시도함으로써 목표를 향해 노력할 가능성이 가장 높은 유형이다(Lodge & Taber, 2000: 188). 직관적 과학자 유형의 동기화된 추론은 정보 처리의 편향에 대해 능동적으로 조정하고, 때로는 의도적으로

과하게 조정하기도 한다. 즉, 자신의 신념을 객관적으로 검증하기 위해 반대되는 증거를 과장하여 이해해 보기도 하고, 자신의 입장에 기여하는 근거와 반대되는 근거를 양과 질의 측면에서 광범위하게 수집하고자 한다. 이들은 자신이 접근할 수 있는 환경적 조건의 제약 내에서 최적화된 결론을 도달하고자 한다는 점에서 과학자의 탐구 모습으로 은유할 수 있다. 그러나, 과학자들은 최적의 결론에 도달하기 위해 항상 연역적인 방식으로 철저한 논증을 통해 순차적으로 결론에 도달하지 않는다. 이들은 베이지안의 과정을 거치면서 목표에 도달하고자 한다. 이 과정에서 신념은 계속해서 최신화되고 더 나은 결론을 향하게 되는 것이다. 그러나 로지와 테이버는 이러한 유형에 도달하는 사람은 고도의 훈련을 받은 소수의 추론 전문가일 것이라고 예측한다. 기존의 추론에 관한 담론에서 가장 많은 유형으로 나타나는 것은 다음에 살펴볼 고전적 합리성 유형과 당파적 추론가 유형이다.

고전적 합리성 유형은 높은 정확성 목표와 약한 방향성 목표의 결합으로 동기화된 추론이다. 이 유형에서의 추론은 '최적의 정확성'에 대한 관심으로 크게 동기부여 되기 때문에 정보를 냉정하고 완전하게 처리하고자 한다. 또한 특정한 결론을 전제하지 않기 때문에 방향성에 의한 동기부여는 매우 미미하며, 따라서 이러한 추론은 탈맥락적이고, 순수한 논리에만 의존할 가능성이 높다. 이러한 유형은 고대 그리스 아테네 철학의 이성에 대한 신뢰로부터 데카르트의 절대주의적 인식론에 해당하는 것으로 논리학에서 말하는 엄밀한 의미의 추론 개념에 가깝다. 아리스토텔레스 논리학은 추론의 전형적인 사례로 정언적 삼단논법을 제시한다. 이를 통해 타당한 연역적 추론은 체계화될 수 있다. 자명한 진리로부터 철학적 탐구를 시작하고자 했던 데카르트의 방법적 회의 또한 이 유형에 속한다. 이처럼 철저한 연역적 추론과 같은 추론의 사용 방식은 고전적 합리성의 규범적 이상을 의미한다(Lodge & Taber, 2000: 188). 고전적 합리성 유형의 동기화된 추론은 인간이 가진 합리성에 대한 신뢰를 바탕으로 논증을 통해 정확한 결론에 도달하도록 추론을 사용하는 방식이다. 엄격한 철학적 논증을 통해 새로운 결론을 도출해내는 것은 인간 이성의

고유한 사용 방식으로 널리 인정되어 왔다. 그러나 의사결정에 대한 다양한 철학적, 심리학적 연구에서는 고전적 합리성 유형의 추론이 현실에서 매우 제한적으로만 나타나며(Haidt, 2001), 이를 허수아비(straw man)로 비유하기도 한다(Lodge & Taber, 2000). 그러나 고전적 합리성 모델은 인지 과정의 이중 처리 모델의 등장과 함께 인간의 의사결정과 관련된 모든 연구 분야에서 수정되거나 대체되었다.

끝으로, 당파적 추론가 유형은 동기화된 추론의 가장 전형적인 양상이다. 좁은 의미의 동기화된 추론의 개념을 사용하고 있는 대부분의 동기화된 추론 연구는 당파적 추론가 유형을 전제한다. 이 유형은 정확한 목표에 도달하고자 하는 동기보다는, 자신의 신념이나 기존의 입장과 같은 '도달하고자 하는 목표'라는 강한 방향성에 의해 동기화된 추론의 모습이다. 당파적 추론가의 모습은 선호되는 결론을 정당화하기 위한 강한 방향성을 바탕으로 근거를 수집하고 논증을 구성한다. 이 경우에는 정보를 접근함에 있어 강한 확증 편향의 모습을 보인다. 자신의 입장과 정면으로 충돌하거나 모순을 일으키는 새로운 정보나 지식은 무시하거나 왜곡하여 수용하며, 기존의 신념을 지지하는 근거는 무비판적으로 수용하는 방식으로 추론을 사용하기 때문이다.

이러한 유형에 '당파적'이라는 이름이 붙은 이유는 정치적 판단과 관련하여 특히 당파적 방향성은 강력한 동기화된 추론을 촉발하기 때문이다(Lodge & Tabor, 2006). 자신을 보호하기 위한 방향성 목표에 의해 동기화된 추론은 주로 자신의 사회적 정체성과 일치하도록 정보를 해석한다. 특히, 동일한 정책 평가의 과제에서 당파적 정체성을 가진 사람들은 자신이 지지하는 정당이 내세울 때와 자신의 상대 정당이 내세웠을 때 정책의 가치를 다르게 평가한다는 증거는 광범위하게 찾아볼 수 있다. 사람들은 자신의 정치적 신념이나 사회적 정체성에 일치하는 정보를 모순된 정보보다 더 큰 가중치를 부여하여 평가한다. 당파적 추론의 양상은 최근 미디어 컨텐츠의 폭발적인 증가와 함께 이슈가 되고 있는 가짜 뉴스 확산 과정에서도 나타난다. 언론 보도에 포함된 정보의 진실성 여부보다는 자신이 가진 기존의 신념을 기준으로 받아들이기 때문에 가짜

뉴스의 확대 재생산에 동기화된 추론이 기여한다는 것이다. 이와 관련하여 최근 연구는 전세계적으로 창궐한 covid-19 바이러스와 관련된 가짜 뉴스의 확산의 원인으로 동기화된 당파적 추론으로 설명되기도 하였다(Porumbescu et al., 2020). 해당 연구에서는 팬데믹 기간 동안 바이러스 관련 정보를 객관적으로 고려하기보다는 자신의 신념을 보호하기 위한 방향으로 해석하는 경향을 발견했다. 특히, 트럼프 정부와 공화당을 지지하는 사람들은 바이러스 확산의 원인을 중국 정부의 음모와 연결시킨 가짜 뉴스를 공신력 있는 정보로 받아들이는 경향을 보였다[25].

이상의 논의를 통해 동기화된 추론의 유형을 살펴보았다. 동기 요인은 크게 정확성과 방향성 목표로 구분된다. 이는 정확한 결론에 도달하고자 하는 목표와 자신이 원하는 방향으로 결론을 도출해내고자 하는 목표와 동기의 결합으로, 구체적으로 네 가지 유형으로 나누어 살펴볼 수 있다. 이때, 낮은 방향성 목표와 낮은 정확성 목표를 제외한 세 가지 유형은 각각 서로 다른 추론의 양상을 보여주었다. 넓은 의미의 동기화된 추론은 낮은 동기를 제외한 세 가지 유형의 동기화된 추론을 포괄하는 개념으로 동기적 요인에 의해 추동되는 추론 일반을 지칭하는 개념이다. 대부분의 동기화된 추론 관련 연구에서 사용되는 개념은 주로 좁은 의미의 동기화된 추론으로 방향성에 의한 동기화된 추론을 전제하며, 구체적으로는 강한 방향성 목표와 낮은 정확성 목표에 의해 동기화된 추론의 유형인 당파적 추론을 지칭한다. 당파적 추론으로서 동기화된 추론은 자신이 보호해야 할 신념과 정체성 등에 입각한 특정 결론을 지향함으로써 정보와 사실을 왜곡하게 한다. 도덕 판단에서 문제가 되는 동기화된 추론의 개념 또한 이와 같은 당파적 추론의 일종으로 볼 수 있다.

---

25) 한편 이와 관련하여 일부 연구는 이를 당파적 추론의 유형이 아니라 추론의 부족 현상으로 지적하기도 한다(Maguire & Persson, 2022). 추론의 부족은 뉴스 보도나 정보를 무비판적으로 수용하며 따라서 일관적인 패턴을 발견할 수 없는 방식으로 정보를 평가한다. 이는 동기화된 추론의 유형으로 분류했을 때, '낮은 동기부여'의 유형에 해당한다.

제 3 장 도덕 판단에서 동기화된 추론의 문제

제 1 절 동기화된 추론과 도덕 판단의 관계

1. 동기화된 추론에 관한 도덕심리학적 접근

도덕적 추론은 도덕 심리학의 영역에서 활발하게 연구된 주제 중 하나이다. 1940년부터 2017년까지 연구된 1,278건의 연구를 대상으로 도덕 심리학의 거시적 동향을 분석한 결과(Ellemers et al, 2019), 도덕적 추론(moral reasoning), 도덕 판단(moral judgment), 도덕적 행동(moral behavior), 도덕적 정서(moral emotion), 도덕적 자아관(moral self-view)이라는 5대 주제를 중심으로 도덕 심리학 연구가 진행되었으며, 그중 가장 큰 비중을 차지한 주제는 도덕적 추론에 관한 것이었다. 도덕적 추론에 대한 축적된 연구는 두 가지 유형의 결론을 이끌어냈다(Ellemers et al., 2019: 352). 하나는 다양한 사회적 역할과 삶의 경험이 도덕적 원리에 비추어 추론하는 방식에 장기적인 영향을 미칠 수 있다는 점이며, 다른 하나는 즉각적 상황 단서들 또한 도덕적 추론에 영향을 미치며 결과적으로 도덕 판단을 도출하는 데 주요한 변수로 작용한다는 것이다. 이와 같은 결론은 도덕적 추론이 개인의 내적인 능력을 기반으로 하는 심리적 메커니즘에 의해 작동하지만, 그 작동의 과정에서 추론에 영향을 미치는 여러 요인을 고려할 때 온전히 이해할 수 있음을 시사한다. 따라서 도덕 판단 및 도덕적 추론에 대한 도덕 심리학 연구의 거시적 동향은 도덕적 추론을 설명하는 다양한 요인과 그 의미의 복잡성을 포괄적으로 고려하기 위한 방향으로 발전해왔으며 이러한 맥락 속에서 동기화된 추론에 관한 도덕 심리학적 접근 또한 등장했다.

1) 콜버그의 도덕 발달론에서 동기화된 추론의 이해

도덕 심리학자 콜버그(Kohlberg, 1969)는 피아제의 인지발달 연구를 토대로, 정의 추론을 바탕으로 하는 도덕 판단의 발달 과정을 탐구하고자 하였다. 피아제와 콜버그의 연구는 아동을 공정성(fairness)이라는 보편적인 도덕규범의 범주에 관한 질문을 중심으로 의미를 구축해가는 철학자로 가정한다. 콜버그는 '아동이 철학자이다'라는 전제가 도덕 발달 연구의 제 1가정임을 강조하였다.

콜버그는 심리학의 역사 속에서 자신의 도덕 발달 연구가 차지하는 여러 의의 중 하나로 도덕 영역에 대한 심리학적 연구를 확장했다고 자평하였다(Kohlberg et al., 1983/2000: 24). 콜버그와 연구진은 아동의 도덕 발달은 도덕적 감정이나 열망, 행동의 발달이 아니라 정의(justice) 추론의 발달이라는 것을 명확히 한다. 이를 확인하기 위해 콜버그는 서로 다른 입장을 가진 사람들로 구성된 가상적 딜레마를 기반으로 연구를 수행했다. 또한 연구진은 딜레마 이야기를 접한 실험 참가자들에게 공정과 정의의 문제에 초점을 맞추어 질문을 제시했다. 콜버그는 도덕 발달에 대한 자신의 연구를 정의 추론의 발달로 규정하기 위해 철학적으로는 칸트의 도덕철학을 내용으로 삼고, 심리학적으로는 피아제의 인지발달론을 연구 방법으로 삼고 있음을 밝힌다(Kohlberg et al., 1983/2000: 43).

정의 추론은 콜버그 도덕 발달 연구의 핵심 아이디어이다. 정의는 개인 간의 갈등을 해결하는 방법을 제공해주고, 충돌하는 이해관계를 공정하게 고려하는 기준이 되며, 상호호혜성에 근거하여 원리화된 방식으로 분배하는 방법을 제시해준다(Lapsely, 1999/2000: 106). 따라서 도덕 발달의 관점에서 정의 추론은 사람들 사이에서 발생하는 딜레마의 해결책을 제공해주는 개념적 틀이다. 이와 관련하여 콜버그는 가역성이라는 준거가 정의의 궁극적 준거임을 강조한다(Kohlberg et al., 1983/2000: 155-160). 가역성이란 도덕적 조작(operation) 가운데 정의 추론의 구조가 갖는 속성이다. 도덕 판단은 가역성이라는 속성으로 인해 그 판단과 관련된 모든 사람들의 관점을 고려하며 가장 평등하고 호혜적인 방식으로 딜레마의 해결책을 구성할 수 있는 것이다. 3수준 6단계의 계열성에

서 높은 단계의 정의 추론 구조는 낮은 단계의 정의 추론 구조에 비해 가역성이 더 높다. 가령, 관계적 조화성 단계(3단계)의 정의 추론 구조는 개인주의 및 도구주의 단계(2단계)의 정의 추론에 비해 더 많은 당사자의 관점을 고려하여 도덕 판단을 도출해낸다. 2단계에서 상호성은 호의와 보복의 교환을 의미한다. 따라서 2단계의 도덕 추론자는 '자신이 받은 만큼 되받아 때려야 한다'는 판단을 내릴 것이다. 그러나 3단계의 도덕 추론자는 그러한 끝없는 보복의 악순환이 가역성에 어긋난다는 것을 이해하며 한층 더 넓은 의미의 가역적인 상호성 개념에 근거하는 것이다 (Kohlberg et al., 1983/2000: 164-166). 콜버그는 가장 높은 단계인 보편적 도덕 원리 지향의 도덕성(6단계)의 정의 추론 구조를 설명함에 있어 롤스의 무지의 베일 속 원초적 입장 개념과 하버마스의 논증대화윤리에서 전제하는 대화의 상호작용 개념을 다음과 같이 설명한다.

> *우리는 가역적인 도덕 판단에 이를 수 있는 과정을 '이상적 역할채택'이라고 부른다. 6단계 도덕 판단은 모든 행위자들이 황금률에 지배·조절된다는 가정 아래에서 각각의 행위자의 주장을 역할채택하는 것에 기초를 두고 있다. 또한 이상적인 역할채택이란 인간존중과 공평성으로서의 정의의 자세를 필요로 하는 궁극적인 의사결정 절차이다. 관련된 모든 사람들이 도덕적 태도 또는 정의의 개념에 지배를 받거나 그들이 기꺼이 다른 사람의 역할을 채택하려는 한, 이러한 방식으로 얻어진 의사결정은 그들의 관점에서 볼 때, '옳다'는 의미에서 '평형화' 상태에 있다(Kohlberg, 1973: 643, Kohlberg et al., 2000: 148에서 재인용).*

즉, 6단계의 도덕 추론자는 딜레마 안에서 직접적인 이해관계를 갖는 사람들의 관점을 고려할 뿐만 아니라, 자신의 선택으로 영향을 받을 수 있는 간접적인 이해관계자까지 고려하여 최종적인 판단을 도출해낼 것이라고 예측할 수 있다. 이처럼 가역적인 도덕 판단에 이르는 과정은 곧 이상적인 역할채택의 과정인 것이다.

정의 추론으로서 도덕적 추론을 전제한 콜버그의 도덕 발달 연구에서 동기화된 추론의 영향은 어떻게 해석할 수 있을까? 콜버그는 동기화된 추론이라는 개념은 명시적으로 언급하지 않았으나[26], 추론에 영향을 미

치는 감정 변수에 대한 그의 견해를 바탕으로 도덕 판단에서 동기화된 추론의 문제에 대한 입장을 파악해볼 수 있다.

콜버그는 인지적 성숙과 도덕적 추론의 비대칭적인 상관관계를 보여준 연구 결과에 주목한다. 해당 연구에서 단계 5, 6의 도덕적 추론의 구조를 보여준 모든 청년과 성인들은 고차원적인 형식적 추론 문제 또한 해결했지만, 반대로 형식적 추론 문제를 해결했던 다수의 청년과 성인들은 단계 5, 6의 도덕적 추론을 보여주지 못했다. 그는 이 결과를 두고 우리가 높은 수준의 도덕적 추론을 하기 위해서는 반드시 인지적으로 성숙해야 하지만, 아무리 영리하더라도 도덕적으로 추론하지 못할 수 있다는 결론을 내렸다. 콜버그는 이러한 비대칭성의 원인이 도덕 판단에 개입하는 동기(motives) 요인에 있음을 지적한다.

> *인지 수준과 도덕 판단 수준 사이에 가로놓인 이러한 비대칭에 대한 한 가지 이유는, 인지 능력이 의지나 욕망과 같은 요인의 방해를 받아 도덕 판단에서 현실화되지 않기 때문이다. 인지 영역에서 자신이 할 수 있는 최상의 수준에서 추론할 수 있는 까닭은 명백히 그 추론이 자신의 이해관계에 속해 있기 때문일 것이다. 도덕 영역에서 이 점은 그렇게 분명하지 않다. (중략) 도덕 영역에는 의지, 욕망, 정서 등이 개입하기 때문이다(Kohlberg, 1981: 138-139).*

우리는 올바른 높은 수준의 정의 추론을 할 수 있는 능력이 있음에도 불구하고, 도덕적 추론을 산출하는 과정에서 의지나 욕망과 같은 동기적 요인으로 인해 현실에서 발현되지 못한다. 이 지점에서 동기화된 추론에 관한 콜버그의 입장을 유추할 수 있다. 앞서 살펴본 것처럼, 콜버그는 진정한 의미의 도덕 발달을 심리학적으로 이해하기 위해서는 추론의 구조에 주목해야 함을 강조한다. 그의 이러한 핵심적인 전제는 이후 여러 경험적, 실증적 연구들에 의해 수많은 비판의 대상이 되었다. 콜버그를 비판적으로 계승한 여러 연구자들은 도덕 판단에서 정서와 동기적 요인에 주목한다(Rest, 1999; Gibbs, 1991; Lind, 2016). 레스트와 그의 동료들이

---

26) 동기화된 추론(motivated reasoning) 연구는 콜버그가 생을 마감하던 해(1989)에 등장했기 때문에 시기적으로 그가 이 개념을 사용하기는 어렵지만, 그의 연구를 통해 추론에 영향을 주는 동기 요인에 대해 고려하고 있음을 확인할 수 있다.

제안한 도덕적 행동을 산출하는 4구성 요소 모형이나, 공감(empathy)과 정의(justice)가 도덕적 동기화의 기본적인 두 원천을 밝히고, 도덕 판단에 있어 자기본위적인 인지 왜곡으로 발달의 실패 요인을 강조한 깁스의 연구, 도덕적 정서의 측면과 인지의 측면을 도덕적 지향과 역량으로 구분하여 이를 동시에 측정하는 도구를 개발한 린트의 연구가 대표적이다. 이들은 공통적으로 콜버그의 연구가 간과한 정서적 요인에 주목하고 보완하고자 하였다. 그러나 콜버그는 도덕 판단에 있어 정서적 요인을 결코 무시하지 않았다. 그는 도덕 판단을 끌어내는 인지와 도덕적 정서는 하나의 도덕적 반응에 대한 두 가지 측면이며, 도덕 판단에 영향을 주는 정서는 도덕적 추론을 통해 표현된다고 보았다.

> *뒤르켐이나 프로이트와 같이 도덕 발달에 있어 정서에 관한 비합리적 이론과 달리 인지발달론적 관점에 따르면, 인지와 정서는 동일한 정신적 사건(mental events)의 서로 다른 두 측면(aspect)이다. 도덕 판단이 강력한 정서적 요인을 포함하는 것은 분명하지만, 이것이 결코 도덕 판단의 인지적 요인의 영향력을 약화하는 것은 아니다. 정서적 요인의 양적인 강도(quantitative strength)가 과학적 판단의 발달을 이해하기 위한 인지적 요인의 이론적 중요성을 인정하는 것과 무관하듯, 도덕적 감정의 정량적 차원에서의 역할(quantitative role)은 도덕 판단의 발달과 구조를 이해하는데 무관하기 때문이다(Kohlberg, 1981: 139-140).*

이처럼 콜버그는 도덕 판단에 있어 정서적 요인의 영향력에 대해 인정하고 있다. 그러나 아동의 도덕 발달 수준을 확인하고, 더 높은 차원의 정의 추론 구조를 파악하기 위해 인지적 요인에 주목해야 하며, 정서적 요인은 그 작업을 훼손시키지 않는다. 그는 도덕 판단에 있어 정서적 역할에 대해 다음과 같이 말한다.

> *도덕 판단은 주로 사람들의 감정과 직관에 대한 것이다. 나아가 대부분의 도덕 판단은 판단을 내리는 주체의 감정에 의해 표현되고 그 감정을 바탕으로 정당화되는 것이다(Kohlberg, 1981: 140).*

따라서 콜버그의 관점에서 동기화된 추론은 낮은 수준의 정의 추론 구

조에 기반을 둔 도덕 판단과 관련된다. 그는 같은 감정적 반응에 대해 서로 다른 도덕 발달의 단계는 감정에 대한 서로 다른 대응을 나타낸다고 말한다. 가령, 도둑질할 때 느끼는 '불안감(feeling of anxiety)'에 대해 2단계에 있는 아동은 그 느낌을 '겁쟁이가 되는 것(being chicken)'이라고 해석하여, 이를 무시하는 판단을 내린다. 반면, 4단계에 있는 아동은 같은 느낌을 '양심의 경고(warrant of conscience)'로 해석하여, 그 감정에 따라 도덕적인 결정을 내릴 것으로 예측할 수 있다. 그는 이러한 반응의 차이가 정서적 역동성의 차이가 아니라 도덕 판단의 인지 구조적 단계의 차이에서 기인하는 것임을 강조한다(Kohlberg, 1981: 141). 이처럼 정서는 발달 단계에 따라 질적으로 다른 도덕적 추론에 영향을 준다.

[표 3-1] 도덕 발달 단계에 따른 도덕적 행위의 동기(Kohlberg, 1981: 130-131)

단계	도덕적 행위의 동기(정서적 요인)
1단계	행위는 벌을 피하려는 동기에 의해 이루어지며, 양심은 벌에 대한 터무니없는 공포이다.
2단계	행위는 보상이나 이득에 대한 욕망으로 동기화된다. 있을 수 있는 죄의식 반응은 무시되고 실용적 방식으로 벌이 나타난다.
3단계	행위가 현실적이든 상상에 의한 가설적인 것이든(예컨대 죄의식), 다른 사람의 불승인을 예상하면서 동기화된다(불승인은 벌, 공포, 고통과 분화된다).
4단계	행위는 불명예를 예상함으로써 동기화된다. 즉 의무의 태만에 대한 제도적 질책에 의해 동기화되고, 다른 이에게 돌아간 구체적 피해에 대한 죄의식으로부터 동기화 된다(형식적 불명예를 비형식적 불승인으로부터 분화시킨다. 사악한 결과에 대한 죄의식을 불승인으로부터 분화시킨다).
5단계	공동체에 대한 존경과 평등한 관계에 대한 존경을 지켜가려는 관심이 있다(평등한 관계의 존경을 가정하는 것은 정서보다는 오히려 이성 위에 기초해 있다). 또한 자기 존경에 대한 관심도 있다. 즉, 자아를 비합리, 비 일관, 무목적으로서 판정하는 일을 피하려는 것이다.
6단계	자신의 원리를 범한 데 대한 자기비난에 관심이 있다(공동체의 존경과 자기 존경을 분화시킨다. 일반적인 합리성의 성취에 대한 존경과 도덕원리의 고수에서 오는 자기존경을 분화시킨다).

전 인습 수준(1, 2단계)의 아동은 벌의 공포를 피하려는 마음이나 자기 이익이라는 욕망을 충족하고자 하는 정서적 동기에 기반하여 도덕 판단

을 내린다. 인습 수준(3, 4단계)에 있는 아동의 도덕 판단과 도덕적 행위는 주로 자신의 사회적 평판과 정체성을 유지하고자 하는 동기에 기반한다. 후 인습 수준(5. 6단계)에서 아동은 공동체와 자신을 분화시키며 최종적으로는 도덕적 행위로 나아가는 데 정서적 동기를 극복하는 모습을 보인다. 이처럼 인지와 정서는 도덕 판단의 서로 다른 두 측면이며, 도덕적 행위의 동기화로서 정서적 요인은 도덕 판단을 촉발하는 계기가 될 수 있다. 인지적 성숙에 따른 정의 추론 구조의 질적인 차이는 도덕적 상황에 대한 정서적 반응을 이해하는 데 있어 질적으로 다른 해석의 틀을 제공한다. 콜버그의 도덕 발달 연구에서 동기화된 추론은 아동의 도덕 발달을 파악하기 위한 직접적인 요인은 아니나, 도덕 판단에 영향을 미치는 요인으로 파악하고 있음을 확인할 수 있다. 특히, 도덕 발달의 단계는 정의 추론에 관여하는 자기본위적인 동기를 극복해가는 과정으로 이해할 수 있다. 성숙한 도덕 발달의 단계에서 도덕적 추론은 자신을 보호하고자 하는 동기에서 벗어나 모든 사람을 공평하게 대하는 가역성에 근거한 도덕 원리로부터 합리적으로 연역되기 때문이다.

2) 하이트의 사회적 직관주의 모델에서 동기화된 추론의 이해

정의 추론을 중심으로 도덕 판단 연구를 수행한 콜버그의 도덕 발달 연구는 '인지 혁명'이라는 인지과학의 거대한 학문적 조류 속에서 30여 년 간 활발하게 이루어져 왔다. 그러나 인지과학은 2000년대에 들어 새로운 연구 방법론 및 다양한 접근법과 함께 '정서 혁명'과 '자동성 혁명'의 변화를 맞이하게 된다. 도덕 심리학 연구 또한 새로운 국면으로 전환되었고, 사회 심리학자 하이트는 이를 도덕 심리학의 '새로운 종합(new synthesis of moral psychology)'의 시대라고 명명하였다(Haidt, 2007).

새로운 종합의 시대는 사회생물학자 윌슨(Wilson, 1978/2000)에 의해 처음으로 제안되었다. 그는 그동안 인간의 본성과 사회성에 대한 연구가 진화적 관점을 간과한 채 철학의 전유물로 여겨졌음을 지적한다. 그에 따르면 대부분의 철학자들은 진화적 관점에 관심을 기울이지 않았으며,

도덕 규범의 기원이 아닌 결과에 초점을 맞추어 연구했다. 그는 대표적으로 롤스와 노직의 관점을 인용하며 자신의 입장을 전개한다. 이들은 근본적으로 다른 처방을 이끌어내지만, 두 처방이 갖는 정서적 반응을 암묵적인 가정으로 처리하며 그 기원에 대해서는 탐구하지 않았다(Wilson, 1978/2000: 28). 윌슨은 도덕에 대한 정서적 반응의 기원을 탐구함에 있어 대뇌 피질의 사고 영역 바로 아래에 있는 감정 중추 및 신경 세포와 호르몬 분비 세포의 복합체인 변연계에 주목한다. 인간의 정서적 반응은 수천 세대 동안에 걸쳐 자연 선택을 거치며 적응과 생존의 결과로 우리 뇌의 변연계 깊은 곳에 프로그래밍 되어왔다는 것이다(Wilson, 1978/2000: 29). 이처럼 도덕성에 대한 진화심리학 및 생물학적 접근은 도덕 심리학의 연구를 인간 도덕성의 기원에 초점을 전환시켰고, 도덕의 기능이 어디에서 유래되었는지를 규명하고자 하였다. 특히, fMRI 등과 같은 뇌에 대한 연구 방법 및 기술적 향상으로 인해, 기존의 도덕 심리학 연구가 의존해 온 실험 참가자의 자기보고식 측정이 획기적으로 개선되었다. 도덕 심리학 연구에서 뇌-신경과학적 접근의 접목은 도덕 판단의 실제적 과정을 규명하는 데 크게 기여했다. 그리고 그 결과는 콜버그식 도덕 발달 연구를 반박하는 근거들을 제공했다.

하이트(Haidt)의 연구는 도덕 심리학 분야에서 자동성 및 정서 혁명의 전환을 가져왔다. 그는 콜버그 이후에도 여전히 심리적 합리주의가 도덕 심리학의 원칙으로 자리잡고 있으며, 이 때문에 '도덕성'이 도덕적 표준들에 대한 아동의 지식의 구조화 상태로 규정되었다고 본다. 특히, 하이트는 콜버그가 전제한 정의 우선성에 비추어 도덕성은 정의롭지 못한 것, 해로움, 권리 침해를 유발하는 행동을 포함한 것을 인식하도록 하고, 다른 종류의 사회적 규칙 위반과는 구별하여 취급되어 왔음을 지적하였다(Haidt, 2001; 2012/2014). 즉, 정의롭지 못한 것, 해로움, 권리의 침해를 제외한 것은 사회적 관습 혹은 개인적 문제의 위반으로 취급해온 것이다(Turiel, 1983). 하이트의 초기 연구는 일상의 도덕 판단에 있어 문화의 영향력과 정서적 반응에 주목하며 콜버그 중심의 도덕 발달 모델의 설명력을 경험적으로 평가하는 것에 초점을 맞추었다. 그는 도덕 판단에

있어 반성적이고 의식적인 추론의 인과적 역할을 검증하고자 한 것이다.

도덕 심리학에서 합리주의 모델은 추론과 반성적 사고의 과정을 통해서 도덕적 지식을 획득하고 도덕적 판단을 할 수 있다고 전제한다(Kohlberg, 1969; Piaget, 1965/2000; Turiel, 1983). 이 모델에 따르면 도덕적 정서는 추론의 과정에 종종 영향을 주긴 하지만, 도덕 판단의 직접적인 원인으로 고려되진 않는다(Haidt, 2001: 814). 그러나 하이트가 보기에 일상의 도덕 판단은 추론의 과정 보다는 직관의 과정을 통해 도출된다. 그는 이를 입증하기 위한 실험에서 도덕적 말 막힘 현상(moral dumbfounding)을 발견하였다. 도덕적 말 막힘 현상이란, 자신이 내린 도덕 판단에 대해 타당한 근거를 제시하지 못함에도 최초의 판단을 수정하지 않고 고집하려는 과정에서 나타나는 말문이 막히는 상태를 의미한다(Haidt et al., 2000: 1689). 하이트는 이를 근거로 도덕 판단에 있어 도덕적 추론이 인과적 기능을 수행하기보다는 자신이 도달하고자 하는 결론(이미 내려진 도덕 판단)에 부합하는 방향으로 동기화된 추론의 일종임을 보였다(Haidt et al., 2000; Haidt, 2001; 2007; 2012/2014).

하이트는 도덕 판단을 '문화 혹은 하위문화가 지닌 책임으로서의 덕에 대한 존경심으로 형성되는 행동 및 인격에 대한 평가'라고 정의하며, 도덕적 추론은 본질적으로 직관적인 신념에 대한 사후 정당화에서 비롯되는 것으로 본다(Haidt, 2001: 817). 정서적 반응은 해로운 결과에 대한 요구보다도 도덕 판단에 더 나은 예측자이다. 그러나 소위 합리주의를 표방하는 콜버그의 도덕 판단 모델은 판단과 이를 지지하는 신념 사이의 상관관계로서 신념이 판단을 유발한다는 것을 보여주지 못한다. 또한 하이트는 도덕적 추론을 '사람들이 도덕 판단에 도달하기 위해 주어진 정보를 변형하는 의식적이고 정신적인 활동'으로 정의한다(Haidt, 2001: 818). 즉, 추론은 판단과 관련된 증거를 탐색하고, 증거의 비중을 조사하고, 증거를 조합하며 결정에 도달하는 인지 처리 과정이다.

하이트와 그 연구진(Haidt, 2001; Haidt et al., 2000)이 수행한 도덕 판단에 대한 실험은 '무해한 금기 위반'에 초점을 맞춘다. 남매간의 합의된 성관계, 국기로 화장실을 청소하는 것, 교통사고로 죽은 반려견을 먹는

것 등이 대표적이다. 하이트는 이러한 이야기에 대해 실험 참가자들이 보이는 부정적 판단을 옹호해 달라고 요청했을 때, 사람들은 쟁점이 되는 해당 이슈의 범위를 벗어나면서까지 자신의 도덕적 직관과 신념에 부합하는 근거를 찾아내어 이유를 대는 모습을 발견한다.

무해한 금기 위반 이야기를 접한 사람들은 대부분 즉각적으로 그 행위가 도덕적으로 용납될 수 없다고 판단하지만, 해당 이야기의 구체적인 상황에 기반했을 때 적절한 이유를 제시하지 못했다. 가령, 줄리와 마크 이야기에서 남매의 행위가 도덕적으로 잘못된 이유를 대라고 압박하면 많은 실험 참가자들이 근친상간으로 인해 기형아를 낳는다거나 정서적으로 상처를 받을 것이라고 응답하지만, 이야기 속 단서들에 의해 해당 근거가 기각되었고 그들은 '근친상간'이라는 사실 이상으로 그들의 도덕 판단에 대한 그럴듯한 이유를 명확히 말할 수 없었다(Haidt et al., 2000). 즉, 도덕적 추론은 올바른 도덕 판단을 연역적으로 도출해내는 역할을 수행하기보다는, 자신의 기존 도덕적 신념에 기반하는 판단을 지지해주는 근거를 수집하여 제시하는 방식으로 이루어진다는 것이다.

하이트는 이 실험을 통해 사람들이 옳고 그름에 대한 판단을 내릴 때 종종 도덕적 확신을 경험한다는 것을 발견했는데, 도덕적 말 막힘 현상은 심지어 그들이 도덕 판단을 내리고도 명확하고 합리적인 설명을 제공할 수 없을 때도 확신에 찬 모습으로 묘사된다. 해당 연구와 관련한 인터뷰에서 연구진은 사람들이 도덕적 말 막힘 현상을 보이는 상황에서도 편안한 상태(comfortably dumbfounded)를 유지한다고 답했다(Sommer, 2009). 하이트 연구진이 수행한 일련의 도덕적 말 막힘 연구는 추론이 도덕 판단의 형성에 중요한 역할을 하지 않는다는 결론을 도출해냈다. 도덕적 추론은 판단을 직접 도출하는 역할을 하는 것이 아니라, 주로 사후 정당화를 만들어내는 역할을 수행하기 때문이다.

이처럼 도덕적 말 막힘 연구는 도덕 판단에서 도덕적 추론이 동기화된 추론의 형태를 보인다는 것을 알 수 있는 대표적인 연구이다. 하이트의 연구 이후, 많은 연구는 도덕적 말 막힘이 다소 흔한 현상임을 밝혔다. 대표적으로는 도덕 판단 연구에서 가장 많이 활용되는 트롤리 딜레마를

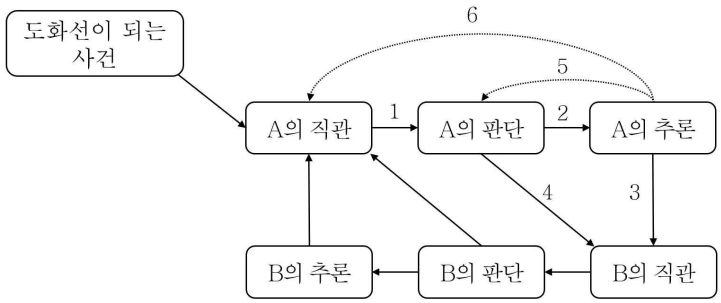
활용하여 도덕적 말 막힘 현상을 발견한 여러 연구가 있다(Hauser, 2006; Hauser et al, 2007; Cushman et al., 2006; Young, & Hauser, 2006). 트롤리 딜레마의 기본 형태는 다음과 같다. 선로에 몸이 묶여 움직일 수 없는 다섯 명의 사람이 있는 상태에서 전차 한 대가 그 무리로 질주하고 있다. 이때, 이들을 구하기 위해 전환 스위치를 당겨 선로를 바꿀 수 있다. 그러나 그렇게 되면 다른 선로에 있는 다른 한 사람이 죽게 된다. 이 상황에서 선로 전환 스위치를 당기는 것은 도덕적으로 허용될 수 있는가? 하우저(2006)에 따르면 대부분의 사람(약 90%)은 스위치를 누른다고 판단하며, 이에 대한 이유를 묻는 질문에는 일반적으로 인명 피해의 수적인 차이를 꼽는다. 그러나 동일한 응답자에게 초기 트롤리 사례와 구조적으로 유사한 변형으로 보이는 것을 판단하도록 요청했을 때, 그들은 사례를 매우 다르게 판단한다. 대표적인 변형 사례는 인도교(foot bridge) 사례이다. 트롤리는 선로를 따라 달려오고 있고, 선로에는 다섯 사람이 있다. 당신은 선로 밖에 서 있고, 바로 옆에는 상당히 무거운 사람이 한 명 서 있다. 다섯 사람을 구하는 유일한 방법은 옆에 서 있는 사람을 선로 위로 밀쳐서 그 무게로 트롤리를 멈추게 하는 것인데, 이 경우 트롤리는 멈추게 되지만 그 사람은 죽게 된다. 이는 도덕적으로 허용 가능한가? 대부분의 경우(70%)에서, 응답자들은 두 경우를 다르게 판단하기 위한 일관된 정당성을 명확히 말하지 못했다(Hauser, 2006: 128). 문제는 철학자들이 차이를 설명할 도덕적 원칙을 찾을 수 없다는 것이 아니라, 응답자들이 이러한 다른 도덕 판단을 하지만, 그에 대한 일관된 이유를 제시할 수 없다는 점이다. 그들은 도덕적 말 막힘 상태에 빠진 것이다. 후속 연구(Cushman, Young, & Hauser, 2006)에서는 다소 다른 방법을 사용하여 유사한 결과를 발견했다. 연구진은 대부분의 사람들이 일관되게 서로 다른 판단을 일으키는 한 쌍의 도덕 딜레마 상황을 제시했다. 가령, 첫 번째 딜레마에서 응답자들은 도덕적으로 허용 가능한 행동이라는 판단을 내리지만, 두 번째 딜레마에서는 도덕적으로 용납할 수 없다는 판단을 내릴 것으로 예상되었다. 각각의 딜레마 쌍에서, 서로 다른 판단은 다음의 세 가지 도덕적 원칙 중 하나를 참조하여 설명할 수

있다. 세 가지 도덕적 원칙은 ①행동 원칙(직접적인 위해는 누락에 의해 야기된 위해보다 도덕적으로 더 나쁘다), ②의도 원칙(의도적인 위해는 부작용으로 야기된 위해보다 도덕적으로 더 나쁘다), ③접촉 원칙(피해를 야기하기 위한 물리적 접촉의 사용은 그러한 접촉 없이 초래되는 피해보다 더 나쁘다)이다. 해당 연구에서 발견한 것은 응답자들이 이 세 가지 원칙에 따라 사례 쌍(dyad)을 신뢰성 있게 판단했지만, (행동 원칙을 제외하고) 판단에 대한 정당성에는 사례 쌍(dyad)을 인용하지 않았다는 것이다. 많은 경우에, 응답자들이 제공한 정당성은 명백히 '판단의 패턴을 정당화할 수 없는 무능함'에 의해 촉발되었다(Cushman et al, 2006: 1086). 다시 말해서, 그들은 도덕적으로 말문이 막히는 상태에 빠졌다.

이처럼 도덕적 말 막힘 현상의 발견은 도덕 판단의 연역적 모델에 심각한 문제를 제기하는데, 이는 사람들이 도덕 판단에 대해 제시하는 이유가 항상 그러한 도덕 판단의 기초가 아니라는 것을 드러내기 때문이다. 대신, 사람들이 도덕 판단의 일부에 대해 제시하는 이유는 명백하게 거짓이거나 불충분하고, 무관한 것으로 드러났다. 심지어 이것이 지적될 때조차도 대부분의 사람은 그들의 도덕 판단을 의미 있는 방법으로 재고하기를 꺼렸다. 따라서 이 현상은 도덕 판단에 대한 합리주의적 모델을 경험적으로 반박하는 근거로 제시될 수 있다[27]. 이처럼, 도덕 판단에서 도덕적 추론은 자신의 기존 도덕적 신념이라는 방향성에 의해 동기화된 추론의 일종으로 드러날 수 있다.

이는 하이트가 심리적 합리주의 모델을 대체하는 새로운 도덕 판단 모델로 제시한 사회적 직관주의 모델 내에서 명료하게 이해될 수 있다 (Haidt, 2001).

27) 이와 관련하여 하이트는 연구결과 대상자의 16%만이 마크와 줄리가 아이를 가질 수 있거나 정서적으로 상처받을 가능성을 배제하고 있다고 지적했다(Haidt et al., 2000).



[그림 3-1] 도덕 판단의 사회적 직관주의 모형 (Haidt, 2001: 815)

1번 링크: 직관적 판단(the intuitive judgment link)
2번 링크: 사후 추론(the post hoc reasoning link)
3번 링크: 추론된 설득(the reasoned persuasion link)
4번 링크: 사회적 설득(the social persuasion link)
5번 링크: 추론된 판단(the reasoned judgment link)
6번 링크: 사적인 반성적 사고(the private reflection link)

사회적 직관주의 모델에 따르면 일상의 도덕 판단은 주로 직관적 판단(1번 링크)을 통해 즉각적으로 나타나며, 도덕적 추론은 이러한 판단에 뒤따른 사후 추론(2번 링크)의 과정이다. 이처럼 도덕 판단에서 추론은 직관적 판단을 지지하는 근거를 탐색하는 방식으로 전략적으로 수행된다. 도덕 판단에서 도덕적 추론은 일종의 동기화된 추론으로 이해할 수 있는 것이다.

1번과 2번 링크가 도덕 판단의 '직관적' 과정을 나타내는 것이라면 3번과 4번 링크는 도덕 판단의 '사회적' 상호작용 맥락을 묘사한다. 도덕 판단이 사회적이라는 것은 도덕 판단의 과정을 이해하기 위해서는 사회적 맥락과 개인들 사이의 관계를 연구해야 한다는 의미를 내포한다(Haidt, 2001: 814). 추론된 설득(3번 링크)은 직관에 의해 도출된 한 개인의 도덕 판단을 명시적인 도덕 언어를 통해 타인에게 정당화하는 과정을 설명한다. 이 과정은 한 개인의 도덕적 논증이 다른 개인의 도덕적 직관과 판단에 영향을 미친다는 사실을 보여준다. 그러나 하이트는 도덕적 추론을 통한 타인의 도덕적 변화에 대해 부정적이다(Haidt, 2001: 819). 그는 도덕적 토론이나 논쟁이 타인의 직관에 영향을 주는 것은 도덕적 추론이

갖는 논증 자체에 있기보다는, 그것이 갖는 정서적인 설득력에 있다고 본다. 한편 사회적 설득(4번 링크)은 사람들이 문화적으로 축적되어 온 집단의 규범에 따르기 때문에 한 개인의 도덕적 직관은 공동체가 제공하는 사회적 판단에 의존하는 경향을 나타낸다. 사람들은 자신이 속한 집단을 지배하는 규범의 영향 하에 있기 때문에 타인이 내린 도덕 판단에 직접적인 영향을 받는다.

1~4번 링크는 사회적 직관주의를 구성하는 핵심적인 링크이다. 하이트는 추론이 인과적으로 도덕 판단을 내릴 수 있다는 가능성 또한 거부하지 않는다. 그러나 추론된 판단(5번 링크)은 도덕적 주체의 인지 능력이 직관의 영향을 압도하는 범위 내에서 가능하다. 가령 철학자들은 성찰의 훈련을 받지 않은 사람들에게서 공통적으로 나타나는 직관과 충돌하는 이론을 명확하게 표현할 수 있다. 또한 주체가 단순한 딜레마를 해결하려고 할 때와 같이 두 가지 약한 직관 사이에서 고민하고 사회적 영향력이 없거나 큰 도움이 되지 않는 경우에 내적인 반성적 사고(6번 링크)를 통해 추론의 결함에서 벗어날 수도 있다. 하이트는 5~6번 링크가 콜버그 학파의 숙고의 과정을 나타내지만, 일상의 도덕 판단에서는 매우 드물게 일어난다고 보았다.

하이트의 연구는 도덕 판단에서 추론은 항상 상황에 대한 철저한 분석을 반영하는 것이 아니라(5번 링크), 개인적인 신념과 가치에 의해 동기화될 수 있다(2번 링크). 하이트는 도덕적 추론이 진리를 탐구하는 과학자의 모습이나 객관적이고 사심 없는 판사의 역할과 같은 기능을 하는 것이 아니라 의뢰인의 승소를 목표로 편향적으로 증거를 수집하고 때로는 사실을 왜곡하거나 은폐하기도 하는 변호사와 같은 역할을 담당한다고 비유적으로 표현한다. 특히, 그는 도덕 판단에서 동기화된 추론의 양상이 사후 정당화의 모습으로 나타난다는 것을 지적하며 다음과 같이 서술한다.

언어적 진술을 하도록 요구할 때, 사람들은 진실을 탐색하는 심판자라기보다는 그 사례에 관하여 변호사처럼 논리적으로 설명하려고 고심한다(Haidt 2001: 814).

추론 과정은 과학자가 진리를 탐색하는 과정이라기보다는 법률가가 의뢰인을 방어하는 것과 유사하다(Haidt, 2001: 820).

직감과 직관 그리고 즉각적인 판단은 지속적으로 동시에 일어나지만 기수(합리적 추론)는 문장을 연결하여 상대방에게 제시할 논거를 만들어낼 수 있을 뿐이다. 도덕적 논쟁에서 기수는 코끼리의 조언자 역할을 넘어 남에게 코끼리의 입장을 옹호하기 위해 여론의 법정에서 싸우는 변호사가 된다(Haidt, 2006/2010: 53)

즉, 기수(이성적 추론)는 주인 밑에서 일하는 하인이라기보다는 고객을 위해 일하는 변호사에 더 가깝다. (Haidt, 2012/2014: 140)

이처럼 도덕적 추론이 도덕 판단의 인과적인 심리 기제라고 분석한 콜버그와 달리, 하이트는 문화적 토대에서 기인하는 도덕적 직관의 강력한 반응으로 도출된 도덕 판단을 사후적으로 정당화하는 기능이 추론의 주된 역할임을 강조한다. 도덕 판단에서 도덕적 추론은 사실에 대한 철저한 탐구에 기여하기보다는, 자신의 기존 신념에 일치하는 방향으로 근거를 수집하고 그럴듯한 이유를 찾는 역할을 수행하기 때문이다.

동기화된 추론의 유형 구분에 따르면 하이트가 발견한 도덕 판단에서의 동기화된 추론의 양상은 방향성 동기에 의해 작동하는 추론의 일종으로 분석될 수 있으며, 구체적으로는 낮은 정확성을 지향하기 때문에 '당파적 추론'에 해당한다. 실제로 하이트의 연구에서 실험 참가자들이 보인 추론의 양상은 자신의 직관적인 반응으로서 도출된 도덕 판단이라는 특정한 결론을 합리화하기 위해 사용되었다는 점에서 높은 방향성을 보였다. 또한 실험 참가자들이 제시한 논증은 연구자가 설계한 상황 내 단서를 통해 쉽게 반박될 수 있는 것이었음을 미루어볼 때, 그들의 동기화된 추론은 낮은 정확성을 보였음을 확인할 수 있다. 이와 같은 하이트의 연구를 계기로, 동기화된 추론 연구자들은 도덕 판단에서 작동하는 동기화된 추론의 기제에 주목하였다.

2. 도덕 판단에서 동기화된 추론의 특징

인간은 옳고 그름에 대해 깊은 관심을 가지고 있으며, 자신의 도덕적 신념을 보호하고 정당화하기 위해 인지적인 노력을 기울인다. 도덕 판단을 도출해 내거나 자신의 판단을 정당화하기 위해서는 적절한 논증을 구성해야 하고 이것이 도덕적 추론의 역할이다. 그러나 동시에 도덕 판단과 정서는 불가분의 관계에 있다. 정서와 직관이 인간의 판단과 의사결정에 중추적인 역할을 한다는 것은 최근의 경험적 연구를 통해 규명되고 있다. 일상생활에서의 도덕 판단이 도덕적 문제에 대해 가지고 있는 정서에 영향을 크게 받는다는 사실이 널리 받아들여지며, 도덕 판단에서 정서의 역할을 이해하는 것은 이제 도덕 심리학 연구의 중요한 초점이 되었다(Forgas & Smith, 2003; Greene et al., 2001; Haidt, 2001; 2007; Monin et al., 2007).

도덕 판단에서 정서의 역할에 주목한 하이트 연구는 동기화된 추론의 개념을 통해 도덕적 추론이 도덕 판단에 대한 사후적 정당화의 기능을 수행함을 밝혔다. 도덕 판단에서 동기화된 추론은 자신의 도덕적 직관에 부합하는 정보와 증거를 수집하고, 그와 반대되는 것들은 무시하거나 축소함으로써 자신이 도달한 결론을 설명하거나 정당화하는 방식으로 사용되었기 때문이다. 이는 도덕 판단을 내리는 우리가 무엇이 옳고 그른지에 대해 그 이유를 즉시 설명할 수는 없지만, 도덕적 추론을 통해 개인이 판단을 내린 이후에 자신의 도덕적 신념과 일치하는 근거를 제공할 수 있다는 것을 의미한다. 동기화된 추론의 맥락에서 볼 때, 사람들의 도덕적 신념은 종종 그들의 개인적인 목표, 가치, 정서에 의해 영향을 받기 때문에 합리적으로 느껴지지 않을 수 있고, 따라서 사후적인 정당화 작업이 수행되는 것이다.

도덕 판단에서 동기화된 추론의 사후 정당화 양상은 크게 두 가지 기능을 수행한다. 하나는 사후 정당화를 통해 도덕적 신념에 대해 보다 일관되고 논리정연한 설명을 제공해준다는 점이고, 다른 하나는 자신이 내

린 도덕 판단에 대한 자신감과 자기 확신을 향상시킨다는 점이다. 전자와 관련하여, 동기화된 추론은 도덕 판단에 대한 사후 정당화를 통해 개인들이 그들의 도덕적 신념에 대해 더 일관성 있고 일관된 설명을 제공할 수 있게 해준다. 이러한 양상은 자신의 도덕적 신념을 공고히 하고 외부 사건으로 인한 신념 변화에 저항하는 데 기여한다. 후자와 관련하여 동기화된 추론은 자신의 도덕 판단에 대해 그럴싸한 근거를 제시해주기 때문에, 자신감을 느끼게 해줄 수 있다. 사람들은 도덕 판단을 내릴 때, 자신의 신념에 대한 이유에 대해 불확실하거나 양면적인 감정을 느낄 수 있다. 그러나 사후 정당화를 통해, 자신의 가치, 정서, 그리고 그들의 판단에 영향을 준 다른 요소들과 일치하는 그들의 신념에 대한 이유를 제공함으로써 자신의 판단에 더욱 확신을 갖게 된다.

나아가 하이트에 따르면 도덕 판단에서 동기화된 추론에 전제된 동기의 내용, 즉 목표는 두 가지-관계성 동기(relatedness motives)와 일관성 동기(coherent motives)-로 분석될 수 있다. 먼저, 관계성 동기는 다른 사람과의 관계 속에서 자신의 평판을 유지하고자 하는 욕구로서 추론을 동기화한다. 하이트는 진화적으로 인간은 자신이 속한 집단이 추구하는 가치 체계에 적응해왔으며, 적대적인 집단의 관점과는 대립하는 방향으로 도덕 판단을 하도록 설계되어 있음을 강조한다. 즉, 우리는 자신과 가까운 동료나 친구, 자신이 준거를 두고 있는 집단과 일치하려는 동기가 존재하기 때문에 도덕적 추론은 그러한 신념에 부합하는 방향으로 동기화해 자신의 도덕 판단을 사후 정당화하는 모습으로 나타난다는 것이다. 일관성 동기는 주어진 도덕적 상황과 자신의 도덕적 신념 사이에서 발생하는 충돌을 해소하고, 자신의 도덕적 신념을 일관되게 유지하고자 하는 욕구로서 추론을 동기화한다.

한편, 일관성 동기를 설명할 때 하이트는 프로이트의 방어기제의 개념과 페스팅거의 인지부조화 이론[28]을 제시한다. 방어기제는 지그문트 프로이트가 그의 정신분석 이론에 도입한 용어이다. 프로이트에 따르면, 방

28) 인지부조화 이론은 동기화된 추론 연구의 모태가 되는 개념으로 본고 2장에서 상세히 논의 되었으므로, 이 절에서는 동기화된 추론의 사후 정당화의 양상을 설명하는 방어 기제에 대해서만 서술하였다.

어기제란 개인이 그들의 행동이나 생각에서 비롯될 수 있는 불안, 죄책감, 또는 다른 부정적인 정서로부터 자신을 보호하기 위해 사용하는 무의식적인 심리학적 전략이다. 사람들은 저마다 자신의 도덕적 신념을 구성하고, 그 신념이 위협을 받을 때 이것이 훼손당할 것이라는 상당한 불안함을 경험하게 되는데, 이에 대한 방어기제로써 사용되는 전략 중 하나가 합리화(rationalization)이다. 합리화는 자신의 무의식적 욕망, 충동, 갈등에 의해 나타난 자신의 행동이나 판단에 대해 그럴듯하고 사회적으로 수용 가능한 설명을 제공하는 과정을 의미한다. 동기화된 도덕적 추론의 사후 정당화는 이처럼 합리화의 모습을 띤다.

이처럼 관계성 동기와 일관성 동기는 도덕 판단에서 사후 정당화의 기능으로 수행되는 동기화된 도덕적 추론의 양상이 판사나 과학자의 모습보다는 법률가의 모습을 보이는 이유를 설명해 준다. 즉, 도덕 판단에서 도덕적 추론은 정확성 목표에 의해 동기화되기보다는 자신의 기존 신념에 부합하는 결론으로 도달하기 위한 방향성 목표에 의해 동기화되는 모습을 띤다. 특히, 하이트는 동기화된 도덕적 추론의 사후 정당화 양상을 사회인지이론에서 제시하는 '내 편 편향(my-side bias)'과 '확증 편향(confirmation bias)'의 모습으로 묘사한다. 추론의 심리학 영역의 권위자 스타노비치는 내 편 편향을 "우리가 자신의 사전 견해와 태도에 우호적인 방식으로 증거를 평가, 생성하고, 가설을 검증할 때 나타나는 편향"으로 정의한다(Stanovich, 2021/2022: 39). 스타노비치에 따르면 내 편 편향은 동기화된 추론의 하위 범주로서, 우리가 강력하게 고수하고 있는 신념, 즉 확신과 일치하도록 증거를 해석하도록 한다. 확증 편향이란 개인이 기존의 신념과 태도를 뒷받침하는 정보에 더 비중을 두고, 이에 반하는 정보는 무시하거나 경시하는 경향을 설명하기 위해 사용되는 개념이다(Nickerson 1998: 175). 확증 편향은 사람들이 의견을 형성하고, 결정을 내리는 심리적 과정에서 주요한 역할을 한다. 가령, 사람들은 자신의 기존 신념을 확인(confirm)하기 위한 정보를 찾고, 자신의 기존 신념을 반증하는(disconfirm) 정보는 무시하는 방식으로 자신의 신념을 유지하게 해준다[29]. 이와 같이 내 편 편향과 확증 편향은 도덕 판단에서 사후

정당화의 기능을 수행하는 동기화된 추론의 양상을 설명해주며, 도덕 판단에서 동기화된 추론으로 인해 발생할 수 있는 다양한 도덕적 문제를 분석하는 틀을 제공한다.

한편, 최근 사회인지이론을 중심으로 수행된 도덕 심리학 연구들은 도덕 판단에서 도덕적 추론이 정서의 영향을 받으며, 방향성 목표에 의해 동기화되는 양상에 주목한다(Cusimano & Lombrozo, 2020; 2021; Ditto et al., 2009; Uhlmann et al., 2009). 이들은 도덕 판단을 구성하는 요소들이 가지고 있는 특징이 동기화된 추론의 요소와 밀접한 관계가 있음을 지적한다.

인간의 인지과정이 정서적인 측면에 결합 되어 있다는 전제 속에서 동기화된 추론 연구자 디토(Ditto et al., 2009)는 판단과 관련된 심리학 연구를 바탕으로 동기화된 추론의 과정이 도덕적 문제에 대한 사고에 미치는 영향에 주목하였다. 그는 '상대적인 정서적 가치에 기반을 둔 도덕 체계는 단지 환상일 뿐이며, 그 안에 아무 소리도 없고 어떤 진실도 없는 완전히 저속한 개념'이라는 소크라테스의 『파이돈』의 일부를 인용하며, 옳고 그름의 문제를 다루는 도덕 판단의 영역은 기본적으로 동기 부여된 추론의 맥락에서 이해될 수 있음을 분명히 한다(Ditto et al., 2009: 307). 도덕적 상황에 대한 정서적 반응은 인간의 본능적인 속성이며 이는 직관의 형태로 즉각적으로 나타난다. 특히, 도덕적으로 중요한 질문을 제시하였을 때, 냉철한 이성적 판단보다는 정서적 이해관계가 즉시 높아지는 모습을 보였다(Haidt, 2001). 그러나 자신의 도덕적 신념을 토대로 내려진 판단을 옹호함에 있어 단순히 주관적인 선호에 호소하는 것은 설득력이 없다. 왜냐하면 그것은 다른 사람들이 동의해야 하는 설득력 있는 논증을 제시하지 못하기 때문이다. 도덕 판단은 일반적인 선호를 판단하는 것과는 질적으로 다른 판단에 포함된다. 따라서 우리는 자신의 도덕적 신념이 기대고 있는 도덕 원칙과 원리를 제시하며 도덕 판단을 설득력 있게 정당화하고자 한다. 여기에서 발휘되는 것이 도덕적 추론의 역할이

29) 내 편 편향과 확증 편향에 대한 개념 간 구분은 연구자들마다 다르게 설정하고 있으나, 주로 확증 편향이 '자신의 신념'에 일치하는 방향으로 정보를 처리한다는 점에서 '내 편 편향'과 동일한 것으로 보기도 한다(Mercier & Sperber, 2017).

다. 이러한 이해를 바탕으로, 디토 연구진(Ditto et al., 2009; Uhlmann et al., 2009)은 개인이 도덕적 문제 상황에서 어떤 행위 혹은 사람의 성품에 대한 도덕 판단을 내리는 과정에서 정서적 이해관계가 있을 경우, 도덕적 추론이 어떤 양상을 보이는지 확인하고자 하였다.

먼저, 디토 연구진은 사람들의 도덕적 신념이 정서적 이해관계에 따라 도덕적 추론에서 어떻게 선택되는지 확인하고자 개발된(Knobe, 2003) 이야기를 활용했다(Ditto et al., 2009). 이 이야기는 CEO가 선택한 사업이 직접적인 이익을 가져다주지만 시나리오에 따라 환경에 해를 끼치거나 (유해 시나리오) 환경에 도움이 되는 부수적인 효과가 뒤따르는 경우(도움 시나리오)로 구성되었다. 실험 참가자들에게 환경을 파괴하면서 이익을 얻는 선택을 할 의향이 있거나 이를 원하는 정도를 조사하였고 이익과 환경의 충돌하는 경우 우선순위에 대한 태도 관련 설문 조사하였다. 이 조사를 바탕으로 연구자들은 실험 참가자들을 두 그룹으로 나눴다. 한 그룹은 이익보다 환경을 우선적으로 생각하고, 자신에게 이익이 되더라도 환경을 파괴하는 사업은 절대적으로 꺼리는 그룹(환경 절대주의)과 환경을 파괴하더라도 부분적으로 자신에게 이익이 된다면 이를 수용하는 그룹(환경 비절대주의)이다. 환경 절대주의 그룹은 환경에 영향을 미치는 결과에 더 큰 도덕적 가치를 두었기 때문에 환경 비절대주의 그룹에 비해 유해 시나리오를 '의도적으로 환경을 해치는 것'으로 판단하는 경향을 보였으며, 반대로 도움 시나리오를 '의도적으로 환경을 돕는 것'으로 판단하는 경향이 낮았다. 또한 참가자들에게 그 행동이 의도적이거나 비의도적이라고 믿는 이유를 정당화하도록 요구했고 명확한 패턴이 발견되었다. 시나리오 속 환경을 파괴한 결과를 낳는 CEO의 선택이 '의도적'으로 행동했다고 보고한 실험 참가자 중 61%는 그 선택의 부작용에 대한 시나리오 속 단서(환경에 해를 끼치지만, CEO는 그것을 하기로 결정했다)를 언급했다. 반면, CEO의 선택이 의도적이지 않았다고 보고한 실험 참가자 99%가 목표 지향적인 상태(CEO는 환경에 관심이 없다)에 주목했다. 즉, 참가자들은 의도성에 대한 두 가지 다른 정의를 사용하고 있었다. 환경 절대주의자들은 CEO의 행동이 의도적이라는 것을 강조하기 위

해 의도성을 매우 느슨하게 적용한 반면, 환경 비절대주의자들은 CEO의 행동에 의도성이 없다는 것을 강조하기 위해 의도성을 매우 엄격하게 적용한 것이다. 이처럼 사람들은 자신의 도덕적 신념을 부합하는 방식으로 도덕 판단을 이끌어내는 과정에서 '의도성'이라는 개념을 선택적으로 적용하여 추론했다. 이처럼 서로 다른 도덕적 가치를 지향하는 사람들에 의해 인식된 의도성의 차이는 행위에 대한 의도를 평가하는 데 서로 다른 도덕 판단을 도출해낸다. 만약 어떤 행동이 우리의 도덕적 신념을 훼손하지 않는다면, 우리는 행위의 '의도성'의 의미를 엄격하게 적용하면서, 그 행동을 정당화할 것이다. 그러나 어떠한 행위가 우리의 도덕적 신념을 훼손시키고, 도덕적인 불쾌감을 일으킨다면 우리는 행위의 '의도성'을 느슨하게 적용하여, 그 행위의 예견적 책임까지를 포함할 것이다. 의도성에 대한 서로 다른 정의를 활용하는 것은 객관성에 대한 환상을 유지하는 강력한 방법이다(Ditto et al., 2009: 321).

또 다른 연구에서 디토 연구진은 사람들이 자신의 정치적 성향과 일치하는 판단을 지지하는 도덕적 원칙을 유연하게 지지한다는 증거를 제공한다(Uhlmann et al., 2009). 일련의 연구에서 그들은 참가자들에게 소수의 무고한 생명이 더 많은 것을 구하기 위해 희생될 수 있는 딜레마로 트롤리 문제를 변형하여 제시하였는데, 이는 결과론과 의무론에 대한 대립으로 이해할 수 있다. 디토의 연구에서 참가자들의 딜레마 해결 방식은 그들의 정치적 성향과 밀접한 관련이 있었다. 예를 들어 정치적 자유주의자들 사이에서 인종적 편견에 강한 반감을 가진 사람들은, 딜레마 속 주인공을 백인으로 추정되는 이름으로 제시하였을 경우보다 흑인으로 추정되는 이름을 제시하였을 경우 결과론적인 도덕 원리를 적용하여, 무고한 사람을 살해하는 것을 지지하는 경향이 더 낮다는 것을 발견하였다. 무고한 민간인 살해와 관련된 딜레마 상황에서, 정치적 보수주의자는 피해자가 미국인에 의해 살해된 이라크인일 때, 그 반대의 경우보다 군사작전 시 살해 상황을 용인할 가능성이 더 높다는 사실을 발견하였다. 이와 같은 연구 결과는 도덕 판단에 결합된 동기화된 추론이 단순히 세상이 어떤지에 대한 서술적 믿음(Dunning & Cohen, 1992; Kunda, 1987;

Norton et al., 2004; Simon, 2004; Simon et al., 2004; Uhlmann & Cohen, 2005) 뿐만 아니라, 이 세상이 어떻게 되어야 하는지에 대한 규범적 믿음에도 영향이 미칠 수 있다는 증거를 제공해준다(Uhlmann et al., 2009: 489에서 재인용). 보다 구체적으로는 도덕 판단에서 도덕적 추론의 실제적인 역할은 하이트가 전제한 전략적 추론인 당파적 추론의 유형에 가까웠다. 심지어 일부 연구에서는 사람들은 도덕적 추론의 과정에서 도덕 원리를 전략적으로 이용하는 모습을 발견하였다(Bargh et al., 1996). 즉, 사람들은 일반적인 도덕 원칙을 통해 특정한 판단으로 추론하는 고전적 합리성 유형의 추론보다는, 도덕적 직관에 대한 지지를 구축하는 데 도움이 되는 주장을 선택적으로 이용하는 당파적 추론 유형의 양상에 더욱 부합했다. 도덕 판단을 내림에 있어 대부분의 사람들은 자신이 선호하는 결론에 도달하기 위해 도덕적 원리를 선택적으로 적용하는 경향을 보이기 때문이다.

이와 같은 연구들을 종합적으로 고려하여, 디토 등(2009)은 '동기화된 도덕적 추론(motivated moral reasoning)'의 개념을 도출해내었다. 동기화된 도덕적 추론이란 특정한 도덕적 결론에 도달하려는 욕구에 의해 도덕 판단의 과정에서 작동하는 도덕적 추론이 동기화된 상태를 의미한다(Ditto et al., 2009: 312). 일반적으로 어떠한 도덕적 원칙은 자신의 도덕적 신념에 부합하는 방식으로 동기화되어 적용되기 때문이다. 어떤 행동이 도덕적 주체에게 부정적 정서를 불러일으키면, 그 주체는 행동을 통제할 수 있으며, 의도적이라고 생각할 수 있다. 또한 시간이 지남에 따라 행동 자체에 대한 부정적 기억을 갖게 되어, 그 행동을 한 사람을 도덕적으로 무책임하다고 비난할 수 있다. 도덕적 추론은 강한 방향성에 의해 동기부여 된다. 디토 연구진의 연구 결과에 미루어볼 때, 도덕적 추론이 기반하고 있는 도덕적 원리는 딜레마 상황의 설정값에 따라 유연하게 선택적으로 적용되었다.

도덕 판단을 도출해내거나 내려진 도덕 판단을 정당화하는 과정에서 사용되는 대표적인 도덕원리는 공리주의를 포함한 결과론과 칸트의 의무론이다. 동기화된 추론 연구자들은 의무론과 결과론을 토대로 도덕 판단

을 내리는 경우에도, 도덕적 추론은 자신의 신념을 보호하려는 동기가 작동하는 모습을 발견하였다. 이는 도덕적 추론의 과정이 동기적 요인에 의해 크게 영향을 받고 특히, 사람들이 자신의 결정을 정당화하기 위해 적용하는 도덕적 원칙이 매우 유연하다는 것을 의미한다(Ditto & Lopez, 1992; Kunda, 1987; Ehrich & Irwin, 2005;). 가령, 트롤리 딜레마 상황에서 의무론적 원칙에 의존하는 사람은 비록 그 행위가 다른 다섯 명을 구할지라도 무고한 개인은 도덕적으로 잘못되었다고 결론지어야 하는 반면, 결과주의 원칙에 의존하는 사람은 반대의 결론에 도달해야 한다.

이처럼 도덕적 추론의 양상은 정확성 목표에서는 비교적 약하게 동기화되며, 자신의 기존 신념에 부합하고자 하는 강한 방향성에 의해 동기화되는 모습을 보인다. 이는 동기화된 추론의 네 가지 유형 중 당파적 추론가 유형과 유사하다. 디토 연구진은 그들의 연구 결과가 하이트의 도덕 심리학 연구 결과와 부합하는 모습을 발견하였음을 밝히고 있다(Ditto et al., 2009). 앞서 하이트의 연구에서 확인한 것처럼, 도덕 판단은 본질적으로 직관적으로 도출되며, 따라서 이를 명확히 이해하기 위해서는 정서적이고 동기적 요인을 고려해야 한다. 하이트는 도덕적 추론이 도덕 판단에 뒤따른다고 주장하며, 추론이 도덕적 평가에 선행하고 도덕 판단을 도출해낸다는 콜버그의 합리주의적 관점에 반대한다(Haidt, 2001). 이는 흄의 도덕철학을 지지하는 경험적 연구로서, 도덕 판단은 일반적으로 직관적이고, 정서적인 과정을 통해 발생한다는 주장이다. 어떤 행동이 단지 잘못된 것이라고 '느꼈다면', 이 깨달음은 이성적인 추론이라기보다 미적인 판단에 가깝다고 할 수 있다. 이처럼 사회적 직관주의의 관점에서 도덕적 추론은 근본적으로 동기화된 추론이다. 도덕 판단은 정서의 개입이 결여된, 즉 데이터 그 자체에 대해 탁월한 이성의 능력을 발휘하여 추상적 원리로부터 구체적 사안에 적용하는 하향식으로 구동되는 것이 아니다. 오히려 도덕 판단은 전형적으로 구체적 사안들로부터 추상적 원리를 찾아가는 상향식으로 구동된다. '올바른' 도덕적 결론에 대한 직관을 가진 개인들이 논쟁을 통해 이를 더욱 확고히 하는 것이다. 이때 도덕적 추론은 정서적 선호도를 지원하는 방식으로 도덕 판단에 영

향을 미치면서도, 개인이 객관성을 드러낼 수 있도록 허용한다. 사람들은 '객관적인 판사'로서 자신의 관점에 도전하는 정도로만 데이터를 객관적이고 논리적으로 분석한다. 동기화된 추론의 효과는 타당성 제약이 느슨하고 모호한 상황에서 가장 두드러지게 나타난다(Ditto et al., 2009: 313). 이는 콜버그의 도덕 발달 연구에서 사용된 딜레마가 사람들로 하여금 '객관적인 판사'가 될 수 있는 상황을 조성하고 높은 수준의 타당성을 요구한 반면, 하이트의 도덕 판단 연구에서 사용된 이야기는 판단의 근거를 도출해내기 어려운 상황을 구성하는 것과도 일치한다. 이처럼 동기화된 도덕적 추론의 개념은 도덕 판단이 사회적 욕망이나 동기에 의해 영향을 받을 수 있고, 도덕 판단에 사용된 도덕 원칙이 다른 원인에 의해 합리화될 수 있음을 시사한다.

동기화된 추론에 대한 도덕 심리학적 접근의 거시적 동향과, 디토 등에 의해 수행된 동기화된 도덕적 추론에 관한 연구를 종합했을 때, 도덕 판단에 영향을 미치는 동기화된 추론의 특징은 다음과 같은 세 가지로 종합될 수 있다. 첫째, 도덕 판단에서 동기화된 추론의 인지 과정은 정서와 매우 밀접하게 결합 되어있다는 점이다. 인간의 정신 과정에서 이성의 합리적 사용으로 이해되어 오던 전통적인 합리주의적 견해는 정서와 독립된 추론의 과정을 규명하고자 하였다. 그러나 동기화된 추론의 개념은 추론의 과정이 정서적 영향으로부터 분리될 수 없음을 전제한다. 이는 fMRI 기법 등을 통해 뇌의 기능을 분석한 결과, 추론의 인지 과정이 정서적 반응과 복잡하게 얽혀있다는(entanglement) 최근의 신경과학 연구의 결과와도 상응한다(May & Kumar, 2018; May, 2021). 도덕 판단에서 추론의 인지 과정과 정서의 관계에 대해 주목한 메이는 도덕적 추론 자체가 정서와 기능적으로 분리될 수 없다고 주장한다. 그는 신경과학적 접근을 통해 도덕적 추론을 포함한 인지 과정이 행위자의 행동, 의도, 맥락과 관련된 규칙이나 규범을 추론하기 위한 뇌의 영역-일반(domain-general) 능력에서 발생한다는 것을 밝혀냈다(May, 2021: 1). 그는 복내측 전전두엽 피질(vmPFC)과 편도체(amygdala)에 주목한다. 복내측 전전두엽 피질은 일반적으로 복잡한 선택을 계획하고 결정하는 능

력을 담당하는 전두엽 피질의 안쪽 아래쪽에 자리 잡고 있으며, 이는 판단과 결정을 안내하는 직감을 수용하는 기능을 담당한다. 한편 변연계(limbic system)의 일부인 편도체는 감정, 특히 두려움과 강하게 연관되어 있다. 뇌의 이 두 부분을 중심으로 수행된 광범위한 도덕에 관한 신경과학 연구 결과물들은 도덕 판단에서 추론과 정서가 얽혀있음을 시사한다.

도덕적 추론은 단순히 추상적인 원칙을 적용하는 이론적 차원이 아니라 목표와 동기에 크게 영향을 받는다. 따라서 메이는 우리의 도덕 판단은 정서의 영향으로 인해 목표와 동기에 의한 희망적 사고, 확증 편향, 사후 합리화로 이어져 옳지 못한 결론을 도출할 수 있다고 보았다. 그럼에도 불구하고 문제가 있는 추론의 형태 또한 여전히 추론이다. 그는 추론[30]의 패러다임이 의식적 숙고에 국한하여 좁은 종류의 정신 현상으로 동일시하려는 것을 경계해야 한다고 주장한다. 그렇지 않으면, 우리는 도덕적 판단이 근본적으로 느리고 의식적인 숙고에 의해 완전히 정서로부터 벗어나있다는 어리석은 논제에 대한 입증의 책임을 도덕적 추론을 강조한 모든 이론가들에게 부당하게 떠넘기기 때문이다(May & Kumar, 2018: 140; May, 2021: 4). 동기화된 추론의 개념은 추론의 과정을 탈맥락적이거나 정서가 배제된 순수 형식 논리적인 차원의 차가운 인지 과정으로 제한하지 않으며, 사회적 상호작용의 맥락에서 자연스럽게 발생하는 정서에 강하게 결합되어 있음을 시사한다.

둘째, 도덕 판단에서 동기화된 추론의 기본적 동기는 자기 본위적인 것에 있다. 도덕 판단에서 추론의 인지 과정이 정서의 영향에 노출되는 까닭은 도덕적 추론에 의해 도달하고자 하는 목표가 자기의 도덕적 신념을 보존하고자 하는 동기에 있기 때문이다. 동기화된 추론 연구의 모태

30) 메이(J. May)는 기존의 심리학 연구들에서 reasoning과 inference 개념을 엄격하게 구분하지 않고 있으며 상호교환적으로 사용되고 있다고 밝혔다(2022. 8. 16. 개인 서신교환). 그럼에도 그는 기존의 합리주의 관점에서 추론(reasoning)의 의미를 너무 좁게 해석하고 있다는 문제 의식에 입각하여 넓은 의미의 관점을 취하고 있다. 이러한 메이의 문제의식과 용어 사용은 추론을 추리의 하위 범주로 방식으로 접근하고 있는 본고의 입장과 유사하다. 이러한 용어 사용은 후술 될 메르시에(H. Mercier)와 스퍼버(D. Sperber)나 모쉬맨(D. Moshman)의 용어 사용과도 같은 맥락으로 이해된다.

가 되는 인지부조화 이론에 따르면, 인지 요소들 간의 충돌이 발생할 때 자기 개념(self-concept)을 보존하고자 하는 동기가 강할수록(Thibodeau & Aronson, 1992), 긍정적 자아상(self-image)을 보존하고자 하는 동기가 강할수록(Mezulis et al., 2004), 자존감을 지켜내고자 하는 욕구가 강할수록(Pyszczynski & Greenberg, 1987) 왜곡된 방식으로 환경을 해석하고자 한다고 보고한다.

도덕 판단에서 동기화된 추론에 영향을 끼치는 정서적 요인이 기본적으로 자기 본위적이라는 사실은 자신이 속한 사회적 관계 속에서 유지하고자 하는 도덕적 자아상의 모습에서 비롯된다. 타인과 세계에 대한 판단은 동기의 영향을 받는다(Ditto et al, 2009: 310). 사람들은 싫어하는 사람들보다 좋아하는 사람들의 행동에 대해 자선과 같은 더 많은 도덕적 행동을 하고(Reeder et al, 2005), 사회적으로 인정되는 도덕적 태도에 관한 증거를 그러한 태도에 도전하는 증거보다 더 효과적이고 설득력 있는 것으로 인식한다(Munro & Ditto, 1997, Ditto et al, 2009: 311에서 재인용). 나아가 동기화된 도덕적 추론 연구자들은 우리가 하나의 도덕적 신념을 채택하는 것이 특정한 도덕적 결과에 도달할 가능성을 높이는 현상에 주목한다(Cusimano & Lombrozo, 2021). 상황에 대한 잘못된 판단을 통해 자신의 기존 도덕적 신념을 현실화하는 방향으로 도덕 판단을 도출하기 때문이다. 특히, 자충적 신념(self-fulfilling belief)은 신념을 채택하는 도덕적 이유와 증거적 이유를 혼동하게 한다. 이로 인해 도덕 판단에서 동기화된 추론은 자신의 도덕적 신념에 위협을 가하는 정보나 상황을 통해 신념을 수정하기보다는, 이를 왜곡함으로써 오히려 자신의 이익에 부합하는 방식으로 도덕적 결론에 도달하도록 한다.

끝으로, 도덕 판단에서 동기화된 추론은 마음의 이중 처리 이론에서 구분하는 시스템 1과 시스템 2 사고 유형의 두 특징을 모두 공유한다. 원하는 결론에 도달하려는 동기는 인지 과정의 작동을 촉발하는데, 이는 의식적으로도 무의식적으로도 나타나기 때문이다. 동기는 접근 가능한 지식의 구조(신념, 기억, 정보를 이해하는 스키마)와 사용되는 인지 과정에 영향을 줌으로써 추론에 영향을 미친다. 많은 이론가는 인간의 사고

유형을 시스템 1과 시스템 2 추리로 유형화한다(Evans, 2002; Kahneman, 2003; Stanovich, 2004). 시스템1 추리 과정은 자동적이고 직관적으로 빠르게 이루어는 사고 메커니즘으로, 주로 무의식적으로 작동하여 이를 통제하거나 알아차리기 어려우며, 인지적 노력이 거의 수반되지 않는다. 시스템 1 추리는 전체적이고, 상황과 맥락에 따라 유연하게 작동하며, 휴리스틱(heuristic, 어림짐작 ; 추단법)적으로 작동한다. 반면, 시스템 2의 추리는 논리적이고, 숙고적이며 상대적으로 느리게 이루어지는 사고 메커니즘으로, 주로 의식적으로 작동하며, 이를 위해 일정한 인지적 자원의 소모가 요구된다. 시스템 2의 추리는 규칙의 논리나 원칙의 엄격한 적용을 포함하며, 분석적(analytic)으로 작동한다. 도덕적 상황에 대한 감정적 반응은 주체가 기반하고 있는 사회적, 문화적, 진화적으로 축적된 도덕적 토대 또는 만성적으로 접근 가능한 도덕 스키마에 기반하기 때문에 동기화된 추론은 즉각적이고 무의식적으로 나타날 수 있다. 그러나 도덕 판단에서 동기화된 추론은 때로는 의도적이고 의식적인 숙고의 과정을 거쳐 나타나기도 한다. 하이트의 연구에서 다수의 실험 참가자는 동기화된 추론을 통해 자신의 도덕 판단에 대한 정당화를 제공하기 위해 다양한 근거들을 수집하고, 인지 전략을 사용하기도 하였다.

동기화된 추론의 개념을 처음으로 도입한 쿤다(Kunda)가 초기 연구에서는 동기화된 추리(motivated inference)라는 개념을 사용했던 것 또한 이러한 맥락에서 이해할 수 있다. 최초의 연구는 동기화된 추론이 시스템 1의 휴리스틱의 일종으로 확인되었으나, 연구가 거듭되고 다양한 동기화된 추론 관련 연구 업적이 쌓이면서, 동기화된 추론의 모습이 시스템 2의 분석적인 과정을 거치기도 하고, 이 둘의 상호작용의 결과로도 나타난다는 것을 발견할 수 있었다. 이처럼 이중 처리 이론은 인간의 사고 과정을 범주화하여 이해하는데 탁월한 설명력을 지니지만, 시스템 1과 시스템 2로 이분하여 사고의 메커니즘을 이해하게 되면, 추론의 메커니즘을 온전히 담아내기 어렵다는 문제가 있다(Moshman, 2004b; Mercier & Sperber, 2017; Mercier, 2017). 따라서 도덕 판단에서 동기화된 추론의 문제를 정확히 이해하고, 이를 근본적으로 개선하기 위한 방

안을 탐색하기 위해서는 추론의 메커니즘에 대한 새로운 이해의 틀이 필요하다.

이처럼 도덕 판단에서 동기화된 추론의 개념은 추론의 과정이 정서와 강하게 결합해 있으며, 그러한 정서적 동기가 주로 자기 본위적인 것에 있다는 특징을 지니고, 그 과정이 이중 처리 이론의 두 가지 유형의 사고 중 어느 하나로 범주화하기 어렵다는 특징을 보여준다. 도덕 판단에서 동기화된 추론의 세 가지 특징은, 명시적으로 '동기화된 도덕적 추론'이라는 개념을 상정하지 않은 여러 연구에서 간접적으로 이를 확인할 수 있는 기준으로 활용될 것이다.

제 2 절 도덕 판단에서 동기화된 추론으로 인한 문제

1. 비도덕적 행위의 정당화

동기화된 추론 연구는 사람들이 객관성의 환상(Kunda, 1990)을 유지하면서 자신이 선호하는 결론에 도달할 가능성이 높다는 것을 시사한다. 이로 인해 우리는 자신의 도덕적 신념이나 행동에 부합하는 해석을 뒷받침하는 증거와 규칙을 찾도록 동기 부여받으며, 추론을 통해 논증을 구성하고, 이를 객관적이라고 믿는다. 나아가 동기화된 도덕적 추론 연구들에 따르면, 도덕 판단에서 동기화된 추론의 양상은 도덕성의 환상을 유지하도록 해준다(Cusimano & Lombrozo, 2020; 2021; Gino et al., 2016; Uhlmann et al., 2009). 즉, 동기화된 추론은 자신이 도덕적이지 못한 행동이나 판단을 내리더라도 여전히 자신을 도덕적 주체로 생각하게 만든다. 이는 도덕 판단에서 동기화된 추론이 비도덕적 행위의 정당화 문제를 유발함을 시사한다.

동기화된 추론 개념이 주로 다뤄지는 연구 분야인 사회 심리학 영역에서는 동기화된 추론으로 인한 비도덕적 행위의 정당화 문제를 도덕적 위선(moral hypocrisy)의 개념으로 설명한다(Batson, 2017; Batson &

Thompson, 2001; Collins & Batson, 2007). 도덕적 위선은 “도덕적으로 행동함으로써 발생하는 비용을 피하면서 가능한 도덕적으로 보이려는 동기”를 의미한다(Collins & Batson, 2007: 589). 이는 도덕적 원칙에 따라 행동하고자 하는 진실한 동기로서 도덕적 자기충실성(moral integrity)과 상반되는 마음의 상태이다. 뱃슨 연구진은 다양한 실험을 통해 도덕적 위선의 모습을 발견한다. 가장 대표적인 연구는 동전 던지기 실험이다(Batson & Thompson, 2001). 이들은 대학생들에게 익명으로 자신과 동일한 성별의 파트너(실제로는 허구의 인물)에게 두 개의 서로 다른 과제를 분배하도록 요청했다. 일반적으로 약 80%의 실험 참가자는 두 개의 과제 중 분명히 다른 과제보다 더 좋은 과제(쉽거나 보상이 큰 과제)를 자신에게 주었다. 이때, 연구진은 학생들에게 공정성이라는 도덕 원칙을 상기시키면서 과제를 결정하기 위해 동전 던지기를 할 수 있으며, 그 방식은 자유롭게 정할 수 있다고 설명했다. 또한 과제를 배정받은 파트너 그 결정이 우연히 내려진 결정으로 알릴 것이라고 말했다. 80%의 실험 참가자는 동전을 던지는 선택을 했다. 그러나 놀랍게도 동전을 던진 사람의 90%가 자신에게 좋은 과제를 배정했다. 도덕적 위선과 관련하여 주목해야 하는 부분은 과제 분배 이후 수행된 자기 평가의 결과이다. 동전 던지기를 했음에도 자신에게 더 좋은 과제를 배정한 실험 참가자들은(M= 7.1) 동전 던지기의 속임수를 거치지 않고 자신에게 더 좋은 과제를 배정한 사람들(M=4.0)보다 그들의 행동을 더 도덕적이라고 평가한 것이다. 공정함(자신에게 좋은 과제를 부여할 가능성과 상대방에게 부여할 가능성을 50%로 만드는 것)의 비용(자신에게 좋은 과제를 부여하지 못하는 경우)을 피하면서 공정한 것처럼 보이려는 모습(동전을 뒤집는 행위)은 도덕적 위선의 증거라고 볼 수 있다. 뱃슨 연구진은 이와 같은 실험을 10여 년 동안 반복적으로 수행하였다. 동일한 방식으로 다른 대상에게 진행한 연구나, 약간의 조건을 변형한 연구에서 모두 도덕적 위선의 모습을 발견할 수 있었다(Batson, 2017). 이와 같은 일련의 실험을 통해 뱃슨은 도덕적 인간의 비도덕적 행위를 이해하기 위해서는 그 동기의 본질인 도덕적 위선의 측면에 주목해야 함을 강조한다.

동기화된 추론 연구는 사람들이 이기적이고 비도덕적인 행동을 하면서도 도덕적으로 행동하는 것처럼 느끼도록 합리화하고 정당화하는 도덕적 위선의 심리 기제를 설명한다(Gino et al., 2016; Valdesolo & DeSteno, 2007). 동기화된 추론으로 인한 편향된 도덕 판단은 자신이 저지른 도덕적 위반이 다른 사람이 저지른 동일한 위반보다 덜 불쾌하다고 평가하는 도덕적 위선을 정당화해준다(Valdesolo & DeSteno, 2007). 사람들은 종종 자신이 하는 일에 대해 도덕적으로 '잘못되었다'는 느낌 없이, 부정직하거나 비도덕적으로 행동하기 위해 의사결정 환경에 존재하는 정당성과 불확실성을 이용하는 경향을 보인다(Gino et al., 2016). 즉, 사람들은 자기행동의 의도나 결과가 지닌 도덕적 가치보다는 자신이 취하는 행동이 도덕적 존재로서의 자신에 대한 평가에 일관성을 유지하는 것에 더 관심이 있으며, 이러한 목표에 따라 도덕 판단에서 동기화된 추론이 작동하는 것이다. 우리는 스스로를 도덕적이라고 믿고 싶어하며, 때로는 자신의 행동이 실제로 도덕적인지 아닌지와는 별개로 이 신념을 지지하는 행동을 도덕적인 것이라고 생각한다. 이러한 신념을 유지하도록 기여하는데 사람들은 무엇이 (비)도덕적인지에 대한 단서나 정보를 자기방어의 방식으로 처리한다(Gino et al., 2016: 190). 지노 등은 도덕 판단에서 동기화된 추론의 작동 메커니즘을 '동기화된 베이지안 추론' 모델로 설명한다. 베이지안 추론은 통계적 추론의 한 방법으로 추론 대상에 대한 사전 확률과 이와 관련된 추가적인 증거 등 새롭게 획득한 데이터를 기반으로 해당 대상의 사후 확률을 추론하는 방법이다. 즉, 사람들은 의사결정 과정 전에 자신의 신념에 관한 정보를 수집하여 추론하며 자신의 이익을 추구하면서도 자신의 행동이 도덕적이고 정직하며, 공정하다고 느끼도록 돕는 방향으로 정보를 수집하는 경향이 있다는 것이다. 즉, 기존의 베이지안 추론이 의사결정에 가장 유익한 증거를 찾고 편향되지 않은 방법으로 그것을 처리하는 반면, 도덕판단에서 동기화된 베이지안 추론은 자신을 도덕적인 행위자로서의 지위를 유지하기 위한 방향으로 새로운 증거를 수집하고, 정보를 편향되게 해석하는 것이다.

동기화된 베이지안 추론은 도덕적 위선의 과정, 즉 도덕 판단에서 동

기화된 추론이 비도덕적 행위를 정당화하는 과정을 두 가지 방식으로 설명해준다. 하나는 사람들이 스스로 도덕적 혹은 정직한 행위나 그 결과를 구성하는 것이 정확히 무엇을 지칭하는지에 대해 동기화된 추론이 사용된다는 것이다(Gino et al., 2016: 195). 즉, '옳음과 그름' '도덕과 비도덕' 등의 의미를 해석하는데 있어 기준을 유연하게 적용한다. 이때, 행위의 도덕성에 대한 판단은 주로 자신의 도덕적 지위를 유지하기 위한 방향으로 동기화된다. 다른 하나는 긍정적인 도덕적 자아상에 도달하기 위해 관련된 정보와 증거 자체에 대한 인식을 바꾸는 것이다. 이 방식은 도덕과 무관한 판단에서 많은 사람들이 자신의 특성과 능력을 평가할 때 일반적으로 나타나는 특징이다. 사람들은 자신의 지능과 매력과 관련된 긍정적인 정보를 과대평가하고, 부정적인 정보를 과소평가한다(Gino et al., 2016: 199)[31]. 도덕의 영역에서도 유사한 양상이 나타나는데, 이는 자신의 비도덕적 행위를 덜 나쁜 것으로 보이게 한다. 이와 같은 두 가지 방식은 사람들이 도덕 판단과 관련한 추가적인 정보와 증거를 자기에게 유리한 방식으로 해석하여 비도덕적인 행동을 조장한다.

한편, 도덕 판단에서 동기화된 추론을 통한 비도덕적 행위의 정당화 문제는 도덕적 이탈(moral disengagement)[32]의 메커니즘을 통해서도 체

31) 가령 사람들이 IQ 테스트 결과에 비례하여 높은 보상을 얻게 될 때, IQ 테스트를 위한 시험지 하단에 정답이 인쇄된 경우에서 그렇지 않은 경우보다 통계적으로 유의미하게 높은 IQ를 보였고 더 많은 보상을 획득했다(Chance et al., 2011). 심지어 부정행위로 더 높은 IQ를 얻은 사람들은 하단에 정답이 인쇄되지 않았던 두 번째 테스트의 결과를 예측하라는 물음에 자신의 타고난 지능을 근거로 높은 예측 점수를 부여했다. 두 번째 테스트의 결과는 자신의 사전 예측에 비례하여 보상을 해주었음에도 이들은 여전히 자신의 IQ를 높게 예측했다. 또한 각 테스트의 결과를 알 수 없게 한 상태에서 자신의 IQ를 예측하도록 했을 때 첫 테스트에서 하단에 정답이 인쇄된 문제를 푼 실험 참가자들은 자신의 부풀려진 능력에 대한 교정을 하는데 매우 더뎠다(Chance et al., 2015).

32) 국내 문헌에서 반두라의 'moral disengagement' 개념은 도덕(적) 이탈(정창우, 2020a; 추병완 2020) 혹은 도덕성의 분리(김의철 등 역, 2015/2018)으로 번역되었다. 반두라는 이 개념을 통해 도덕적 주체가 비도덕적 행위를 하면서도 자존감을 지켜낼 수 있다고 보았다. 자신의 도덕적 기준으로부터 이탈(disengagement)하는 현상을 설명한다는 점에서 도덕(적) 이탈이라는 번역은 타당하다. 한편, 이 개념은 자신의 행위와 그 행위에 내재된 도덕적 가치를 제거하는 양상으로 나타난다는 점에서 도덕성 분리 현상으로도 번역할 수 있다. 두 번역어 모두 반두라의 이론을 적절히 담아내고 있기 때문에, 본 논문에서는 도덕 교육계에서 상용되고 있

계적으로 설명된다. 사회인지이론가인 반두라에 의해 제시된 개념인 도덕적 이탈의 메커니즘은 자신에 대한 도덕적 평가를 수행하는 도덕적 자기 제재가 선택적으로 활성화되어 비도덕적인 행위를 하면서도 여전히 자신을 도덕적 행위자(moral agency)로서의 지위를 유지하도록 하는 심리적 과정을 지칭한다(Bandura, 2016/2018).

반두라는 기존의 도덕 심리학이 의도, 신념, 책임의 개념을 각각 분리하여 하나에만 초점을 두는 방식으로 연구해왔던 것을 지적하며 사회인지 이론의 관점에서 도덕적 '주체(agency)'에 초점을 둘 때 도덕성을 온전히 이해할 수 있다고 주장한다(Bandura, 2016/2018). 사회인지이론은 자아에 대한 신념이 행동을 발생시키는 역할에 주목한다. 즉, 자기 효능감과 같은 자신에 대한 신념은 생각과 행동을 연결하는 자기 조절의 과정을 통해 나타난다. 특히, 자기 개념에 중심적인 역할을 차지하는 도덕적 기준을 바탕으로 이루어지는 자기 제재는 도덕적 행위를 일으키는 주요 동인이다. 자신의 도덕적 기준을 위반하면서도 자존감(self-esteem)에 손상 없이 만족해하는 모순적인 행위를 설명하는 심리적 기제는 단지 도덕적 기준을 엄격하게 지키는 것을 설명하는 것보다 복잡한데, 반두라는 이러한 비도덕적 행위의 정당화 방식을 도덕적 이탈의 메커니즘이라는 개념을 통해 설명하고자 한다(Bandura, 2016/2018).

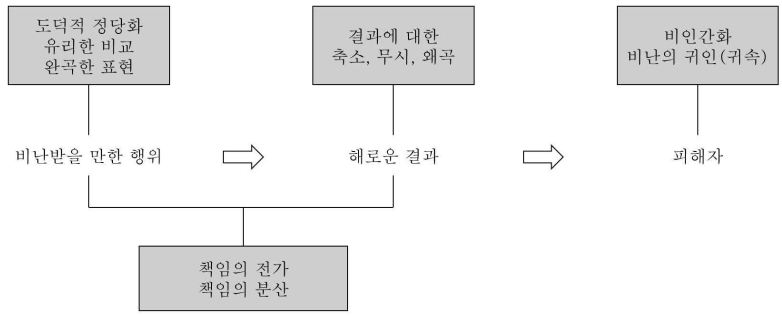
비도덕적인 행위를 하거나, 그러한 행위를 목격했을 때 도덕적 기준은 활성화되며 죄책감과 양심의 형태로 비도덕적 행위자를 비난한다. 이처럼 행위의 옳고 그름에 대한 도덕적 기준은 자기 조절을 통해 도덕적 행위를 하거나, 비도덕적 행위를 제어한다. 이때 도덕적 기준의 적용은 일관적으로 적용되지 않는다. 일상의 도덕적 문제 상황에서 주체는 자신에게 유리하도록 상황을 해석하기도 하고, 자신의 자존감을 유지하고자 하는 강한 방향성을 가지고 있기 때문에 도덕적 기준을 선택적으로 적용하는 모습을 보인다. 길거리의 모퉁이가 쓰레기 더미로 가득한 상황을 가정해보자. 쓰레기가 가득한 길거리는 자연적으로 생긴 것이 아니라 사람들이 쓰레기를 버린 행위로 생겼을 것이다. 그러나 많은 사람의 도덕적

는 도덕(적) 이탈을 기본적으로 따르되, 문맥에 따라 도덕성 분리라는 표현도 병용하였다.

기준은 공공장소에 쓰레기를 버리는 행위를 도덕적으로 비난받을 행위로 판단할 것이다. 이 경우, 의식적이든, 무의식적이든 쓰레기를 버린 사람들은 자신의 도덕적 기준을 어겼기 때문에 자책감을 느끼거나 괴로움을 느낄까? 반두라는 대부분의 경우, 그렇지 않을 것이라고 예측한다. 사람들은 자신이 쓰레기를 버리는 행위를 판단하는 데 도덕적 기준을 선택적으로 활성화함으로써 행위와 도덕적 가치를 분리하기 때문이다. '(쓰레기장이 아니라는 것을 알면서도) 이곳은 원래 쓰레기를 버리는 곳이야'라고 생각하거나, '다른 사람들도 다 버리는데', 혹은 '난 그렇게 더러운 쓰레기를 버리진 않았어' 등 다양한 방식으로 자신의 행위로부터 도덕적 기준을 이탈시킬 수 있다. 이 과정을 통해 사람들은 비도덕적인 행위를 하면서도 자책이나 괴로움으로부터 벗어날 수 있고 도덕적 주체로서의 지위를 유지할 수 있게 된다. 즉, 인간이 비도덕적인 행위를 하고도 도덕적 주체로서 지위를 잃지 않는 것은 오히려 도덕적 기준을 자신의 행동을 분리해 그에 대한 자기 제재의 책임을 면제받고자 하는 동기에서 비롯되는 것이다(Bandura, 2016/2018: 9). 이처럼 자기 제재는 도덕적 행위의 강한 동기이다. 자기 제재는 자신의 도덕적 기준에 부합하는 행동을 하도록 유도한다. 그러나 이 기준은 높은 정확성을 목표로하지 않는다. 즉, 이를 통해 우리는 진실에 도달하고자 하기보다는, 자신이 '도달하고자 하는 결론' 즉 도덕적인 주체로서의 지위를 잃지 않고자 하는 방향에 의해 강하게 동기화되는 모습이다. 이러한 방향성 목표에 의해 동기화된 추론은 도덕적 이탈의 메커니즘을 통해 자신의 행위 결과를 축소하거나 왜곡하고, 유리하게 비교하거나 자기중심적으로 재해석하며, 때로는 책임을 전가·분산하는 등의 방식을 보이는 것과 같이 다소 편향적인 방식으로 상황을 해석함으로써 비도덕적 행위의 정당화 문제를 유발한다.

도덕적 이탈의 메커니즘은 도덕적 자기 조절 과정의 네 지점(four locus)에서 발생한다. 각각은 자신이 저지른 비난받을 만한 행위의 측면(behavioral locus), 유해한 행위를 저지른 당사자로서 주체의 측면(agency locus), 자신의 행위로 인한 해로운 결과의 측면(effects locus), 해로운 영향으로 인한 피해자 측면(victim locus)이다(Bandura,

2016/2018). 다음의 그림은 각 지점에서 나타나는 도덕적 이탈의 메커니즘을 여덟 가지로 세분화한 것이며 각각의 메커니즘은 동기화된 추론 개념으로 설명된다.



[그림 3-2] 도덕적 자기제재가 선택적으로 활성화되는 메커니즘(Bandura, 2002: 103)

도덕적 이탈은 비난받을 만한 행위의 측면에서 ①도덕적, 사회적, 경제적으로 정당화하고, ②자신에게 유리하게 비교하고, ③완곡한 표현을 사용함으로써 도덕적 이탈이 일어난다. 이 세 가지 메커니즘은 행위의 인지적 재구성을 촉진하여 비도덕적 행위에 가담한 개인에게 덜 해롭게 보이도록 한다. 동기화된 추론은 비도덕적인 행위를 어떤 면에서 유익한 것처럼 보이게 함으로써 작동한다.

먼저 ①도덕적(사회적, 경제적) 정당화(moral, social, economic justification)이다. 비도덕적 행위를 하는 경우, 일반적으로 사람들은 스스로의 행동에 다양한 방식으로 정당성을 부여함으로써 도덕적 주체로서의 지위를 유지한다. 특히, 도덕성의 측면에서 자신의 비난받을 만한 행위를 정당화함으로써 자신이 앞으로 수행할 행동이 도덕적으로 명예로운 목적을 지향하고, 사회적 가치를 위해 스스로 헌신하기 위한 것처럼 위장하는 것이다. 반두라는 이 양상의 대표적인 사례로 군사적 임무를 수행하는 군인들에게서 찾는다. 그는 앨빈 요크 병장 에피소드, 십자군 전쟁, 이슬람 극단주의자의 테러 행위 등의 사례를 제시하며 전쟁 당사자들이 자신의 행위를 도덕적으로 정당화하는 방식의 도덕적 이탈 메커니즘을 보여준다(Bandura, 2016/2018). 군인들은 군사적인 임무를 수행하는

과정에서 불가피하게 살상을 하게 된다. 이들은 살상(殺傷)의 의미를 도덕적(사회적, 경제적)으로 재규정함으로써 자신의 군사적 행동을 정당화하였다. 자신의 행위로 인해 누군가의 생명에 위협을 가하게 됨에도 불구하고, 그 행위의 명분을 되새기며 스스로를 설득하는 것이다. 또한 그들은 적국의 군사적 행위에는 강한 도덕적 기준을 적용함으로써 그들에 대한 자신들의 도덕적 우월성을 확보하기도 한다. 이처럼 사람들은 자신이 저지른 비난받을 만한 행위에 도덕적 가치를 부여하는 방식으로 정당화함으로써 비도덕적인 행위를 하면서도 여전히 도덕적 지위자로서의 자존감을 유지하는 것이다. 이는 디토 등에 의해 수행된 동기화된 도덕적 추론 연구에서 도덕적 원리를 선택적으로 적용함으로써 타인의 행위 '의도성'에 대해 서로 다른 방식으로 해석했던 결과(Ditto et al., 2009)와 일치한다.

다음은 동기화된 추론을 통해 ②자신에 대한 유리한 비교(palliative comparison)를 함으로써 비도덕적 행위를 정당화하는 과정이다. 동기화된 추론은 비난받을 만한 행위를 보다 다른 사람의 더 심각한 비인간적 행위와 비교함으로써 책임을 회피하고자 한다. 유리한 비교를 통해 사람들은 자신의 악행을 왜곡하여 해석함으로써 자신의 비도덕적 행위를 사전적으로 정당화한다. 반두라는 편법적인 역사 왜곡의 사례를 들어 유리한 비교의 과정을 설명한다. 가령, 자신이 지지하는 정치인의 위법 사실을 변호하기 위해 반대 측 정당의 정치인들이 관여한 역대 위법 관련 증거들을 제시하는 경우이다. 이 경우 양측 모두 비도덕적 행위인 것이 자명하지만, 다른 사람과의 비교를 통해 자신의 비도덕적 행위에 대한 책임을 면제하고자 하였다. 또한 베트남 전쟁 당시 미군의 군사 개입 과정에서도 유리한 비교를 찾아볼 수 있다. 당시 수많은 인명 피해가 동반되었음에도 미국 정부는 공산주의 정권에 예속된 민중들을 구하기 위한 것이라는 명분을 바탕으로 군사 개입으로 인한 보편적 인권의 향상과 불가피한 인명 피해를 비교함으로써 비도덕적 행위를 합리화하였다(정창우, 2021: 101). 반두라는 이 과정에서 도덕적으로 폭력이 용인되는 경우를 실용주의의 관점에서 두 가지 제시한다. 하나는 비폭력적인 방식으로 행

위 했을 때, 원하는 결과에 도달하지 못하는 경우 이는 고려 사항에서 제외되는 경우이며 다른 하나는 실용적 분석을 통해 비도덕적 행위로 인해 발생할 피해보다, 이로 인한 고통의 예방적 가능성이 크다고 판단하는 경우이다. 이러한 방식을 통해 사람들은 비난받을 만한 행위를 하면서도 더 심각한 유해행위와 비교함으로써 자신의 유해성을 중화시키고, 행위에 대한 면책을 주장하는 것이다. 이 과정에서도 역시 동기화된 추론이 작동한다. 추론이 정확성 목표에 의해 동기부여되지 않고 단지 도덕적 주체로서의 지위를 유지하고자 하는 방향성 목표에만 동기부여되는 경우, 상황이나 정보를 객관적으로 인식하지 못하고 자신이 도달하고자 하는 목표에 부합하는 방식으로 왜곡하여 인식하기 때문이다.

③완곡한 언어 표현(euphemistic labelling) 또한 비난받을 만한 행위의 측면에서 발생하는 동기화된 추론을 통한 비도덕적 행위의 정당화 양상이다. 언어의 사용은 인간의 사고방식에 영향을 받기도, 영향을 미치기도 한다. 자신의 행동을 어떠한 언어로 표상했느냐에 따라 그 행동의 가치 표현이 달라진다. 사람들은 완곡한 언어의 사용을 통해 비도덕적 행위로부터 행위자를 분리시킴으로써 도덕적 기준의 적용을 피해가도록 한다. 반두라는 비도덕적 행위를 직접 표현하는 언어 대신 이를 탈가치적인 언어로 바꿈으로써 도덕성을 자신의 행위로부터 분리한다고 본다. 가령, 대부분의 군사작전에서 잔혹하고 파괴적인 행위는 '목표 수행'이라는 행정적인 단어로 완곡하게 표현함으로써 자신의 비도덕적 행위로부터 도덕적 기준을 분리시킨다. 이를 통해 자신이 행할 행위에 대해 스스로를 도덕적으로 납득시키게 되는 것이다. 또한 유해한 상품의 상업 광고는 완곡한 언어 표현을 통해 동기화된 추론이 비도덕적 행위를 정당화하는 모습을 보여주는 대표적인 사례이다. 건강에 유해하고, 특히 아동과 청소년의 모방 가능성에 취약한 술, 담배 등의 경우 그 유해성을 강조하기보다는 다양한 좋은 이미지를 내세우며 가치 중립적인 것으로 보이게 만든다. 멋진 배우가 고된 하루 끝에 담배를 찾는 모습을 보여주거나, 축하하는 자리에서 술이 빠질 수 없다는 식의 광고 문구는 시청자들이 술이나 담배의 유해성을 생각하기 어렵게 만든다.

지금까지 살펴본 비난받을 만한 행위의 측면에서 발생하는 세 가지 도덕적 이탈의 메커니즘에서 동기화된 추론이 비도덕적 행위를 정당화하는 문제는 우크라이나 침공 사태 당시 러시아 대통령 푸틴의 담화문에서도 확인할 수 있다.

> *우크라이나는 러시아의 역사, 문화, 종교의 필수적인 일부 중 하나였습니다. 이들은 우리의 친구이며 혈연관계로 이루어진 친척이나 다름없습니다. (중략) 동시에 우크라이나 당국은 국가를 건설하기 시작했던 첫 단계부터 역사를 왜곡하려 했습니다. 수천만의 사람이 우크라이나에 살고 있고, 당연히 우크라이나는 극단적인 민족주의에 직면했습니다. 이는 빠르게 공격적인 공러증과 네오 나치즘의 형태를 가졌습니다. 북캅카스의 테러 집단에서 우크라이나 민족주의자들과 네오 나치들이 발견되는 것은 이상한 일이 아닐 것입니다. (중략) 우크라이나에 대량 살상 무기가 등장함에 따라서 세계와 유럽, 특히 러시아와 우크라이나의 상황이 가장 급진적으로 바뀔 것입니다. 우리는 이 실질적 위협에 대응하지 않을 수 없습니다. 특히 서방의 후원자들이 우크라이나에 또 다른 파란을 일으키기 위해 그러한 무기가 등장할 수 있게 하는 데 기여한다는 점을 다시 강조합니다. (후략) - 2022.2.21. 푸틴 담화문 中*

위 담화문에서처럼 러시아는 우크라이나 침공에 대한 도덕적 정당화를 통해 자신의 군사적 행위를 옹호하고 있다. 오히려 유리한 비교를 통해 우크라이나의 야만적 행위나, 우크라이나를 지원하는 국가들의 신무기 도입에 대해 도덕적으로 비난을 하는 모습도 나타난다. 또한 러시아와 우크라이나의 관계를 '친척'으로 완곡하게 표현함으로써 비도덕적 행위를 합리화하는 모습도 드러난다. 이처럼 자신이 저지른 비도덕적 행위에 도덕적 가치를 부여하는 방식으로 정당화하고, 자신의 행위를 더 잔혹한 다른 행위와 유리하게 비교하고, 완곡한 표현을 통해 정교하게 도덕적 기준에서 이탈함으로써 비도덕적인 행위를 하면서도 여전히 도덕적 주체로서 지위와 자존감을 유지하는 것이다. 이는 동기화된 도덕적 추론 연구에서 도덕적 원리를 선택적으로 적용함으로써 타인의 행위 '의도성'에 대해 서로 다른 방식으로 해석했던 결과와도 상응한다.

다음은 두 번째 측면으로, 행위자의 측면에서 발생하는 두 가지 도덕

적 이탈의 메커니즘이다. 두 메커니즘은 ④책임 전가(displacement of responsibility)와 ⑤책임 분산(diffusion of responsibility)이다. 비난받을 만한 행위를 저지른 당사자가 본인임에도 불구하고, 그 행위의 책임을 다른 사람에게 전가하거나 분산하는 방식으로 자신의 행위로부터 도덕적 기준을 분리시키는 것이다. 이러한 인지 메커니즘은 자신의 행동으로 인해 발생하는 피해에서 개인의 역할을 최소화한다. 책임을 전가하거나 분산하는 이러한 인지는 자신의 행동으로 인해 발생할 피해를 암묵적으로 부인하거나, 명시적으로 지시했을 수 있는 권위 있는 인물에게 그 책임을 돌리는 경향으로 동기화된 추론을 보인다.

권위 복종 실험으로 유명한 밀그램의 연구 결과(Milgram, 1974)는 도덕적 이탈의 메커니즘으로서 책임의 전가를 보여주는 전형적인 사례이다. 사람들은 자신의 행위가 다른 사람에게 고통을 주는 행위임을 알면서도, 그 행위는 연구자의 명령에 의한 것이므로 그 책임의 소지가 행위자 자신이 아님을 강조한다. 이를 통해 자신이 저지른 행위임에도 책임을 지지 않고자 하며, 또한 행위의 실질적인 주체가 본인이 아니기 때문에 자기 비난으로부터도 면제받는다(정창우, 2020a: 103). 특히, 밀그램의 연구는 해악을 유발하는 명령을 내린 명령권자의 권위와 친밀함과도 강한 양적 상관관계를 보였다.

도덕적 주체는 때로는 책임의 소지를 다양하게 분산시킴으로써 도덕적 기준으로부터 분리한다. 특히 집단적 수준에서 비도덕적 행위가 산출될 때, 행위자 개개인은 그러한 행위의 책임을 다른 사람들에게 분산시킴으로써 책임의 소지를 모호하게 만든다. 가령, 과속의 사례에서도 책임 분산의 모습이 발견된다. 자신의 과속 행위는 명백한 도로 규칙 위반임에도 불구하고, 다른 모든 사람들이 과속을 한다는 사실을 근거로 자신의 잘못된 행위를 다른 사람들에게 분산시킴으로써 책임을 분산시킨 것이다. 이처럼 동기화된 추론은 사후적인 정당화뿐 아니라, 자신의 비도덕적 행위로 인해 발생할 여러 상황에 대해 예견적으로 정당화하는 역할을 수행하기도 한다.

세 번째 측면은, 자신의 행위로 인해 발생한 해로운 결과의 측면으로

⑥결과의 축소, 무시, 왜곡(minimizing, ignoring or misconstruing the consequences)을 통한 도덕적 이탈의 양상이다. 사람들은 유해한 결과를 초래하는 선택을 할 때, 그로 인해 발생할 피해를 회피하거나 최소화하기 위해 노력한다. 행위의 결과가 무해한 것으로 판명이 나는 경우 도덕적으로 면책이 된다고 느끼기 때문이다. 이처럼 행위의 결과가 축소, 무시, 왜곡되는 경우 자기 비난이 일어날 가능성이 거의 없다(정창우, 2020a: 104). 특히 해로운 결과의 측면은 행위로 인한 결과가 비가시적일 때 더욱 활성화된다. 타인의 고통을 직접적으로 확인할 수 없거나, 그 행위로 인한 파괴적인 결과가 시간적, 공간적으로 거리를 두고 있다면 비도덕적 행위를 하기 더 용이해진다. 반두라는 해로운 결과의 측면에서 발생하는 도덕적 이탈의 대표적인 예로 지구 온난화를 제시한다. 일상에서 우리의 수많은 행동은 지구 온난화 가속에 직·간접적으로 영향을 미친다. 그러나 그 행위의 결과는 즉각적으로 드러나지 않으며 눈에 보이지도 않는다. 따라서 도덕적 기준이 엄격하여 책임감이 강한 사람일지라도 자신의 평소 생활 습관이 유해한 결과를 초래할 가능성이 높다는 것을 알면서도 무시하는 경우가 발생하는 것이다. 나아가 동기화된 추론 연구에 따르면, 강력한 과학적 증거에도 불구하고 사람들의 기후 변화에 대한 신념은 분열되어 있는 모습을 강한 방향성 목표에 의한 당파적 추론의 유형으로 분석한다(Bayes & Druckman, 2021). 사람들은 정치적 신념에 따라 환경 문제에 대한 서로 다른 과학적 증거들을 과장하여 해석하고, 자신과 반대되는 과학적 사실들의 신뢰도에 의문을 제기하였다. 정치적 보수주의자들은 기후 변화의 위험성에 대한 합의된 과학적 증거에 대해 회의적으로 평가하였다. 특히 친기업적인 성향의 사람들은 개발로 인해 발생하는 환경적 부정적 요인을 축소하거나, 기술적으로 해결 가능한 범위 내에 있는 것으로 인식하였다. 이처럼 자신의 행위로 인해 발생하는 해로운 결과를 축소, 무시, 왜곡하는 과정에서 기화된 추론은 강한 방향성 목표를 토대로 비도덕적 행위를 정당화한다.

마지막으로 네 번째 측면은 자신의 행위로 발생한 유해한 결과에 의해 영향을 받을 피해자 측면과 관련하여 피해자를 ⑦비인간화

(dehumanization)[33]하고, ⑧비난을 오히려 그들에게 귀인(attribution of blame)시킴으로써 도덕적 이탈의 양상을 보인다. 비도덕적 행위에 대한 도덕적 자기 비난으로부터 벗어나는 것은 행위로 인한 피해자들을 인식하는 방식과 관련이 있다. 자신의 비난받을만한 행위가 자신과 직접적으로 관련이 있는 사람들인 경우 도덕적 이탈은 억제된다. 인간에 대한 기본적인 동정심이 발현되기 때문이다. 자신의 행위로 인해 자신이 아는 구체적인 인물이 고통을 받는 모습에 쉽게 공감할 수 있으며 이때, 도덕적 기준은 행위를 한 자기 자신을 비난하도록 한다. 사람들은 자신의 비도덕적 행위로 인해 발생할 피해자들의 인간성을 제거하는 방법을 통해 자기 비난으로부터 벗어난다. 즉, 인격적 특질이 제거되면 피해자에 대한 죄책감 수준이 낮아지며, 도덕적 기준의 적용이 느슨해지기 때문이다. 인간 이하의 존재로 가치를 평가 절하하면 더 이상의 공감적 고통이나 양심의 가책은 느끼지 않아도 되기 때문이다. 피해자 측면에서 동기화된 추론으로 인한 비도덕적 행위의 정당화 문제의 대표적인 사례로 나치의 홀로코스트를 떠올릴 수 있다. 히틀러 집권 당시 독일 나치당은 우생학에 근거하여 독일인을 '아리아인의 인종'의 일원으로 찬양하고, 유대인을 포함한 흑인, 산티족 등을 '비아리아인'으로 규정하여 인간적으로 열등한 존재이자, 인종 청소의 대상으로 설정하였다. 유대인 대학살을 일컫는 '홀로코스트(Holocaust)'는 그리스어 holokauston에서 비롯되었는데, 이는 고대 그리스에서 신에게 동물(olos)을 태워서(kaustos) 제물로 바치는 것을 의미한다. 이처럼 잔혹한 행위로 인한 피해자들을 비인간화함으로써

33) 반두라는 비인간화(dehumanization)와 비인격화(depersonalization)를 다음과 같은 의미로 구분하여 사용하고 있다. 비인간화는 타인의 인간으로서의 자질을 박탈함으로써 악마나 야수적인 존재로 간주하는 것을 지칭하며, 비인격화는 다른 사람과 정서적인 거리를 두어 냉대하여 인간으로서 거의 고려하지 않는 것을 지칭한다. 반두라는 비인격화는 건설적인 목적을 위해 사용되는 경우가 있음을 제시한다. 예를 들어, 사회복지사들은 그들이 만나는 빈곤, 질병으로 고통받는 사람들을 매일 마주해야 한다. 이때, 정서적으로 깊게 공감을 하게 되면 정서적 고통이 너무 강렬하여 그들에게 적절한 도움을 주기 어려울 수 있다. 이 경우, 자신의 업무의 일환으로 그들을 비인격화함으로써 정서적으로 적절한 거리를 유지함으로써 객관적인 입장에서 그들에게 적절한 조치를 취할 수 있기 때문이다(Bandura, 2016/2018: 107-108). 따라서 비도덕적 행위를 정당화하는 것을 예방하기 위해서는 비인격화에서 비인간화로 확대되는 것을 경계해야 한다.

도덕적 주체로서의 자신의 지위를 유지할 수 있는 것이다.

이상에서 살펴본 여덟 가지 도덕적 이탈의 메커니즘은 개인이 결정을 내리고 선택을 경험하는 방식을 재구성하는 동기화된 추론의 과정으로 분석될 수 있다. 도덕적 이탈은 자신이 '하고 싶은 것'과 자신의 도덕적 기준이 허용 가능한, 즉 '해야만 하는' 혹은 '해서는 안 되는 것' 사이에 진정한 충돌이 있다는 생각을 약화시키는 것을 목표로 동기화되기 때문이다. 도덕적 이탈은 누군가가 자신의 도덕적 기준을 무시하고, 그로 인해 자기 일관성을 유지하려는 그의 욕구가 좌절되는 상황에서 발생한다. 상상된 행동과 (도덕적) 기준 사이의 실제 또는 상상의 인지적 충돌로 인해, 이 사람은 행동을 수행하기도 전에 이미 모종의 죄책감을 경험한다. 죄책감은 일종의 예견적 자기 제재(anticipatory self-sanctioning)를 형성하며, 이는 행위자가 인식한 갈등을 더 잘 이해하게 한다. 이러한 예견된 죄책감은 인지부조화를 구성한다. 인지부조화는 결국 갈등을 해결하기 위해 동기화된 추론의 과정을 촉발하는 경향이 있다.

이처럼 동기화된 추론이 비도덕적 행위를 정당화하는 과정은 한 사람의 행동에 대한 잠재적인 도덕적 저항을 예상하고, 그 일이 발생하기 이전에 자신의 행동을 방어하기 위한 이유나 설명을 만들어 내는 방식으로 수행되기도 한다(Hindriks, 2015). 가령, 한 사람은 자신이 속한 집단에서 비도덕적 행위를 하기 전에 사전적 정당화를 보일 수 있다. 그는 자신의 비도덕적 행위가 집단의 목표를 달성하거나 이익을 보호하는 데 도움이 될 것이기 때문에, 그 행동을 하기에 앞서 그 행동이 정당하다고 스스로를 사전적으로 설득할 수 있다. 이러한 사전적 정당화를 통해 도덕적 주체로서 개인은 자신의 행동에 대한 책임감을 줄일 수 있고, 심지어 그러한 행동 전후에 긍정적인 자아상을 유지할 수 있게 해준다.

도덕 판단에서 동기화된 추론의 사전적 정당화 양상에 주목한 힌드릭스는 사후적 정당화와의 차이를 동기화된 추론이 향하는 방향에서 찾는다. 사후적 정당화가 타인을 향하는(other-directed) 방향성을 띠는 반면, 사전적 정당화는 자신을 향하는(self-directed) 방향성을 띤다(Hindriks, 2015: 242). 즉, 도덕 판단에서 동기화된 추론의 사후 정당화 양상은 자

신과 다른 사람들에게 자신의 행동을 정당화하고 설명하기 위한 방법으로, 행동이 일어난 후에 발생하는데 반해, 사전적이고 예견적인 정당화의 양상은 자신의 판단과 행동에 앞서 스스로를 납득시키기 위한 방법인 것이다. 다음의 사례를 살펴보자.

> *고속도로에서 운전하고 있고, 자신이 과속하고 있다는 것을 알아차린다고 상상해 보자. 당신은 교통 법규를 위반했다는 것을 깨닫고 아마도 당신이 사고를 당할 위험을 증가시킨다는 것을 인지할 것이다. 그러나 동시에 다른 사람들도 과속하고 있다는 것을 알게 된다. 이 두 번째 고려 사항에 기초하여 당신은 지금처럼 과속 운전하는 것이 괜찮다고 결론짓는다. 이 시나리오에서 당신은 도덕적 추론을 하고 있다. 당신은 도덕적으로 관련이 있는 문제를 고려한다 - 당신이 다른 사람들을 노출시키는 위해의 위험을 증가시키는 것을 고려할 때 속도를 내는 것이 괜찮은지 여부. 그리고 당신은 이 문제를 관찰한다. 다른 사람들도 마찬가지이다. 당신은 이 관찰을 통해 속도를 내는 것이 괜찮다는 주장을 공식화한다 - 다른 사람들도 그렇게 하기 때문에 속도를 내는 것은 괜찮고 - 그리고 당신은 도덕적 판단을 형성한다. 속도를 내는 것은 괜찮다는 주장을 공식화한 것이다(Hindriks, 2015: 238).*

도덕적 주체는 어떤 주체가 허용, 금지 또는 의무적으로 무엇을 해야 하는지에 대한 결론에 도달하기 위해 도덕적 고려를 평가할 때 도덕적 추론에 참여한다. 평가의 주체가 곧 평가의 대상과 항상 일치하진 않기 때문에 도덕 판단에서 동기화된 추론은 타인을 향할 수도, 자기 자신을 향할 수도 있는 것이다. 도덕 판단이라는 결론에 도달하는 것은 판단을 구성하는 일과 내려진 판단을 재확인하는 문제로 구분될 수 있다. 이러한 관점에서 본다면 앞서 살펴본 하이트의 '도덕적 말 막힘' 연구는 동기화된 추론이 이미 내려진 도덕 판단을 재확인하는 문제임을 보여준다고 할 수 있다. 자신의 도덕적 신념에 부합하는 도덕 판단의 근거를 대기 위해 사용된 동기화된 추론은 다른 사람을 설득하고자 하는 목표, 즉 타인을 향하는 방향을 갖는다. 자신의 도덕적 입장을 고수하며 타인을 설득하기 위해 도덕적 추론이 사용되기도 하지만, 설득은 또한 자신을 향할 수도 있다. 위의 사례에서 판단의 주체는 그 순간 과속을 해도 괜찮

다는 결론에 도달하기 위해서 자기를 향하는 추론을 수행한 것이다. 그리고 그의 목표는 자신의 비도덕적 행동이 괜찮다고 스스로를 납득시키기 위한 것이다. 이와 같이 동기화된 추론은 주로 사후적인 방식으로 자신의 비도덕적 행위를 정당화하지만, 때로는 예견적인 방식으로도 이루어지기도 한다.

2. 도덕적, 정치적 불일치의 양극화 심화

도덕 판단에서 동기화된 추론이 유발하는 주요 문제 중 하나는 기존에 가지고 있는 도덕적 신념과 태도의 집단적 분열을 강화함으로써 도덕적, 정치적 신념의 양극화를 심화시킬 수 있다는 것이다. 이는 중요한 문제에 대해 생산적인 대화와 타협에 참여하는 시민들의 능력을 약화시킨다는 점에서 민주주의에 심각한 결과를 초래할 수 있다. 민주주의 사회는 도덕과 종교 및 시민들의 신념에 대한 합당한[34] 불일치가 존재하는 자유로운 사회이다. 시민들은 양립할 수 없는 교리에 동의하기 때문에, 무엇이 좋고, 중요한지에 대해 불가피하게 의견이 다르다. 그러나 현재 극심하게 치닫고 있는 집단 간 갈등과 혐오, 배타주의는 무지와 비이성적 편견에 뿌리를 둔 도덕적, 정치적 불일치에서 비롯된다. 일련의 연구들은 이와 같은 도덕적, 정치적 불일치의 극단적인 양극화가 동기화된 추론에 의해 심화되고 있음을 지적한다. 또한 지적이고 양심적인 사람들조차 분열적인 도덕적, 정치적 문제에 대해 합리적이고 편견이 없는 방식으로 추론하기 어렵게 만드는 판단의 부담과 같은 장애물에 직면해 있다고 분석한다(Ancell, 2017: 121). 우리가 지적이고 양심적인 사람들에게 판단의 부담을 안정적으로 극복할 것을 기대할 수 없듯이, 불합리한 근원을 안

34) 정치철학자 롤스(J. Rawls)는 그의 저서 『정치적 자유주의(J. Rawls, 2005, 장동진 역, 2016)』에서 합리적인 것(the rational)과 합당한 것(the reasonable)을 구분하여 논의를 전개한다. 합당한 것은 평등한 존재들 간의 사회적 협동의 조건으로서 원칙과 기준을 제시하고 그것을 기꺼이 준수할 태도가 있다는 것을 의미하며(140-141), 합리적인 것은 합당한 것과는 구분되며 자신의 고유한 목적과 이익을 추구하는 데 있어 판단과 숙고를 통해 가장 효율적인 수단을 채택할 수 있다는 것을 의미한다(143). 이 두 개념은 시민들의 도덕적 능력을 나타낸다. 특히, 합당한 것의 기본적인 측면은 상호성과 공적 이성의 사용과 관련이 깊다.

정적으로 극복할 것을 자연스럽게 기대할 수는 없다는 것이다.

도덕 판단의 과정에서 나타나는 동기화된 추론이 도덕적, 정치적 양극화를 촉진하고 심화시키는 데 중요한 역할을 한다는 연구가 증가하고 있다(Ditto et al., 2010; Huemer, 2016; Kahan, 2013). 사람들은 자신의 신념에 도전하는 정보와 주장을 편향되거나 결함이 있는 것으로 무시하면서 자신이 지지하는 도덕적, 정치적 신념을 뒷받침하는 정보와 주장을 신뢰할 수 있고 설득력 있는 것으로 볼 가능성이 더 높다. 그 반대 측의 입장은 상대적으로 평가절하하고 과소평가하기 때문이다(Kunda, 1990).

또한 케이헌(Kahan, 2013)은 동기화된 추론과 정치적 양극화의 관계에 대한 연구를 수행하여 개인이 정치적 신념에 도전할 때 동기화된 추론에 참여할 가능성이 더 높다는 것을 발견하였다. 그의 연구는 사람들이 자신의 신념을 뒷받침하는 정보를 찾고, 모순되는 정보를 무시하는 경향이 있어서 그들의 정치적 견해의 강화로 이어진다는 것을 보여주었다. 이러한 현상은 개인이 점점 더 자신의 신념을 확고히 하고 대안적 관점을 고려할 가능성이 적기 때문에 정치적 양극화로 이어질 수 있다. 나아가 디토 연구진은 동기화된 추론 연구에 대한 메타분석을 통해 미국의 대법관들, 즉 원칙적인 법적 추론의 전문가들이 광범위한 헌법적 법률 원칙에 호소함으로써 종종 그들이 선호하는 법적 결론을 합리화했다는 것을 보여주는 연구들을 제시하였다(Ditto et al., 2010: 324).

현실에서 정치적 양극화의 모습에서도 이러한 양상은 쉽게 찾아볼 수 있다. 사람들은 자신의 정치적 견해를 대중적으로 호소할 때, 개인이나 집단 이익의 측면보다는 주로 도덕적 이유를 제공하기 때문이다. 누군가에게 '더 좋은 복지 혜택', '법인세 인상', '이민 제한의 확대' 정책을 지지하는 이유를 물으면, 그들은 '바로 그것이 나에게 이익이 되기 때문'이라고 대답하지 않을 것이다. 대신, 그들은 보통 가난한 사람들을 돕기 위해 우리가 어떻게 더 많은 것을 해야 하는지, 혹은 기업들이 그렇게 적은 세금을 내는 것이 얼마나 불공평한지, 혹은 이민자들이 열심히 일하는 자국민들로부터 일자리를 빼앗는 방법에 대해 이야기한다(Ancell, 2017: 77). 개인적인 이익보다는 도덕적, 사회적 차원의 명분은 자신의 주장을

공정한 것으로 보이게 만들기 때문이다. 이처럼 도덕 판단에서 나타나는 동기화된 추론은 우리가 어떤 의미에서는 진실하기를 원하는 결론에 기여하는 이유와 증거에 대한 우리의 평가를 체계적으로 왜곡함으로써 자신의 집단 혹은 개인의 도덕적 신념을 강화하는 정보는 확대 해석하고, 그와 위배되는 정보들은 거부하도록 한다. 휴머(Huemer, 2016)는 특정한 결론에 대한 이러한 선호를 '비인식론적 신념 선호' 현상이라고 부르는데, 이는 '우리가 믿고 싶은 것들, 그 명제들의 진실과 무관한 이유들 또는 그것들이 증거에 얼마나 잘 뒷받침되는지'를 지칭하는 개념이다(Huemer, 2016: 460). 사람들은 믿고 싶어 하는 것이 매우 다양하지만, 심리학에서의 풍부한 증거는 사람들이 일반적으로 자신의 이익을 증진하고 그들의 기존 신념을 지지하고, 자신이 가진 편견을 정당화하고 그들의 당파적인 충성을 정당화하는 신념을 선호한다는 것을 보여준다.

역사적인 사건들 또한 도덕적 원칙의 동기화된 사용이 도덕적 불일치와 정치적 양극화의 원천을 제공함을 보여준다. 가령, 노예제도의 도덕성에 대한 갈등을 살펴본 앤더슨(2015)은 노예제도 찬성론자와 폐지론자가 당시의 지배적인 도덕적 원칙인 '황금률'을 선택적으로 적용하고 있음을 지적한다. 황금률은 다른 사람이 우리에게 해주기를 바라는 대로 다른 사람에게 하라고 말한다. 폐지론자들은 노예 제도가 잘못된 것이라고 주장하기 위해 황금률을 적용했다. 노예가 되고 싶지 않다면 다른 사람을 노예로 삼아서는 안 되기 때문이다. 그러나 원칙에 입각한 추론에 능숙하고 황금률을 지지하는 노예제 찬성론자들은 황금률을 적용하면 모든 사회적 불평등의 철폐와 같은 직관에 반하는 결과를 초래할 수 있기 때문에 이를 적용해서는 안 된다고 반박했다. 나아가 그들은 신이 부여한 사회 질서이므로 사회적 불평등이 폐지되어서는 안 된다고 주장했다. 이 사례는 동기화된 추론을 통해 노예제도를 합리화하는 도덕 원칙을 선택적으로 적용하는 이기적인 편향의 양상을 보여준다.

동기화된 추론의 고전적인 증명 중 하나인 로드 등의 연구(1979)는 동기화된 추론이 도덕적, 정치적 양극화에 기여한다는 것을 설명한 최초의 연구다. 해당 연구에서는 강한 도덕적 신념을 가진 사람들이 복잡한 사

회적 문제를 검토하는 데 관련 경험적 증거를 편파적으로 검토할 가능성이 있다는 가설을 세웠다. 이들은 자신의 신념에 반하는 비확증적 증거는 비판적인 평가를 하는 반면, 자신의 신념에 부합하는 확증적 증거는 액면 그대로 받아들이는 경향이 있고 그 결과, 제공된 다양한 지식과 정보들을 바탕으로 자신의 초기 입장을 더욱 강하게 유지하고자 했다. 이를 검증하기 위해 연구자들은 윤리적 이슈 중 하나인 사형제도에 대한 강력한 찬성 또는 반대 의견을 가진 사람들을 실험 참가자로 모집하였다. 실험 참가자들은 사형 억제 효과를 조사한 두 가지 연구의 짧은 요약을 읽는다. 한 연구는 사형이 억지력으로 작용한다고 제안한 반면, 다른 연구는 그 반대 주장을 제안했다. 그리고서 연구 참가자들에게 연구의 질과 그 연구가 그들의 신념을 뒷받침하는지 또는 부정하는지의 여부를 평가하도록 요청했다.

해당 연구는 사람들이 일반적으로 그들의 기존 견해와 상충되는 연구들이 더 열악하다고 생각했고, 전반적으로 그 증거가 그들의 기존 믿음을 뒷받침한다는 것을 발견했다. 사형제 지지자들은 이 연구가 억제 효과가 있다는 것을 보여주는 것이라고 해석했지만, 반대자들은 억제 효과를 보이지 않는 것으로 해석한 것이다. 더욱이 실험의 말미에, 실험 참가자들에게 사형에 대한 그들의 견해를 물었을 때, 연구진은 사람들이 그들의 견해가 더 극단적으로 변했다는 것을 발견했다. 그들은 '사형 찬성자와 반대자를 동일한 증거(표면적으로는 동등한 수준의 지지와 반증을 제공했음)에 노출시킨 순수 효과는 그들의 견해 차이를 더욱 증가시켰다'라고 말한다. 만약 이들이 자신의 초기 신념과 관계없이 정보를 합리적이고 객관적으로 처리하고 있다면 실험 참가자들은 자신의 신념을 수정할 가능성을 가지고 있거나, 적어도 자신이 가지고 있는 신념과 관련된 이슈가 이전의 생각보다는 훨씬 더 복잡했으리라는 것을 깨달을 것이다. 그렇게 된다면, 윤리적 이슈와 관련한 다양한 정보를 접하게 하는 것은 해당 이슈에 대한 서로 다른 견해의 차이가 좁혀질 가능성을 발견할 수 있을 것이다. 그러나 인지부조화 이론 연구 결과와 마찬가지로, 사람들은 동기화된 추론에 기반하여 정보를 왜곡하여 수용하거나 거부함으로

써 양극화를 심화시켰다.

한편, 또 다른 연구에서(Cohen, 2003) 실험 참가자들은 어떤 정책을 선호하는지에 대한 질문 이외에도 다음의 요인들이 그들의 판단에 얼마나 영향을 미치는지 평가하도록 요청받았다. 그 요인으로는 ①정책의 세부 내용, ②사회 문제에서 정부의 역할에 대한 개인적 철학, ③전형적인 민주당원 혹은 공화당원이 믿는 것, 그리고 ④복지에 관한 사전지식/배경이 제시되었다. 실험 참가자들은 압도적으로 ①정책의 세부 내용과 ②개인적 철학에 영향을 받았고, 자신이 속한 당의 다른 구성원들이 믿는 것에 가장 영향을 적게 받았다고 보고했다. 그러나 결과로부터 그들 당의 입장에 대한 정보의 존재에서 정책의 세부 사항들은 거의 차이를 보이지 않았다. 사람들은 단지 자신이 지지하는 당의 일부 정책들을 선호했을 뿐이다. 동기화된 추론은 코헨의 연구에서 실험 참가자들이 그러한 선호가 단순히 당파에 대한 순응의 결과일 때에도 그들의 정책 선호에 대한 도덕적 정당성을 생성할 수 있었던 이유를 설명한다. 사람들은 자신이 동일시하는 정당이 해당 문제에서 '옳은' 위치라고 믿기를 원한다. 사람들은 지지하는 당이 특정 정책에 반대한다는 말을 들었을 때, 지지자들로 하여금 그 정책이 나쁘다는 결론에 도달하기를 원하게 만든다. 이러한 목표에 의해 동기화된 상태에서 그들은 정책을 해석하고, 그 결론에 도달하기 위해 관련된 이유와 증거를 어떤 방식으로든 저울질 하는 것이다(Ancell, 2017: 78). 가령, 우리가 지지하는 정당이 복지 정책을 확대하길 원한다면, 국민의 인도적인 삶에 대한 정부의 도덕적 책임을 강조하고, 만약 이를 반대하길 원한다면 자신의 삶에 대한 주체성과 개인적 책임의 중요성을 강조할 것이다. 비록 사람들은 자신이 객관적이라고 생각하지만, 그들이 도덕적 가치를 우선시하는 방식은 도달하고자 하는 특정 목표에 의해 동기화된다. 그 경험적 증거는 동기화된 추론이 도덕적, 정치적 이슈에 대한 우리의 생각 대부분에 퍼져있다는 것을 강하게 시사한다.

이처럼 도덕 판단에서 나타나는 동기화된 추론은 도덕적, 정치적 불일치를 더욱 확대하는 문제점을 지닌다. 동기화된 추론은 앞서 살펴본 사

형 제도를 포함하여, 다양한 윤리적 이슈에 대해서도 유사한 양상을 띠었다. 나아가 추가적인 연구들은 과학적 사실로서 입증된 증거에 대한 반응이 자신의 도덕적 신념에 부합하는 방식으로 선택적으로 수용함을 입증했다. 대표적으로 케이헌 연구진(Kahan et al., 2010)은 기후 변화, 핵 폐기물 처리, 총기 소지 허용 등 다양한 이슈를 바탕으로 동기화된 추론이 도덕적, 정치적 불일치를 심화시키는 인지 과정을 확인하였다. 이들에 따르면 사람들은 자신이 가지고 있는 도덕적 신념이나 사회 질서에 대한 관점에 의해 누구를 권위 있는 과학 전문가로 간주하고, 어떠한 과학적 지식을 수용할지 예견해 준다고 보았다. 자신의 입장을 지지하지 않는 전문가의 지식은 거부하였으며 심지어는 반대편 측의 전문가를 비전문가로 취급하기도 하였다(Kahan et al., 2010). 즉, 대중들이 전문가의 말을 이해하지 못하거나, 따르기를 꺼리는 것이 아니라 자신의 도덕적 신념이나 정치적 입장에 의해 전문가의 신념에 반대되는 인식을 형성하는 경향이 있기 때문이라는 것이다. 따라서 윤리적 이슈에 대한 전문가의 사실적인 지식은 사회적 논쟁을 잠재우지 못하며, 동기화된 추론에 의해 기존의 도덕적 불일치와 정치적 양극화에 기여하는 방향으로 해석된다는 것이다.

전문가에 의해 합의된 사실적 증거가 있음에도 동기화된 추론으로 인해 양극화가 심화되는 가장 대표적인 사례는 기후 변화에 대한 신념의 차이이다(Dieckmann et al., 2017; Kahan, 2015; McCright & Dunlap, 2011). 미국에서 기후 변화에 대한 신념과 태도는 과학적 증거들을 토대로 전문가 집단에서 합의한 사실에도 불구하고, 당파적인 분열 상태로 남아 있다. 이는 비단 미국의 문제만으로 국한되지 않는다. 56개국에서 이루어진 25개의 여론 조사와 171개의 학술연구에 관한 메타분석은 '기후 변화에 대한 신념에 대한 가장 큰 인구학적 상관관계는 정치적 소속'이라는 결론을 내렸다(Hornsey et al., 2016: 622). 이러한 사회적 양극화 문제에 대해 동기화된 추론 연구는 강한 설명력을 제공한다. 초기의 연구들은 미국 내 두 정당의 구성원들이 서로 다른 정당 규범, 가치, 이전의 기존 신념에 기반을 두고 서로 다른 원하는 결론을 달성하기 위해 기

후 변화와 관련된 지식과 정보를 수용하거나 회피하기 때문에 당파적 양극화가 발생한다고 보았다(Dieckmann et al., 2017). 또한 기후 문제를 둘러싼 정치적 맥락이 더욱 양극화되면서 시민들은 자신의 편이 논쟁에서 이기도록 하는 욕구에 의해 동기화되기도 한다(Kahan, 2015). 즉, 기후 변화에 대한 새로운 정보가 등장하면 사람들은 자신이 원하는 결론을 유지하고자 방향성에 의해 동기화된 추론을 사용한다. 가령, 보수주의자들은 기후 변화에 대한 회의적인 입장을 유지하기 위해 인간이 초래한 기후 변화에 대한 과학적 합의에 대한 사실적 정보를 거부하며, 오히려 그러한 정보를 접한 후에 회의적인 입장이 더욱 강화되기도 하였다(McCright & Dunlap, 2011). 반면, 동일한 정보를 제공받은 진보주의자들은 비판적 검토 없이 수용한다. 나아가 이들은 지구 온도 상승에 대한 더 많은 자극적인 정보들에 주목하며 도덕적 명분을 바탕으로 기후 친화적인 행동에 참여하도록 독려하기도 한다(Luo & Zhao, 2019). 이처럼 동기화된 추론의 작용으로 기후 변화에 대한 신념은 과학적 합의나 사실적 정보와 관계없이 자신의 도덕적 신념이나 정치적 입장에 기반하여 양극화된다. 이와 같은 현상은 비록 다른 주제이지만 앞에서 살펴본 사형제도 이외에도 이민이나 총기소지, 핵 폐기물 처리 등 다양한 윤리적 이슈에서 유사한 구조를 보였는데, 이러한 현상의 공통점은 각 이슈가 가진 메시지가 위험이나 도덕성을 유발했다는 것이다.

도덕적 신념을 포함하여 하나의 핵심적인 신념을 가지고 있는 경우 이것이 정보를 수집하고 해석하는 방식을 결정하는 방향에 의해 동기화된 추론이 나타나며, 이로 인해 도덕적, 정치적 신념의 양극화가 더욱 심화된다. 진보주의자와 보수주의자의 동기화되 추론 양상을 분석한 무니는 동기화된 추론에 의해 정치적 양극화가 발생하는 과정을 '선택적 노출 효과'로 설명한다(Mooney, 2012: 214-219). 선택적 노출 효과는 동기화된 추론의 출발점인 페스팅거의 인지부조화 이론 연구에서 발견한 파생적인 심리적 현상이다. 이 현상은 우리가 자신의 도덕적, 정치적 신념에 맞는 정보에 의식적/무의식적으로 더 많이 노출되며, 신념에 반하는 진실은 회피하도록 선택한다는 것이다. 선택적 노출은 특정 언론사의 뉴스만 보

거나, 특정 성향의 SNS 채널을 구독하는 등 일반적으로 개인적 차원에서 일어난다고 생각될 수 있다. 그러나 자신이 속한 사회적, 문화적 환경에 의해서도 자연스럽게 선택적으로 노출이 되는 경우도 많다. 가령, 동일한 교육과정 속에서도 교사나 학교 분위기에 따른 잠재적 교육과정이나 영 교육과정의 영향은 집단적 선택적 노출의 가능성을 시사한다. 선택적 노출 효과는 자신의 신념에 반하는 불편한 진실에 대해 인지적인 장벽을 높게 세운다는 점에서 동기화된 추론이 도덕적, 정치적 불일치를 심화시키는 양상을 설명해준다. 선택적 노출 효과는 우리의 도덕적 판단과 의사결정 과정이 동기화된 추론에 의해 확증 편향의 영향을 받는다는 발견(Lodge & Tabor, 2006)이 단지 개인 내적인 심리적 과정에 국한되지 않고 매체를 포함한 다양한 사회적 환경에 강한 영향을 받고 있음을 시사한다. 선택적 노출은 확증 편향을 통해 자신의 의견이나 신념을 뒷받침하는 증거를 찾거나 과대평가하도록 만들고, 이를 통해 자신을 위협하는 정보를 공격하고 회피하게 만들기 때문이다.

동기화된 추론의 선택적 노출 효과는 페스팅거의 이론이 발표된 직후인 1960년대 이래로 활발히 연구되었으며, 2000년대 초반에 메타분석이 이루어졌다(Hart et al., 2009). 하트 연구진은 선택적 노출 효과가 정확성 동기가 아닌 방향성 동기[35]에 의해 동기화된 추론에 의해 유도된 효과임을 메타 분석을 통해 밝히고자 하였다. 이들은 2008년 2월 이전까지 연구된 동기화된 추론 연구물 중 선택적 노출 효과와 관련된 키워드를 기반으로 300개의 대상 논문을 선정하고 이를 12가지 변수를 기준으로 코딩하여 동기화된 추론과 선택적 노출 효과의 관계를 밝히고자 하였다.

35) 해당 연구에서는 '정확성(accuracy)' 동기와 대비되는 개념으로 '방향성(direction)' 동기가 아닌 '방어적(defense)' 동기의 개념을 상정하였다(Hart et al., 2009). 이때 '방어적'이라함은 자신의 기존의 신념, 태도 또는 행동을 보호하려는 방향에 의해 동기화된 추론을 의미한다. 방어적 동기는 기존의 태도, 신념 또는 행동에 대한 개인의 헌신과 지속적인 가치에 대한 높은 관련성에 의해 강화된다(Hart et al., 2009: 555). 즉, 자신의 신념을 '방어'하려는 목표라는 일정한 방향에 의해 동기화된 추론으로 이는 방향성 동기에 의한 동기화된 추론의 일종으로 볼 수 있다. 이에 본 논문에서 2장에서 규명한 동기화된 추론의 두 가지 유형이자, 도덕 판단에서 문제가 되고 있는 방향성에 의한 동기화된 추론이라는 용어로 사용하였다.

해당 연구는 실험 참가자들이 가지고 있는 기존의 도덕적, 정치적 태도, 신념 또는 행동을 지지하는 정보와 이러한 성향에 반하는 정보 사이에서 선택을 제공하는 상황에서 기존의 신념 및 행동과 관련된 정보 선호도를 조사하였다. 연구 결과, 자신의 신념과 반하는 정보보다 자신의 신념에 일치하는 정보에 대한 선호도가 통계적으로 유의미하게 높았다. 특히, 방향성 동기에 의한 동기화된 추론을 증가시키는 요인의 함수로 인해 일관성 편향(congeniality bias; 즉, 자신의 기존 신념에 기여하는 방향으로 정보를 수용하고 해석하는 양상)이 증가한다는 것을 보여주었다. 이러한 편향은 정보의 질, 정보를 얻기 위한 노력 정도, 가치 관련성, 폐쇄성과 긍정적인 상관관계가 있었지만, 자신의 기존 태도, 신념 또는 행동에 대한 자신감이나 지지와는 부정적인 상관관계가 있었다. 사람들은 도덕적, 정치적 신념을 위협하는 정보보다 지지하는 정보를 소비할 가능성이 약 두 배 높았다. 특히, 자신의 신념을 보호하고자 하는 방어적 목표[36]가 뚜렷한 경우 선택적 노출에 더욱 취약했다. 이처럼 방향성 동기는 선택적 노출을 촉발시킬 가능성을 높인다는 사실을 시사한다. 선택적 노출 효과는 사람들이 동기화된 추론을 통해 자신의 도덕적 세계관에 기초하여 자신만의 현실을 만들어내는 양상을 설명해준다. 즉, 동기화된 추론은 자신과 자신이 속한 공동체의 신념에 기반하여 그것에 부합하는 방향으로 편향된 판단을 도출하며 자신과 다른 신념을 가진 타인 및 타집단에 대한 배타주의를 강화하는 데 기여한다.

나아가 동기화된 추론과 배타주의에 관한 연구는 방향성 목표가 갖는 사회적 특징이 도덕적, 정치적 신념의 양극화를 가속한다는 사실에 주목한다(Bayes & Druckman, 2021). 기후 변화나 이민 문제 등 윤리적 이슈에 대한 도덕적, 정치적 신념은 타인과의 사회적 연결에 대한 우려와 관련되며 특히 개인적으로 중요한 사회적 집단 즉, 자신의 소속 정당과의 동일성을 유지하려는 욕구 즉, 사회적 정체성에서 비롯된다는 것이다.

36) 이때, 해당 연구에서 확인된 방어적 동기 중 하나는 인지적 폐쇄성이었다. 폐쇄적인 마음은 선택적 노출을 촉진하였으며, 경험에 대한 개방된 마음은 선택적 노출을 약화시켰다(Hart et al., 2009: 577). 이는 동기화된 추론의 문제를 개선할 수 있는 가능성을 시사하기도 한다.

이와 함께 방향성 동기는 정당뿐만 아니라 자신의 친구, 가족을 비롯한 사회적 합의에 일치하고자 하는 욕구에 기인하기도 한다. 사회적 합의에 따라 사람들은 상충하거나 모호한 정보에 대해 자신의 사회적 관계망이나 동료들의 관점에서 정보의 수용 여부를 결정한다. 따라서 사회적 합의는 과학적 합의와 무관하게 기후 변화에 대한 신념에 영향을 미칠 수 있다(Kobayashi, 2018). 특히, 개인의 사회 집단은 종종 당파적으로 동질적일 가능성이 높기 때문에 개인이 사회적 합의에 의존하여 윤리적 이슈에 대한 자신의 신념을 무비판적으로 확립하는 것은 동기화된 추론으로 인한 양극화 심화를 더욱 공고히 하는 데 기여할 뿐이다.

제 3 절 도덕 판단에서 동기화된 추론 문제의 원인

1. 도덕적 신념의 보호

앞선 논의에서 살펴본 것처럼, 추론의 동기로서 방향성은 주로 자신을 보호하고자 하는 자기본위적인 방향을 갖기 때문에 방어적인(defensive) 목표에 의해 동기화된 추론으로 분류하기도 한다. 이처럼 방향성 혹은 방어적 목표에 의한 동기화된 추론은 특정한 결론에 도달하기 위해 주어진 환경과 정보를 선택적으로 수용하고, 거부하는 양상을 보인다. 동기화된 추론 연구의 모태가 되는 인지부조화 이론에 따르면, 사람들은 그들의 믿음과 행동 사이의 일관성을 유지하고, 자신을 이성적이고, 일관되고, 지식이 있는 것으로 인식하는 자연스러운 경향을 가지고 있다. 우리는 자신의 신념, 태도, 가치관에 모순되는 정보에 직면할 때, 심리적인 불쾌함을 경험한다(Festinger, 1957/2016). 인지부조화가 발생하게 되면, 우리는 불쾌함을 줄이기 위해 동기화된 추론을 사용한다. 동기화된 추론을 통해 긍정적 자아상을 유지할 수 있기 때문이다. 특히, 도덕적 신념이나 태도, 가치관에 모순되는 상황 즉, 정체성의 위협 상황에 직면하게 되었을 때, 인지부조화의 정도는 양적으로나, 질적으로 더 강하게 나타났다

(Graham. 2007). 이때, 동기화된 추론은 심리적 불편함을 줄이고 자신의 정체성을 보호하기 위해, 기존의 도덕적 신념과 상충하는 상황을 조화시킴으로써 긍정적인 도덕적 자아상을 유지하도록 한다.

신념은 개인의 정체성 형성에 중요한 역할을 한다. 우리는 자라면서 가족, 문화, 종교, 교육 및 사회적 규범을 통해 다양한 신념 체계에 노출된다. 이러한 신념은 우리의 자아감에 뿌리를 내리고 우리의 가치관과 태도 및 행동을 형성한다. 이를 통해 형성된 도덕적, 사회적 정체성은 우리의 윤리적이고 친사회적인 결정과 행동을 안내하는 일종의 나침반 역할을 할 수 있다. 나아가 신념이 자신의 정체성에 깊이 통합되면 자기 일관성을 경험한다. 즉, 자신의 신념과 정체성이 조화를 이룰 때 우리는 스스로를 더 진실되고 일관된 사람으로 이해하며 자기 확신을 느낀다.

사람들은 기본적으로 자신의 정체성과 관련된 가치관에 부합하는 선택을 하고자 하며, 자신의 신념을 위협하는 요인에 대해서는 거부하는 반응을 보인다. 신념에 대한 위협은 곧 자신의 정체성을 부정하는 것으로 여겨지기 때문이다. 특히 사람들은 자신의 도덕적 가치나 자신이 속한 집단의 규범과 강하게 동일시할 때, 또는 의사결정에서 이러한 가치를 우선시하고 도덕적 가치가 침해되었을 때 죄책감이나 수치심 같은 부정적인 정서를 경험할 가능성이 더 높다. 이러한 방식으로, 누군가의 도덕적 정체성은 개인의 도덕적 신념을 고수하고 그에 따라 행동하려는 동기를 강화할 수 있다. 도덕적 정체성은 그것이 활성화되고 의식적일 때, 또는 접근 가능할 때, 특히 행위자의 도덕적 기준에 따라 행동을 촉진하기도, 억제하기도 한다(Hindriks, 2015). 자기 일관성의 경향은 자아 정체성을 구성하는 핵심 요소이다. 자아개념에 대한 심리학적 기초를 최초로 제공한 제임스(James, 1890/2005)는 자아를 '인간이 자신의 것이라고 할 수 있는 모든 속성'으로 정의한다. 특질(trait) 이론을 제안한 성격심리학의 초기 연구자인 올포트(Allport, 1937)는 고유자아(proprium)라는 용어를 통해 개인에게 속해 있는 독특한 어떤 것, 즉 '느껴지고 알려진 대로의 나'라는 개념으로 자아의 속성을 정의한다. 한편, 상징적 상호작용론을 이야기한 미드(Mead, 1934)는 자아 개념을 사회적인 관점으로 해석한

다. 그는 자아는 다른 사람과의 관계 속에서만 의미가 있으며, 다른 사람의 관점으로부터 자신을 파악하는 반성적 사고를 통해 자아가 규정된다고 보았다. 한편, 정신분석 이론가인 에릭슨(E. Erikson)에 따르면 자아정체성(self-identity)은 청소년 시기의 일련의 위기 극복 과정을 통해 확립되는 것으로, '스스로 생각하는 확고한 자기 자신의 모습'을 의미한다. 동기화된 추론이 비도덕적 행위의 정당화 기제로 작동하도록 하는 매개변수로 거짓된 도덕적 정체성을 제시한 모쉬맨(D. Moshman)은 정체성이 '자기 자신을 한 명의 개별적 인간으로서 이해되도록 하는 명시적 이론'으로 간주될 만큼 충분히 조직적이고 설명적이며 의식적인 자신의 사회적 맥락에서의 자신에 대한 개념을 의미한다고 본다(Moshman, 2005: 89). 이러한 정체성 개념은 에릭슨의 개인적 정체성의 패러다임에 뿌리를 두고 있으면서도, 동시에 개인이 속한 다양한 사회적 집단 및 타인과의 상호관계 속에서 자신의 정체성을 확립하는 과정을 포함한다. 즉, 자신의 정체성을 다양한 사람들 및 자신이 속한 집단과 관련하여 정의내리고, 사회적 관계나 소속, 역할 등의 패턴을 가진 개인으로 스스로를 인식하는 것이다(Moshman, 2007: 117).

도덕 판단에서 동기화된 추론은 도덕적 신념에 부합하는 방향으로 근거를 수집하거나, 무시, 왜곡하는 방식으로 추론함으로써 자신의 긍정적 자아상을 보호하고 유지하는 판단을 도출해낸다. 사회인지이론에서는 이와 관련하여 우리의 도덕 판단에서 동기화된 추론의 양상이 도덕적 정체성을 유지하고자 하는 모습을 보인다는 것에 주목한다. 대표적으로 아퀴노(K. Aquino)는 가치, 목표, 특성, 행동 대본에 대한 헌신을 포함하여 누군가의 도덕적 헌신의 관점에서 도덕적 정체성을 정의한다(Aquino et al., 2009: 124). 강한 도덕적 정체성을 가진 누군가는 자신의 개인적, 공적, 집단적 자아 개념을 포함하여 다른 자아 개념보다 도덕적 자아로서의 자신을 더 중요하게 여긴다. 사람들은, 그들의 도덕적 정체성이 사고에서 중심적인 역할을 할 때, 그들의 도덕적 기준을 고수하는 경향이 있다. 도덕적 정체성을 확립한 사람, 즉 스스로를 도덕적 개인으로 보는 사람들은 자기 개념을 구성하는 요소 중 도덕성이 가장 중요한 부분을 차

지하게 된다(Aquino & Reed, 2002). 해당 연구는 개인의 행동을 형성하는 데 있어 도덕적 정체성의 중요성을 강조하고, 도덕적 정체성이 동기화된 추론의 원천으로 작용할 수 있음을 시사한다. 위 연구와 같은 맥락에서 카를로 등(Carlo et al., 2007)은 청소년기의 도덕적 정체성 발달에 관한 연구를 수행하였다. 그들은 개인적 자기 성찰에 참여할 가능성이 높을수록 강력한 도덕적 정체성의 형성으로 이어질 수 있다는 것을 발견했다. 그들의 연구는 또한 강한 도덕적 정체성을 가진 개인들이 친사회적 행동에 더 많이 참여하고 반사회적 행동에 덜 참여하기 때문에, 강한 도덕적 정체성이 일탈 행동에 대한 보호 요인으로 작용할 수 있다는 것을 보여주었다. 이러한 발견은 도덕적 정체성이 청소년기 동안 개인의 행동을 형성하는 데 동기화된 추론의 원천으로 작용할 수 있음을 시사한다. 한편, 오이저만과 리(Oyserman & Lee, 2008)는 자아 개념과 친사회적 행동의 관계에 대한 연구를 수행하였다. 그들은 사회적 정체성이 긍정적인 자존감의 원천으로 작용할 수 있고 사회적 정체성이 강한 개인들이 친사회적 행동에 더 많이 참여할 가능성이 있다는 것을 발견했다. 그들의 연구는 또한 강한 사회적 정체성을 가진 개인들이 자신을 더 큰 공동체의 일부로 볼 가능성이 더 높다는 것을 보여주었고, 이것은 더 큰 연결성과 집단에 대한 더 강한 헌신으로 이어질 수 있다. 즉, 자신의 집단의 규범을 기준으로 이에 헌신하고자 하는 방향으로 동기화된 추론이 작동될 수 있다. 또한, 하디와 카를로(Hardy & Carlo, 2011)는 동료 압력에 저항하는 도덕적 정체성의 역할에 대한 연구를 수행했다. 그들은 강한 도덕적 정체성을 가진 사람일수록 도덕적 정체성이 행동에 대한 인지적인 지침의 역할을 하기 때문에, 동료의 압박을 받을 때 일탈적인 행동에 참여할 가능성이 적다는 것을 발견했다. 이 연구는 개인의 행동을 형성하는 데 있어 도덕적 정체성의 중요성을 강조하고 일탈적 행동에 대해 도덕적 정체성이 보호 기능을 할 수 있다는 증거를 제공한다. 이러한 발견은 도덕적 정체성이 동료의 압력에 저항하여 동기화된 추론의 원천으로 작용할 수 있음을 시사한다.

이처럼 사회적 정체성의 일종으로 긍정적인 도덕적 자아상, 즉 자기

자신을 도덕적인 존재로 인식하고자 하는 성향은 동기화된 추론이 도덕적 신념을 보호하기 위한 방향을 갖게 되는 원인으로 작용한다. 동기화된 추론은 사람들이 자신의 도덕적 정체성을 유지하기 위한 과정에서 자신의 정체성에 부합하는 정보를 선택적으로 수용하고 그렇지 않은 것은 무시하거나 왜곡하는 형태로 이해하기 때문이다. 자신의 기존 도덕적 신념과 주어진 상황 사이에서 인지적인 균열이 인지부조화로 이어지는 까닭은 사람들이 일관되고 긍정적인 자아 개념을 갖기를 원한다는 사실에서 비롯된다(Aronson, 1999). 애런슨에 따르면 인지부조화는 도덕적 주체가 동기화된 추론을 통해 자기 개념을 유지할 수 있게 해주는 과정에서 해소된다. 자기개념을 회복하고 자신을 도덕적인 사람으로 계속 여기기 위해서는 다양한 방식으로 합리화함으로써 죄책감이 해소되기 때문이다(Aronson, 1999: 111-112). 이처럼 도덕은 자기 개념의 중요한 차원이 되는 경향이 있다. 도덕 판단에서 동기화된 추론으로 문제가 발생하는 까닭은 이처럼 사람들이 자기 일관성을 유지하기를 원한다는 사실에서 비롯된다. 자기 자신을 스스로 '도덕적 존재'로 인식하고자 하는 방향성 목표에 따라 동기화된 추론이 자기 일관성을 유지하는 근거를 마련해주기 때문이다.

자기 일관성을 향한 인간의 경향으로 인해 사람들은 자신을 구성하는 핵심적인 신념과 가치를 일관되게 유지함으로써 자아 정체성을 형성하고 안정감을 느끼고자 한다. 도덕적 상황을 이해하거나 도덕적 이슈와 관련된 정보나 사실적 근거를 평가할 때도 자기 일관성에 대한 욕구가 도덕 판단을 추동하고 동기화된 추론에 참여하도록 만든다. 추론이 이처럼 자기 일관성에 기여하는 방향으로 동기화되는 또 다른 이유는 감정의 역할에 있다. 사람들은 종종 자신의 신념과 가치에 강한 감정적 애착이 있으며, 이에 따라 그러한 믿음과 가치에 도전하는 정보에 저항할 수 있기 때문이다. 예를 들어, 한 정당의 당원이라고 강력하게 밝힌 사람은 특정 정치적, 윤리적 문제에 대한 정당의 입장과 모순되는 증거를 제시했을 때 분노나 좌절감을 느낄 수 있다. 이를 통해 우리는 기존의 신념과 가치와 일치하는 정보를 받아들이고, 자신의 정체성을 위협하는 정보를 거

부하는 방향으로 추론을 사용한다. 사회인지이론에서 동기화된 추론의 이러한 양상은 정체성 보호 인지의 개념으로 구체화 된다(Cohen, 2003).

정체성 보호 인지는 문화적으로 다양한 개인들이 그들의 집단에서 지배적인 신념에 일치하는 방향으로 증거와 정보를 선택적으로 신뢰하고 무시하는 경향으로, 동기화된 추론이 정체성을 매개로 나타나는 양상(Cohen, 2003; Kahan et al., 2007; Kahan, 2017)이다. 이는 특히 그러한 신념이 자신의 사회적 또는 도덕적 정체성과 밀접하게 연결되어 있을 때 더 강하게 나타난다. 이 개념은 집단의 구성원으로서 정체성을 자각하는 것이, 사회 세계의 거의 모든 종류의 자극에 대한 정보를 사람들이 처리하는 방식에 영향을 미칠 수 있다(Baumeister & Leary 1995: 504)는 사회 심리학 연구의 전제에서 비롯되었다. 개인은 자신에게 영향력 있는 집단 내 구성원들의 공통된 신념을 채택하는 경향이 있다. 또한 상반된 사실적 정보 간의 충돌 상태에서, 특히 자신의 신념을 위협하는 정보가 집단 내 정보보다 집단 외 정보에서 비롯되는 경우, 대부분의 경우 신념을 수정하지 않았다(Kahan et al., 2007: 469-470).

정체성 보호 인지의 존재는 집단 내에서 다양한 형태의 추론을 통해 구성원들이 상호작용하는 양상을 드러내 준다. 특히, 정체성을 보호하고자 하는 목표에 의한 동기화된 추론의 사용은 단지 무의식적이고 직관적인 휴리스틱에만 의존하는 것이 아니며 체계적인 추론을 사용하는 정보처리 과정에도 영향을 미친다(Kahan et al., 2007: 470). 집단 내에서 공유되는 신념은 그 집단 구성원이 상대적으로 높은 수준의 숙고와 비판적인 분석을 특징으로 하는 체계적인 추론을 사용하는 경우에도 영향을 미쳤다(Cohen, 2003). 동기화된 추론은 정체성을 보호하기 위한 수단으로써 사용되며, 이를 통해 개인은 신념을 뒷받침하는 방식으로 정보를 처리하는 경향이 있기 때문에 여러 문제를 불러일으키는 것이다.

정치적 이슈와 관련된 도덕 판단에서 나타나는 동기화된 추론을 연구한 캐헌(Kahan, 2017)은 가짜 뉴스가 광범위하게 파급될 수 있는 요인으로 정체성 보호 인지에 따른 동기화된 추론을 주장한다. 가짜 뉴스와 같이 출처가 불분명하거나 사실과 다른 정보일지라도 개인이나 집단의 정

체성을 긍정하는 것일 때, 동기화된 추론은 그 정보를 진실로 받아들여 정체성을 보호하는 목표를 달성한다. 반대로, 자신의 정체성과 신념이 위협받고 있다고 느낄 때, 동기화된 추론은 이를 보호하기 위한 방향으로 작동하여 자신의 신념과 충돌하는 정보를 의도적으로 혹은 무의식적으로 거부한다는 것이다. 특히, 사람들은 자신의 정체성을 구성함에 있어 중요한 영향을 끼친 집단에서 소외되지 않기 위해 동기화된 추론을 토대로 집단의 가치를 위협하는 사실 정보에 저항한다(Kahan at al, 2007). 이처럼 사람들은 정보를 처리하고 형성하는 기준이 되는신념 체계와 정체성을 이미 가지고 있으며, 정체성 보호 인지를 통해 자신이 누구인지 이해하는 방식으로 정보를 처리하여 신념과 정체성을 유지하고자하는 방향에 의해 동기화된 추론을 사용하는 것이다.

다양한 층위에서 도덕 판단에 영향을 주는 추론을 연구한 모쉬맨은 도덕적 행위자의 판단과 행위가 도덕 원칙에 기반하여 도덕적 추론의 형태로 표현되지만, 본질적으로는 자기본위적인 동기에 의해 도달될 수 있다고 보았다. 이처럼 동기화된 추론은 실제로는 도덕적인 모습이 아닐 수 있음에도, 도덕적 정체성을 훼손하거나 포기하지 않고 유지해낸다(Moshman, 2004a).

도덕적 정체성을 갖는다는 것은 자기-의식적인 도덕적인 행위자로서 자신을 명시적으로 이해하며, 자신이 기반하고 있는 도덕적 신념과 가치를 근거로 행위한다는 것을 의미한다. 모쉬맨에 따르면 자신의 도덕적 문제를 인식하고 성찰하며 그에 따라 행동하는 사람이라면, 인지의 정확성이나 추론의 질적인 측면, 판단과 행동에 대한 정당화 여부와 무관하게 그 사람은 도덕적 정체성을 지니고 있는 것이라고 말할 수 있다(Moshman, 2004a: 89-90). 또한 모쉬맨은 도덕적 행위자로서 자신에 대한 명시적 이론이 '거짓'일 경우를 지적하며, 거짓된 도덕적 정체성(false moral identity) 개념을 제안한다. 거짓된 도덕적 정체성이란 자신의 실제 행동과 일치하지 않는 도덕적 신념과 가치를 채택하여 자기 일관성을 유지하는 현상을 말한다(Moshman, 2004a: 91). 모쉬맨에 따르면 도덕적 정체성은 자아의 중심적 측면이며 그것은 자신이 도덕적인 사람이라는

신념에 뿌리를 두고 있는데, 개인이 자신의 도덕적 성격에 대한 신념이 실제 행동과 일치하지 않을 때 거짓된 도덕적 정체성이 형성될 수 있다.

거짓된 도덕적 정체성은 긍정적인 자아상과 도덕적 평판을 유지하려는 개인의 필요에 기반한다. 모쉬맨에 따르면 거짓된 도덕적 정체성은 개인의 자아에 대한 인식이 도덕적인 사람으로서 자신에 대한 이상화된 자아상에서 파생될 때 발생하지만, 그 자아상은 자신의 실제 행동과 일치하지 않는다. 따라서 거짓된 도덕적 정체성은 도덕적 인지부조화를 초래할 수 있다. 즉, 자신의 도덕적 신념과 가치 사이의 불일치와 실제 행동이 다를 때 우리는 심리적 불편함을 느끼고, 이러한 부조화는 죄책감과 수치심을 유발하기도 한다. 이때, 동기화된 추론은 자신이 생각하는 도덕적 자아상에 뒷받침하는 정보를 찾고 그것을 훼손하는 정보를 무시하거나 왜곡함으로써 도덕적 인지부조화를 해소하고자 하는 양상으로 드러난다.

모쉬맨은 사람들이 '도덕적 정체성을 가진다'는 사실과, '정체성의 진위' 여부를 구분해야하며, 이를 기준으로 도덕적 정체성의 세 가지 범주를 제안한다(Moshman, 2004a: 93-94). 첫 번째 범주는 도덕적 정체성 형성이 불충분한 경우이다. 도덕적 자아 개념을 가지고 있거나 도덕적 추론 능력을 지니고 있음에도 정체성에 대한 강한 헌신이 없는 사람이나, 강한 정체성을 가지고 있지만 그 정체성에 도덕성의 요소가 결여되어 있는 사람이 이에 해당한다. 두 번째 범주는 진정한 도덕적 정체성이 형성된 경우이다. 도덕적 자아를 인격의 중심으로 위치시키고, 도덕성의 요소로 구성된 정체성을 바탕으로 도덕적 헌신을 실제로 수행하는 사람이 이에 해당한다. 마지막으로 세 번째 범주는 거짓된 정체성이 형성된 경우이다. 타인의 권리와 복지에 대해 깊이 고려하고, 이에 따라 행동하는 사람으로 자신을 명시적으로 이론화하지만, 행동과 이론이 모순되는 경우이다. 이러한 사람들은 자신을 근본적인 수준에서 도덕적 행위자로 개념화한다는 점에서는 도덕적 정체성을 가지고 있다고 할 수 있지만, 그것은 거짓된 것이다. 즉, 거짓된 도덕적 정체성이 형성된 사람은 자신을 도덕적이라고 깊이 여기면서도 도덕성에 대한 객관적 헌신을 보이지 않는다(Moshman, 2004a: 94).

거짓된 도덕적 정체성을 가진 개인은 도덕 판단에 있어 도덕적 인격으로서의 자아의식을 유지하기 위한 방향으로 동기부여를 받는다. 동기화된 추론은 그 과정에서 자신이 인식한 도덕적 인격에 대한 신념을 뒷받침하는 정보를 찾지만, 그것에 도전하는 정보는 무시하고 과소평가할 수 있다. 동기화된 추론은 거짓된 도덕적 정체성을 유지하고자 하는 신념과 그렇지 못한 행동 사이의 불일치를 해결하고, 그 불일치로 인해 발생할 수 있는 부정적인 정서를 피하기 위한 시도로 볼 수 있다. 나아가 모쉬맨은 긍정적인 자아상을 유지하고자 하는 목표는 우리가 편향된 방식으로 정보를 처리하도록 동기부여하기 때문에 동기화된 추론은 잘못된 도덕적 정체성에 대한 일반적인 반응이라고 제안한다. 이는 현실과 동떨어져 있고 이상화된 자아상에 기반을 둔 자아개념을 만들어내는 결과를 낳을 수 있다.

한편, 동기화된 추론은 거짓된 도덕적 정체성을 매개로 도덕적 위선을 정당화하고 도덕적, 정치적 불일치의 양극화를 심화시키기도 한다. 모쉬맨은 거짓된 도덕적 정체성이 대량 학살과 테러 등 집단적 폭력 행위를 하는 가해자들의 공통된 특징임을 지적한다(Moshman, 2007). 특히, 그는 집단 학살이나 테러와 같은 집단적 폭력 행위[37]가 사악한 개인이나 집단에 의한 희귀한 행위가 아니라 주로 일반인들에 의해 자행되고 있다는 점에 주목하며, 이들의 폭력 행동의 심리적, 사회적 과정을 추적한다(Moshman, 2004a; 2007; 2009; 2011). 그는 홀로코스트 핵심 수행 요원이었던 아돌프 아이히만의 재판 과정을 지켜보며 '악의 평범성(banality of evil)' 개념을 도출해낸 한나 아렌트(Arendt, 1963)의 철학적 통찰을 비롯하여 타 문화, 타 종교, 타 인종에 대한 무자비한 학살이라는 비도덕적 행위의 주체가 일반인들에 의해 자행되었다는 사실을 뒷받침하는 심리적, 사회적 이론들을 검토하여 이들의 공통된 특징을 세 가지 거짓된

37) 모쉬맨은 집단학살과 같은 집단적 폭력 현상을 '추상적으로 정의된 집단을 목표로 하는 파괴 행위나 과정'으로 정의한다(Moshman, 2007: 115). 이러한 폭력의 구체적인 양상은 다양하지만, 많은 가해자들이 가담하며, '특정 집단'에 대한 폭력적인 행위라는 점에서 단일한 행위나 과정으로 분석될 수 있다. 집단적 폭력 현상은 피해자 집단의 생물학적, 사회적, 문화적으로 형성된 정체성을 훼손하는 행위를 반드시 포함한다(Moshman, 2007: 115-116).

도덕적 정체성 과정으로 분석한다. 세 가지 과정은 ①'우리'-'그들'의 이분화, ②'그들'의 비인간화, ③비도덕적 행위의 부정을 의미한다.

먼저 비도덕적 행위에 가담한 사람들은 가해자와 피해자의 정체성을 '우리'와 '그들'로 이분화한다. 사회적 정체성은 국가, 민족, 인종, 종교, 성별, 직업, 이데올로기 등과 같이 다양하고 중첩적으로 구성됨에도 이러한 일반적인 패턴을 무시하고 '우리'와 '그들'이라는 단일한 차원에 따라 구분 짓는다. 이분화는 정체성의 일부 잠재적 차원을 너무 두드러지게 하거나 심지어 다른 것들이 주변적인 것이 되도록 강제하는 사회적, 문화적 이해의 구성을 포함한다(Moshman, 2007: 114). 동기화된 추론은 가해의 대상이 되는 사람들이 가진 정체성을 소수의 특징적인 것으로 축소하며 도덕적으로 열등한 것으로 분류함으로써 자신의 비도덕적 행위에 대한 도덕적 정당성을 유지하게 해준다.

이어서 비도덕적 행위의 주체는 폭력의 피해자들 즉 '그들'을 비인간화하여 도덕적 의무의 영역 밖에 둔다. 비인간화는 이분화된 '우리'와 '그들'의 도덕적 우열을 구분하는 구체적인 방식이다. 도덕적으로 열등하다고 여겨지는 정체성에 대해 적극적으로 오명을 씌우거나 극단적으로는 인간성을 박탈하는 것이다(Moshman, 2007: 120-121). 자신의 거짓된 도덕적 정체성을 유지하고자 하는 목표에 의해 동기화된 추론은 '우리'와 구분되는 '그들'을 비인간화하는 언어를 찾아낸다. 가령, 비인간화된 집단의 구성원들은 '잡초, 쥐, 해충, 개, 바이러스, 미생물, 기생충, 역병, 암세포' 등으로 다양하게 묘사된다(Moshman, 2007: 122-123). 이러한 비인간화의 공통된 특징은 도덕적 세계관을 '우리' 집단으로 제한한다는 점이다. 따라서 '우리' 집단 내 구성원들은 서로에 대한 의무를 인정하고 존중하는 도덕적 개인이지만, '그들'은 우리와 다를 뿐만 아니라 완전한 사람이 아니며 따라서 우리의 도덕적 의무가 확장되는 범위에 해당하지 않는 것이다. 이처럼 비인간화는 이분화를 강화하고 확장한다.

최종적으로 비도덕적 행위를 수행하고도 자신의 도덕적 자아 개념을 보존하기 위해 거짓된 도덕적 정체성의 주체는 스스로 자행한 행위를 부정하는 방식을 취한다. 비인간화의 방법이 자신의 행동을 '비도덕적 행

위'라고 생각하지 않도록 만드는 동기화된 추론의 양상이라면, 부정은 자신의 행위 자체를 인정하지 않는 방법이다. 대량 학살의 사례에서 폭력의 당사자들은 사실에 대해 거부하거나, 자신이 알고 싶지 않은 정보를 무시하거나, 왜곡하여 기억하거나 선택적으로 기억했다(Moshman, 2009: 123-124). 자신의 비도덕적 행위의 피해를 입은 당사자들의 정체성을 부정하도록 동기부여하는 것은 자신을 도덕적 주체로 인식하는 거짓된 도덕적 정체성이다. 거짓된 도덕적 정체성은 자기 자신을 비도덕적으로 보이게 하거나, 비도덕적인 행위를 증거와 해석을 부정하거나 반박하도록 동기 부여한다.

나아가 역사적으로 극단적인 사례에서뿐만 아니라 일상에서 자행되는 집단 폭력에서도 이와 같은 일반적인 심리 과정이 모든 연령대의 사람들에게서 나타나며, 특히 아동기와 청소년기에서도 두드러진 특징으로 나타난다. 아동 폭력 가해자를 직접적 대상으로 연구한 와인립(Wainryb, 2011)에 따르면 피해의 크기와 관계없이 폭력의 가해는 일반적인 아이들의 도덕적 주체의 문제, 즉 대량 학살 가해자의 공통적인 특징으로 제기된다고 지적한다. 그녀는 심지어 특별한 집단 폭력에 못지않게 이러한 정체성 문제가 아동 및 청소년들에게 나타나고 있음을 실증적으로 주장한다. 아이들은 끔찍한 만행을 저지를 때에도 그들의 경험으로부터 계속 의미를 재구성하고 삶의 서사를 구성하여 자신의 도덕적인 정체성을 유지하고자 한다. 거짓된 도덕적 정체성에 기반한 동기화된 추론은 정체성이 비도덕적인 행위를 하는 토대가 되어준다는 점, 그리고 그러한 행위를 하는 이유가 비도덕적인 정체성에 있는 것이 아니기 때문에 스스로의 정체성을 여전히 도덕적인 것으로 인식한다는 점 등을 시사한다.

모쉬맨은 거짓된 도덕적 정체성이 비도덕적 행위를 합리화하는 세 가지 심리적 과정을 질적으로 구별되는 단계라기보다는 실질적으로 중복된다고 파악한다. 이분화는 비인간화를 가능하게 하지만, 동시에 이분화를 강화한다. '그들'을 파괴하는 비도덕적 행위와 피해자에 대한 지속적인 부정은 비인간화를 확대하고 이분화를 정당화해준다. 이와 같이 거짓된 도덕적 정체성을 매개로 작동하는 동기화된 추론을 통해 비도덕적인 행

위를 하는 사람들은 도덕적 행위자로서의 지위를 유지할 수 있는 근거를 계속해서 제공받는다.

이처럼 도덕 판단에서 동기화된 추론의 문제는 도덕적 추론이 자신의 도덕적 신념을 유지하고자 하는 강한 방향성에 도달하고자 하는 성향에서 비롯된다. 이를 통해 도덕적 주체는 상황적 단서와 정보에 대한 정확한 평가보다는 자신에게 유리한 방식으로 왜곡하고 해석한다. 비도덕적 행위를 수행한 당사자는 동기화된 추론이라는 인지적 전략을 사용하여 도덕적 주체로서의 지위를 유지한다. 또한 도덕적 명분을 내세우며 자신의 집단과 구분되는 타 집단에 대해 배타적인 태도를 견지하기도 한다.

2. 편향의 사각지대 현상

도덕 판단에서 동기화된 추론의 사용은 사람들이 자신의 추론이 인식론적으로 올바른 결론에 도달했다고 생각하며, 그 결론이 정당하다고 진심으로 믿게 한다(Ellis, 2022). 이러한 원인은 편향의 사각지대(the bias blind spot) 개념을 통해 적절히 이해될 수 있다. 편향의 사각지대 현상은 모든 사람이 인지적 편향을 보임에도 불구하고, 다른 사람들에 비해 자신을 덜 편향적으로 인식하는 경향으로, 행동 윤리학 영역에서 고안된 개념을 말한다(Pronin et al., 2002; Scopelliti et al., 2015).

편향의 사각지대 현상을 처음 발견한 연구(Pronin et al., 2002)는 인지편향에 대한 자기보고식 평가에서 자신에게 유리한 방향으로 발생한다는 사실을 밝히기 위해 여러 실험을 수행했다. 해당 연구에 따르면 실험 참가자들은 여러 가지 동기적 편향이 그들 자신에게서 보다 남들에게서 훨씬 더 만연하다고 믿는다. 첫 번째 실험에서, 실험 참가자들은 다양한 인지적 편견에 대한 민감성을 포함하여, 여러 특성에 대해 자신과 동료들을 평가하도록 요청받았다. 그 결과는 참가자들이 그들의 동료들보다 편견에 덜 민감하다고 스스로를 평가했다는 것을 보여주었다. 또 다른 실험에서, 참가자들은 그들과 동료들이 단어 목록을 기억하는 것과 같은 다양한 인지적 작업에서 어떻게 수행할 것인지 예측하도록 요청받았다.

이번에도 참가자들은 실제 성과가 동급임에도 불구하고 또래보다 더 좋은 성과를 낼 것으로 예측했다. 연구자들은 편향의 사각지대가 개인이 객관적 증거에 의해 보장되는 것보다 더 유리한 관점에서 자신을 보는 내 편 편향의 한 형태라고 결론지었다. 이러한 편향은 자신의 능력과 판단을 과신하게 함으로써, 자신의 인식과 결정에 편견이 미치는 영향을 과소평가하게 할 수 있다.

이어 스코펠리티 등(2015)은 편향 사각지대와 동기화된 추론의 관계를 탐구하였다. 연구자들은 편향 사각지대가 자신의 신념을 뒷받침하는 정보를 찾고 이에 모순되는 정보를 무시하는 것을 포함하는 동기화된 추론의 일반적인 형태인 확증 편향과 관련이 있다는 것을 입증하기 위해 여러 실험을 수행했다. 한 실험에서, 참가자들은 기후 변화와 같은 논쟁적인 문제와 관련된 증거를 평가하도록 요청받았다. 그 결과, 자신을 확증 편향에 덜 민감하다고 평가한 사람들은 그것에 의해 영향을 받을 가능성이 더 높았고, 그들의 믿음을 확인하는 정보를 찾는 것으로 나타났다. 또 다른 실험에서, 참가자들은 그들과 그들의 동료들이 논쟁적인 문제에 대한 증거를 어떻게 평가할 것인지 예측하도록 요청받았다. 이번에도 참가자들은 실제 결과가 다르게 나타났음에도 불구하고 또래에 비해 확증 편향에 덜 취약할 것으로 예측했다. 연구자들은 편향의 사각지대가 개인들로 하여금 편견에 덜 민감하다고 믿게 하고, 그들의 신념과 결정에 더 자신감을 느끼게 함으로써 동기화된 추론을 악화시킬 수 있다고 결론지었다. 이것은 확고한 신념과 부정확한 정보에 입각한 의사결정의 악순환을 초래할 수 있다. 이처럼 편향의 사각지대는 도덕적 상황에서 문제가 있는 의사결정으로 이어질 수 있는데, 이는 이러한 견해를 가진 개인들이 인지 편향에 영향을 크게 받으면서도 객관적이고 편견 없는 결정을 내린다고 믿기 때문이다. 특히, 동기화된 추론의 맥락에서 편향 사각지대는 개인이 타인의 편향은 엄격하게 인식하면서도 자신의 편향은 느슨하게 인식함으로써 여러 문제를 야기할 수 있다.

먼저, 편향의 사각지대에 빠진 사람들은 자신의 도덕 판단에 더 큰 자신감을 갖고 자신의 신념과 모순되는 정보를 찾을 가능성이 적다. 따라

서 개인이 자신의 신념을 뒷받침하는 정보만 찾고 그에 반하는 정보는 무시하는 확증 편향을 초래할 수 있다. 특히, 이때의 확증 편향은 우리 자신의 기존 신념과 태도에 우호적인 방식으로 증거를 평가하고 생성한다는 점에서 내 편 편향과 유사하다(Stanovich, 2021/2022: 32-39). 편향의 사각지대는 동기화된 추론의 양상인 확증 편향과 내 편 편향의 결과로, 타인의 판단이 우리의 판단과 다를 때 그 차이가 증거에 대한 서로 다른 타당한 해석의 차이 때문이 아니라 타인이 편향적이기 때문이라고 생각하게 만든다(Stanovich, 2021/2022: 175). 가령, 특정 집단에 편향된 개인은 객관적인 결정을 하고 있다는 자신의 신념을 강화하면서, 그들의 판단이 부당하다는 것을 암시하는 증거를 무시하게 된다. 이것은 비윤리적이거나 심지어 차별적인 의사결정으로 이어질 수 있다.

또한, 편향의 사각지대에 빠진 사람은 자신의 동기화된 추론을 의식할 가능성이 더 낮아진다. 그들은 자신이 편향에 영향을 받지 않는다고 믿기 때문에 자기 성찰과 자기 인식에 참여할 가능성이 낮으며, 자신의 편향성이 의사결정에 어떻게 영향을 미치는지 고려할 가능성도 낮기 때문이다. 앞서 살펴본 비도덕적 행위의 정당화 문제나, 도덕적, 정치적 양극화의 심화 현상은 편향의 사각지대로 인해 무의식중에 발생할 수 있는 것이다. 즉, 사람들은 자신의 판단에 대한 객관성을 과대평가하고, 타인의 판단에 대한 객관성을 과소평가함으로써 자신의 비도덕적 행위를 정당화하거나, 자신과 배타적인 타인 및 집단에 대해서는 쉽게 비난하면서도 의식하지 못하게 되는 것이다. 동기화된 추론과 편향의 사각지대와의 관계에 관한 연구에서(Kriegeskorte et al., 2009) 무의식적으로 편향된 신념을 가진 사람들이 그들의 신념과 태도를 유지하기 위해 상충하는 정보를 제시했을 때 동기화된 추론을 더 많이 보인다는 것을 발견했다.

나아가 편향의 사각지대는 자신의 편향을 인식하지 못하고 도덕 판단에서 동기화된 추론에 참여하는 개인이 불평등과 차별을 영속화하는 결정을 내릴 가능성이 높기 때문에, 개인 간의 관계뿐 아니라 사회적 측면에도 상당한 부정적 영향을 미칠 수 있다. 자신의 편향을 인식하지 못하면서, 스스로를 객관적이라 생각하는 집단의 의사결정권자가 무의식적으

로 자신의 정서와 신념이 의사결정에 영향을 미치도록 허용한다면, 소외된 공동체에 해를 끼치는 정책과 결정으로 이어질 수 있기 때문이다. 이처럼 일반적인 대중과 달리 소위 사회 집단의 엘리트층이 편향의 사각지대에 빠지게 되면, 그 파급력이 더 커짐을 알 수 있다.

동기화된 추론 연구자들은 이를 인지적 엘리트[38](cognitive elite)의 문제라 부른다. 인지적 엘리트의 문제란 인지능력과 교육수준이 높은 개인이 자신의 판단과 의사결정을 과신할 가능성이 높고, 자신의 편견이 인식에 미치는 영향을 덜 인식하는 현상(Ditto et al., 2019; Mooney, 2012)을 말하는데, 인지적 엘리트들에게서 유독 광대한 편향의 사각지대가 조성될 가능성이 있다(Stanovich, 2021/2022: 182). 편향의 사각지대에 빠진 인지적 엘리트들은 동기화된 추론을 통해 그럴싸한 도덕적 논증들을 만들어내고, 일반 대중들은 이러한 인지적 엘리트들이 내놓은 편향된 주장을 광범위하게 소비하게 된다. 이처럼 도덕 판단에서 동기화된 추론은 편향의 사각지대와 인지적 엘리트의 인지 편향 문제를 통해 더욱 확대되어, 개인적 차원의 비도덕성을 넘어 사회적 차원의 문제를 만든다.

스타노비치 연구(Stanovich, 2021/2022: 177)에 따르면 편향의 사각지대와 인지적 성숙 사이의 정적인 상관관계가 있었다. 즉, 교육 수준이 높고 인지적으로 더욱 성숙한 실험 참가자들이 그렇지 못한 참가자들보다 편향의 사각지대를 가질 가능성이 더 크다는 것이다. 이는 다른 인지 편향과 동기화된 추론으로 발생하는 확증 편향 및 내 편 편향의 차이를 설명해준다. 교육을 많이 받고 이성적인 사람일수록 편향에 빠질 가능성이

38) 인지적 엘리트(cognitive elite)라는 용어는 심리학자 헌스타인(R. J. Herrnstein)과 정치학자 머레이(C. Murray)가 1994년에 집필한 저서(The Bell Curve: Intelligence and Class Structure in American Life)에서 미국 사회 내의 인지 계층화 현상을 설명하며 사용한 표현이다. 인지적 엘리트란 높은 수준의 교육을 통해 사회 주요 직업군에서 활동하며 계층화 현상을 공고히 하는 사회-경제적 상류층을 의미한다(Herrnstein & Murray, 1994: 25-37). 해당 저서는 인간의 지능이 유전적 요인과 환경적 요인 모두에 의해 실질적으로 영향을 받으며, 높은 지능이 '성공적인 삶'이라고 불리는 일반적인 지표들과 양의 상관관계가 있음을 실증적으로 보였다. 특히, '인지적 엘리트'가 정규 분포 속 평균 및 그 이하의 지능을 가진 사람들과 점점 분리되고 있으며, 사회-경제적 지위가 이들에게 쏠리는 현상이 나타나며, 이와 같은 인지 능력의 차이가 당시 미국 내의 사회적 분열의 원인임을 지적하였다.

낮거나, 정확성을 목표로 추론을 보일 가능성이 높다는 상식과는 반하는 결과이다. 실제로 대부분의 휴리스틱 관련 연구들에서 인지 편향은 인지적 능력과 부적 상관관계임을 살펴볼 수 있다(Stanovich, 2021/2022: 176). 인지적 엘리트들은 스스로를 덜 편향적이라고 믿고, 실제로도 일반적으로 덜 편향적이다. 그러나 정확히 같은 이유 때문에 이들은 편향의 사각지대에 더욱 빠지기 쉽다. 즉, 이성적 사유를 더 잘하거나, 수준 높은 논리 교육이나 과학 교육을 받은 인지적 엘리트들은 스스로를 비판적이고, 합리적이고 객관적으로 생각하는 자신의 능력을 과대평가하는 경향이 더 높다. 따라서 자신의 신념과 모순되는 정보를 '논리적으로 틀린 것'이라는 근거로 무시하게 되므로, 아이러니하게도 편향의 사각지대에 더 쉽게 빠지게 되는 것이다. 인지 엘리트들은 이와 같은 동기화된 추론을 사용하여 자신의 높은 교육 수준에 따른 지능과 인지 성숙도 덕분에 자신이 당연히 다른 사람들에 비해 편향적 사고를 덜 한다고 믿는다(Stanovich, 2021/2022: 177).

나아가 이들은 때로 과학적 사실을 거부하면서까지 자신의 신념을 유지하기 위한 논증을 생성해내고, 이는 대중에 의해 지지를 얻기도 한다. 인지적 엘리트가 만들어낸 동기화된 추론은 특히, 윤리적 이슈와 과학적 사실이 강하게 결부되어 있는 응용윤리의 여러 분야- 가령, 기후 변화, 원자력 발전, 줄기세포 연구- 등의 맥락에서 과학적 증거와 사실을 거부하는 결과를 초래할 수 있다. 이러한 경우에, 비록 그것이 많은 과학적 증거에 의해 잘 확립되고 지지를 받더라도, 개인은 그들의 도덕적 또는 정치적 신념에 반하는 과학적 증거를 거부할 수 있다. 따라서 동기화된 추론과 관련하여 인지적 엘리트 문제는 잘못된 신념을 갖게 되거나, 자신의 행동을 합리화하고 사실적인 정보를 왜곡하여 인식하는 것을 지식과 지능의 결핍에서 찾는 것은 적절하지 않다는 것을 시사한다.

인지적 엘리트의 동기화된 추론 문제에 주목한 크리스 무니는 이를 '똑똑한 바보(smart idiot)' 효과로 비유한다(Mooney, 2012). '똑똑한 바보' 효과는 특정 분야나 주제에 대한 식견이 높고 지적인 사람(인지적 엘리트)이지만 삶의 다른 영역에서는 어리석은 행동을 보이거나 잘못된

결정을 내릴 수 있는 현상이다. 다시 말해, 비록 한 사람이 한 분야에서 지적이고 지식이 있다고 여겨지더라도, 동시에 잘못된 판단을 내릴 수 있다는 것이다. 심지어 가장 똑똑한 사람들도 체계적인 사고의 오류인 인지 편향을 보이기 쉬워서, 그들이 이성적이지 않은 판단을 내리거나, 정확성이 아닌 최선의 이익이라는 방향성에 치우친 결정을 내리도록 이끌 수 있다. 낙태와 같은 윤리적 이슈와 관련된 특정 정책에 대해 반감을 가진 사람들의 지식수준이 낮다면 그들은 이유 없이 반대할 것이다. 그러나 만약 반대자의 교육 및 지식의 수준이 높다면 그들은 그럴싸한 반론을 제기하며 반대할 것이다. 이 반론은 정서가 실려 있기 때문에 말 그대로 그 사람의 일부로서 작용하게 된다. 인지적 엘리트가 특정 윤리적 이슈와 정책에 대해 갖는 의견(지지나 반론)은 더 정교하게 스스로를 설득하게 된다. 그 결과 자신의 신념에 반하는 사실적 차원의 정보가 제시되는 경우에도 그가 가진 초기의 신념을 오히려 강화하는 방향으로 동기화된 추론을 발휘한다.

게다가, 자기 확신과 우월감 또한 똑똑한 바보 현상을 조장할 수 있다. 인지적 엘리트들은 자신의 지식과 전문성이 삶의 모든 분야에 적용될 수 있다고 믿기도 하며, 자신의 견해가 일반 대중들 사이에서 유력한 견해로 자리 잡을 수 있음을 인지하고 있다. 이와 같은 똑똑한 바보 효과는 인지적 엘리트에 의해 발생한 도덕적, 정치적 의견 격차가 일반 대중들 사이에 퍼지게 되며 양극화를 더욱 심화시키는 문제와도 관련됨을 시사한다. 무니는 주로 보수주의적 입장을 취하는 공화당원들의 동기화된 추론 양상을 분석함으로써 인지적 엘리트 문제의 원인을 규명하고자 한다. 교육을 많이 받은 보수주의자일수록 보수적 뉴스나 견해의 소비량이 많을 것이고, 따라서 자연스럽게 특정 이슈에 대한 다른 보수주의자들의 견해와 근거들을 수집할 기회를 많이 얻게 된다. 자신의 견해를 지지하는 논거나 이유에 대해서도 더 익숙하고, 이에 대한 반박 및 재반박의 논리적 구조에 대한 이해도도 높을 것이다. 이들은 그런 논거들을 자신의 정체성과 뇌의 일부로 받아들였으며, 그 과정에서 특정 사실과 정치적 가치관 사이에 강한 감정적 연결망을 구성하게 되는 것이다(Mooney,

2012: 77).

앞서 살펴본 케이헌 연구진이 미국인을 대상으로 수행한 정치적 입장과 기후변화에 대한 신념 연구에서 흥미로운 발견이 있다. 진보적인 민주당 지지자들은 기후 변화의 심각성을 제기한 정보를 적극적으로 수용했고, 보수적인 공화당 지지자들은 이러한 정보를 왜곡하여 이해하거나 자료의 출처를 의심하는 방식으로 거부하였다. 더 나아가 케이헌은 다양한 질문을 바탕으로 실험 참가자들의 과학 지식을 측정하여 이들의 인지적인 능력과 동기화된 추론의 정도를 확인하고자 하였다(Kahan, 2015). 가령, '위젯 5개를 만드는 데 기계 5개로 5분이 걸린다면, 기계 100개로 위젯 100개를 만드는 데는 얼마의 시간이 필요한가?' 등의 질문을 통해 응답자들의 사고 능력을, '레이저는 음파를 집중시켜서 작동한다. 참인가 거짓인가?' 등의 질문을 통해 기초 과학 지식을 물었다.

지식이나 지능을 충분히 갖출 때, 방향성에 의한 동기화된 추론을 극복하고 정확하고 객관적인 추론을 한다면 과학적 지식을 많이 알수록 기후 변화와 관련된 과학적 질문에 관한 답변에 이견이 없으리라 예측할 수 있다. 그러나 케이헌의 후속 연구(Kahan, 2017)는 정반대의 결과를 발견했다. 과학적 지식이 낮은 집단일수록 오히려 의견 일치 비율이 높게 나타났으며, 진보와 보수 모두 33%가 지구온난화의 원인이 인간이라고 말했다. 오히려 과학적 지식이 높은 집단일수록 정치적 입장이 지구온난화에 대한 견해를 구분 짓는 주요 변수로 작용하였다. 과학적 지식이 가장 높은 집단에서 지구온난화의 원인이 '인간'이라고 답한 이들은 진보주의자의 경우 100%였지만, 보수주의자는 20%였다(Kahan, 2017: 1011-1012).

이러한 현상은 기후 변화 이외에도 응용윤리학의 주요 쟁점과 관련된 과학적 사실-줄기세포 연구, 인간 진화에 대한 관점 등-에도 적용되었다(Drummond & Fischhoff, 2017). 해당 연구에서는 과학교육과 리터러시에 있어 높은 수준을 보인 사람들이 오히려 과학과 관련된 주요 쟁점에 대해 심각한 이분법적인 도덕적 신념을 가지고 있음을 확인했다. 이들은 자신의 전문적 지식을 활용하여 입장을 정당화하는 데 더 적극적으로 동

기화된 추론을 사용했다(Drummond & Fischhoff, 2017).

유사한 연구에서도 이와 같은 결론이 도출되었다(Pasek, 2018). 미국인들을 대상으로 수행된 동기화된 추론 연구에서, 과학적 합의에 대한 인식과 모순되는 신념을 주장하는 정도에 영향을 주는 변수는 (도덕적 이슈와 관련된) 종교적 신념과 당파성이었다. 사람들은 자신이 속한 종교나 당파적 입장이 지지하는 도덕적 신념에 근거하여 과학적 사실을 왜곡하여 이해한다. 나아가 이들이 동기화된 추론을 통해 생성해낸 논증은 과학적 사실에 위배된다 하더라도 사회적 지지를 얻기 때문에 신념이 유지될 수 있다. 인지적 엘리트는 이와 같은 방식을 통해 일반 대중에게까지 특정 집단의 신념을 전파하는 데 주요한 역할을 수행한다. 즉, 똑똑한 바보 효과가 기후 변화 및 진화 등 많은 과학적 문제의 고도로 양극화되고 정치적인 특성으로 인해 악화된다. 개인은 자신의 정치적 신념에 반하는 과학적 증거가 잘 확립되고 많은 과학적 증거에 의해 뒷받침되더라도 거부할 가능성이 더 크다. 이처럼 과학적 지식 질문에 대해 오답을 말하는 것은 오랫동안 과학적인 문맹으로 여겨져 왔지만, 동기화된 추론 연구가 축적됨에 따라 무지보다는 동기화된 추론이 오답의 많은 부분을 설명할 수 있게 되었다. 나아가 이는 오히려 교육수준이 높을수록 도덕적, 정치적 신념의 양극화가 증가하는 현상 또한 설명해준다[39].

디토 등에 의해 동기화된 도덕적 추론이라는 개념이 정립되기 이전에, 도덕 판단이 동기화된 추론으로 인해 근본적으로 편향될 가능성을 가지고 있음을 주장한 발데솔로와 데스테노는 두 가지 연구를 통해 동기화된 추론 문제의 원인을 분석한다(Valdesolo & DeSteno, 2007; 2008). 연구진은 도덕 판단에서 동기화된 추론으로 발생하는 도덕적 위선의 본질이 편향의 사각지대라고 보았다. 즉, 개인이 자신의 위반을 다른 사람들과 비교하여 훨씬 관대하게 평가하기 때문에 동기화된 추론으로 인한 문제가 발생한다고 분석한다. 나아가 이들은 편향의 사각지대로 인한 비도덕

39) 또한 해당 연구는 정체성과 지식의 차이가 동기화된 추론을 매개로 신념 양극화를 심화시키는 양상을 제시한다(Pasek, 2018: 800-801). 동기화된 추론에는 강한 정체성, 전문적 지식, 신념을 방어할 수 있는 지적 능력, 자신의 지식에 대한 자신감 등의 영향이 반영되기 때문이다.

적 행위의 정당화 문제가 집단 수준으로 확장되는 경우 도덕적, 정치적 불일치의 양극화가 심화될 것이라고 보았다.

이를 검증하기 위해 첫 번째 연구에서 연구진은 도덕적 원칙의 유연한 적용 즉, 동기화된 도덕적 추론의 존재를 파악하고 도덕적 위선이 집단 수준으로 확장되는 양상을 확인했다(Valdesolo & DeSteno, 2007). 이들은 도덕성이 본질적으로 사회적인 것이라는 사회 심리학적 전제를 기반으로, 도덕 판단이 집단 내의 평판 문제와 직결되어 있기 때문에 사람들은 주변 사람들보다 자신의 도덕성을 높게 평가받기 위해 동기부여 된다는 사실에 주목한다. 해당 연구에서 실험 참가자들은 자신과 집단 내 구성원들의 공정성 위반을 집단 밖 사람들의 동일한 위반과 다르게 판단한다는 것을 보여주었다[40]. 자신과의 관계의 함수로부터 다른 사람들에게 확장되었다. 집단 내 구성원에 의해 범해진 도덕적 위반은 자신의 동일한 위반과 동등하게 수용되었으나 집단 밖 타인이 범한 동일한 위반에 대해서는 거부되며 집단 간 관계에 따른 도덕 판단의 비대칭 양상을 확인하였다

후속 연구에서는 이와 같은 편향된 도덕 판단의 원인이 동기화된 추론에 있음을 확인하였다(Valdesolo & DeSteno, 2008). 연구진은 자신, 집단 내 구성원, 집단 밖 낯선 타인에 대한 우리의 감정과 생각이 상호작용하는 양상을 확인하여 도덕 판단의 편향, 즉 도덕적 위선의 원인을 파악하고자 했다. 특히, 이들은 도덕적 위선이 자동적인 직관에 의해 발생되는 것인지, 의식적인 숙고를 거친 동기화된 추론의 영향에 의한 것인지를 구분하여 검증하고자 하였다. 이를 검토하기 위해 연구진은 인지적 제약의 조작을 도덕적 위반에 대한 판단과 교차하는 요인으로 연구를 설계했

40) 실험의 방법은 다음과 같다(Valdesolo & DeSteno, 2007: 689-690). 76명의 참가자는 네 가지 실험 조건 중 하나에 무작위로 할당된다. 각 조건은 동일한 행위에 대해 공정성을 판단하도록 요청했다. 실험 참가자들은 두 가지 다른 유형의 과제 중 하나를 풀어야 하며, 과제를 완료하는 데 걸리는 시간과 결과에 대해 보고해야 한다는 지침을 듣는다. 녹색 과제는 약 10분 소요되며 짧은 사진 검색을 포함한 간단한 설문 조사이다. 반면 빨간색 과제)은 약 45분 소요되며 다소 지루한 정신적 부담이 큰 수학과 논리 문제들로 구성된다. 이를 바탕으로 연구진은 실험 참가자들이 자신의 도덕적 위반, 자신의 집단 내 구성원의 도덕적 위반, 집단 밖 타인의 도덕적 위반에 대해 각각 도덕적으로 평가하도록 요구했다.

다. 만약 인지적 제약의 조작이 참가자 자신의 위반에 대한 판단에 영향을 미치지 않는다면, 이 모델은 편향된 자동 직관에서 위선이 발생하는 모델임을 시사할 것이다. 그러나 인지적 제약의 증가가 참가자 자신의 위반에 대해 그 행동이 불공평하다고 판단하여 도덕적 위선을 약화시킨다면, 이러한 발견은 도덕적 위선의 정당화를 목표로 하는 모순된 의지적 노력에서 발생한다는 것을 암시할 것이다(Valdesolo & DeSteno, 2008: 1335). 먼저, 통제 조건에서 부당하게 행동한 후 자신의 행동을 판단한 참가자들의 공정성 판단(즉, 개인적 선택을 통해 자신이 선호하는 업무와 다른 덜 선호하는 업무를 할당)과 동일한 방식으로 부당하게 행동하기로 한 집단 내 구성원의 판단 간의 차이를 검증하여 이전 연구에서 확인된 일반적인 도덕적 위선의 효과)와 일관된 결과를 얻었다(Valdesolo & DeSteno, 2008: 1335-1336). 또한 인지적 제약의 증가로 인해 참가자 자신의 위반에 대해 더 불공평한 것으로 도덕 판단을 내림으로써 도덕적 위선의 정당화 경향이 줄어든다면, 이는 도덕적 위선이 다른 사람들과 상대적으로 자신에 의해 범해진 도덕적 위반을 정당화하는 것을 목표로 하는 모순된 의지적 노력에서 발생한다는 것을 암시할 것이다. 연구 결과, 도덕적 위선이 공정성 위반의 사고에 대한 부정적인 정서적 반응이 의사결정을 중재하는 인지적 제약을 통해 감소되었음을 확인하였다. 그러나 이는 주로 타인의 판단에 대한 편향을 중재하는 데 효과가 있었다. 이와 달리 자신의 비도덕적 행위를 정당화하고자 하는 동기는 초기의 부정적인 정서를 약화시키고 오히려 더 관대한 판단으로 이어졌다. 즉, 동기화된 추론의 과정은 부정적 반응을 더 인과적으로 강력하게 만들고 도덕적 위반을 더욱 효과적으로 정당화하는데 사용된 것이다. 이는 도덕적 추론이 맥락 의존적 성격으로 인해 동기화된 추론의 일종으로 분석될 수 있다는 이들의 이전 연구(Valdesolo & DeSteno, 2006)와도 일치했다. 도덕 판단은 결과적으로 자신의 도덕적 평판을 높이는 기능을 수행하며 그 과정은 근본적으로 편향에 취약하다.

이처럼 편향의 사각지대로 인해 발생하는 동기화된 추론을 통한 비도덕적 행위의 정당화 문제는 집단 수준에서 자신이 속한 집단의 동료 구

성원들에게도 자신의 행동을 정당화하는 것과 동일한 도덕적 유연성을 부여하기도 한다. 도덕적 위선은 집단 차원의 사회적 정체성으로 인해 더욱 쉽게 확산될 수 있다. 따라서 동기화된 추론의 문제를 개선하기 위해서는 근본적으로 편향의 사각지대를 경계하고 극복해야 한다.

제 4 장 도덕 판단에서 동기화된 추론의 개선방안

제 1 절 동기화된 추론의 개선 가능성

1. 도덕적 신념의 가변성

도덕적 주체로서 우리는 삶의 경험 속에서 나름의 도덕적 신념을 세우고 이를 바탕으로 판단을 내리고 행동을 동기화한다. 도덕적 신념은 개인의 가치관, 문화적 배경, 양육 환경 및 개인적 경험에 깊게 뿌리를 두고 있기 때문에 개인의 정체성을 이루는 중심이 되기도 한다. 따라서 신념을 위협하는 정보나 근거를 마주했을 때 사람들은 자신의 정체성에 대한 위협을 느끼기도 한다(Cusimano & Lombrozo, 2021; Ditto et al., 2020). 이처럼 동기화된 추론의 문제는 그것의 방향이 자신을 향하며 그 중심에는 개인의 도덕적 신념을 보호하고 긍정적 자아상을 유지하고자 하는 마음의 경향에서부터 유래한다.

신념은 때로는 깊은 숙고나 합리적인 생각을 거치지 않은 채 형성될 수 있다. 자신과 동일시하는 사람이나 집단의 가치와 규범 체계를 무비판적으로 수용할 수 있고, 오랜 역사와 전통을 거쳐 형성된 공동체의 가치 체계 속에서 일상을 살아가면서 자신도 모르게 습득되기도 한다. 특히, 우리의 도덕적 신념과 가치는 진공상태에서 형성되는 것이 아니라 자신이 속하거나 준거로 삼는 사회집단의 규범과 가치에 의해 영향을 받는다. 따라서 사람들은 비록 질적으로 동등한 도덕적 논증이 주어지더라도, 그들의 기존 도덕적 신념을 지지하는 주장을 받아들이고 그들과 모순되는 주장을 거부할 가능성이 더 높은 것이다. 도덕 판단의 맥락에서, 이것은 사람들이 객관적인 증거나 논리를 찾는 것보다 그들의 도덕적 신념을 옹호하는 것에 더 동기부여가 될 수 있다는 것을 의미한다.

도덕 판단에서 동기화된 추론의 존재를 지적한 하이트는 도덕 판단을

형성하는 일련의 도덕적 기반들을 도덕 매트릭스(moral matrix)로 표현하기도 했다. 도덕 매트릭스는 개인의 도덕적 신념에 가장 강력하고 지배적인 영향을 주는 요인이다.

> *도덕 매트릭스는 사람들을 하나로 묶어주지만, 그것은 다른 매트릭스가 가진 논리를 못 보게 하는 면이 있다. 이 때문에 사람들은 세상에 하나 이상의 도덕적 진실이 있다는 사실을 헤아리는 데 무척이나 어려움을 느낀다. 사람을 판단하거나 사회를 운영하는 정당한 틀도 하나 이상 있을 수 있다는 사실 역시 마찬가지이다(하이트, 2012/2014: 215).*

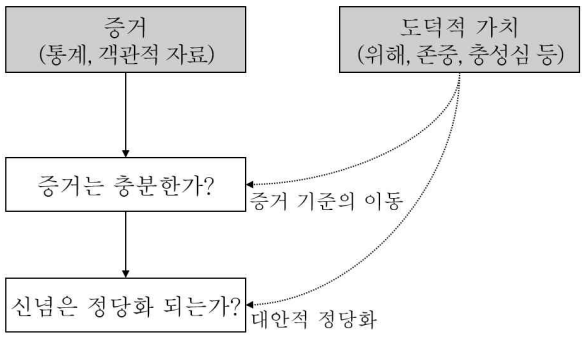
> *어떤 것이든 사람들을 하나의 도덕 매트릭스로 엮을 수 있기만 하면, 그리하여 내부 집단은 미화하고 동시에 타 집단은 악으로 몰 수 있기만 하면, 거기에서 도덕을 내세운 살인은 얼마든지 나올 수 있다(하이트, 2012/2014: 478).*

영화 「매트릭스」속 인물들이 현실 세계로 빠져나오지 못하고 거짓 세계가 주는 환상과 위안 속에 영원히 머무는 선택을 하는 것처럼, 자신이 속한 도덕 매트릭스에 갇히게 되면 마치 그곳의 세계만이 진정한 세계인 양 착각하게 되고 다른 도덕 매트릭스의 존재를 부정할 수 있다. 즉, 우리는 자신의 신념이 어떻게 형성되었는지 검토할 필요가 있으며, 다양한 도덕 매트릭스의 존재를 인정함으로써 자신이 가지고 있는 신념을 수정해나가야 한다.

동기화된 추론 연구가인 디토 등은 사람들이 도덕적 신념과 가치에 부합하는 세상에서 살기를 열망한다는 고전적 연구의 성과에 빗대어(Lerner & Miller, 1978) 정의로운 사회에서 살고자 하는 본성적 도덕감(moral sense)을 설명한다(Ditto et al., 2009). 아담 스미스에 따르면, 도덕감이란 개인이 도덕적 평가와 판단을 할 수 있게 해주는 자연스럽고 보편적인 인간의 능력을 의미한다(Smith, 1798/2009). 그리고 이는 도덕적 상황에 대한 자연스럽고 직관적이며 즉각적인 반응을 제공한다. 사람들은 좋은 도덕감을 추구하고, 이것이 널리 인정되는 사회를 갈망한다. 러너와 밀러는 이러한 도덕감 개념을 바탕으로 정의로운 사회 추구에 대한 사람들의 열망을 실험 심리학적으로 입증하고자 하였다(Lerner &

Miller, 1978). 해당 연구는 정의로운 세상에 대한 사람들의 신념과 욕망이 주변에서 일어나는 사건과 행동에 대한 인식에 어떤 영향을 미치는지 살펴보았는데, 사람들은 세상이 공정하고 예측 가능하며 통제 가능하다고 믿는 강한 선호를 가지고 있으며, 이러한 선호는 사건의 원인과 개인과 집단의 도덕적 특성에 대한 귀인을 만드는 방식에 영향을 미친다고 주장한다. 이는 우리가 자신의 도덕적 가치를 믿고 싶은 강한 선호와 욕구를 가져야 한다는 사실과 동시에, 자신과 동일시하는 집단 내 타인이 훌륭한 도덕적 행위자라고 믿도록 동기 부여되어야 함을 암시한다(Ditto et al., 2009: 313). 이와 같은 맥락에서 우리가 부정적으로 보는 사람들과 그러한 집단에 대해서는 전반적으로 부정적인 정서와 일치하는 도덕적 특성을 보여준다고 생각하게 되는 것이다.

도덕 판단에서 동기화된 추론의 존재에 주목한 쿠시마노 연구진은 사람들의 도덕적 신념이 그가 속한 사회가 중시하는 도덕적 가치를 어떻게 반영하는지를 분석하여 동기화된 추론의 근거가 되는 신념 또한 변화할 수 있음을 보이고자 했다(Cusimano & Lombrozo, 2021). 도덕적 가치가 신념에 영향을 미치는 양상을 자세히 보면 다음과 같다. 도덕적 가치는 도덕 판단에 바탕이 되는 도덕적 신념에 대한 충분한 근거로 여겨지는 증거들을 바꾸거나(증거 기준 이동 가설), 도덕적 고려 사항을 그러한 증거와 직접 비교하여 도덕적 신념에 대한 대안적인 정당성(Cusimano & Lombrozo, 2021: 3)을 제공하는(대안적 정당화 가설) 방식으로 나타난다. 다음의 그림은 도덕적 가치의 특성으로 인해 도덕 판단에서 동기화된 추론이 유발되는 두 가지 양상을 나타낸다.



[그림 4-1] 도덕적 가치가 신념에 영향을 미치는 두 가지 방식

증거 기준 이동 가설은 도덕성이 도덕적으로 유익한 신념과 위험한 신념에 대해 이중적인 기준을 만들어낸다는 것을 시사하고, 대안적 정당화 가설은 도덕적 이익이 동기화된 도덕적 신념으로 이어진다고 가정한다. 도덕적 고려가 충분한 증거를 구성하는 것에 영향을 미칠 수 있다는 아이디어는 이전의 여러 연구를 통해 입증되었다. 예를 들어, 결정을 빨리 내려야 할 때, 사람들은 제한된 증거에 더 많은 비중을 두고, 더 적은 양의 정보에 기초한 믿음을 형성하는 것이 정당하다고 느끼는 경향이 있다. 그러나 정확성이 중요한 경우, 동일한 증거로 인해 신뢰도가 낮아지고 수용 전 시간이 길어진다. 이는 사람들이 다른 양의 증거에 근거하여 믿음을 형성할 수 있지만, 여전히 그들의 신념이 증거에 의해 적절하게 지지된다고 믿는다는 것을 보여준다(Cusimano & Lombrozo, 2021: 4). 그러나 실제로 종종 사람들은 자신이 원하지 않는 신념보다 원하는 신념에 대해 더 낮은 증거 임계값을 설정하는 유해한 이중 기준으로 이어진다(Ditto & Lopez, 1992). 이처럼 도덕 판단에 기준이 되는 도덕적 가치는 사람들이 증거에 기반한 방식으로 동기화된 추론을 하도록 유발한다. 또한 대안적 정당화 가설에 따르면 다른 사람들이 무엇을 믿어야 하는지 결정할 때, 사람들은 도덕적 고려를 그것을 뒷받침하는 사실적 차원의 다른 증거와 비교하여 경중을 따진다. 만약 사람들이 어떤 신념에 대해 근거가 충분하다고 스스로 생각한다면, 이를 통해 자신의 신념을 유익한 것으로 보고 동기화된 추론을 정당한 근거로 취급하도록 한다.

한편, 진화 생물학에서 개발된 '밈(meme)' 이론은 신념과 그것을 소유한 주체 사이의 관계를 이해하는 새로운 틀을 제공함으로써 도덕적 신념의 변화 가능성을 제시한다. '밈' 개념은 리차드 도킨스가 『이기적 유전자』에서 처음 소개되었다. 그에 따르면, 밈은 문화 내에서 사람에서 사람으로 퍼지는 생각, 행동 또는 스타일이다(Dorkins, 1989/2010). 그는 밈을 유전자와 유사한 방식으로 복제하고 시간이 지남에 따라 진화하는 문화 바이러스의 한 종류로 보았다. 도킨스에 따르면, 유기체는 개체로서 자기 자신의 이익이 아니라, 유기체를 구성하는 유전자의 이익을 위해 존재한다. 즉, '이기적 유전자'의 복제를 위해 유기체가 존재하는 것이다.

대니얼 데닛은 밈에 대한 도킨스의 개념을 더욱 확장했다(Dennett, 1995). 데닛은 밈을 "정신의 바이러스"로 생각할 수 있는 문화적 전달의 단위로 묘사했다. 그는 유전자가 우리의 생물학을 형성하는 것과 같이, 밈이 우리의 신념과 행동을 형성하는 데 중요한 역할을 한다고 주장했다. 이처럼 도킨스와 데닛은 밈이 유전자와 유사한 복제자로 생각될 수 있는 방법을 강조했다. 즉, 유기체가 유전자의 복제를 위해 존재하는 것처럼, 밈의 견지에서 본다면 우리의 신념과 행동은 우리의 이익에 부합하기 때문에 나타나는 것이라기보다는, 오직 밈 자신의 이익에 부합하기 때문이다. 즉, 밈은 말, 글, 심지어 몸짓 언어와 같은 다양한 수단을 통해 한 개인에서 다른 개인으로 전달될 수 있다고 언급했다. 일단 밈이 전달되면, 그것은 집단을 통해 퍼지고 진화하면서, 스스로의 삶을 영위할 수 있는 잠재력을 가지고 있다[41].

밈이 우리의 믿음을 형성할 수 있는 중요한 방법 중 하나는 밈플렉스 현상을 통해서이다. 밈플렉스는 더 큰 믿음 체계를 형성하기 위해 함께 작용하는 상호 연관된 밈의 집합이다. 이러한 믿음 체계는 우리의 생각,

41) 밈에 대한 생각은 문화적 진화를 이해하는 데 유용한 도구로 보는 반면, 다른 사람들은 결함이 있거나 축소주의적인 개념으로 보는 등 시작부터 논란이 있어 왔다. 그럼에도 불구하고 밈의 개념은 인류학, 심리학, 사회학 등의 분야에서 영향력을 발휘하여 패션 트렌드의 확산부터 종교 운동의 성장에 이르기까지 광범위한 현상을 설명하는 데 사용되어 왔다. 밈 개념에 대한 논쟁과 관련하여서는, Aunger, 2000; Blackmore, 1999; Dennet, 1995′ Distin, 2005; Sterelny, 2006을 참고할 것.

행동, 심지어 정서를 형성하면서 한 개인에서 다른 개인으로 전달되는 문화 프로그램으로 생각될 수 있다. 예를 들어, 종교적 근본주의의 밈플렉스를 생각해보자. 이 밈플렉스는 성경의 문자적 해석에 대한 믿음, 자연 현상에 대한 과학적 설명의 거부, 개인적인 경건함과 도덕적 순수성에 초점을 맞춘 것과 같은 많은 상호 연관된 밈으로 구성되어 있다. 이 밈들은 이 밈플렉스에 가입한 개인들의 생각과 행동을 형성하면서 일관된 신념 체계를 형성하기 위해 함께 작용한다. 마찬가지로, 정치적 이데올로기는 정부의 역할, 부의 분배, 개인의 권리와 책임의 본질에 대한 우리의 믿음을 형성하는 여러 상호 연관된 밈으로 구성된 밈의 집합체로 생각될 수도 있다. 예를 들어, 신자유주의의 밈플렉스는 자유시장, 개인주의, 제한된 정부 개입의 중요성을 강조하는 반면, 사회주의의 밈플렉스는 생산수단의 집단소유와 부(富)보다 평등한 분배를 강조한다.

도덕적 가치 또한 밈의 관점에서 분석될 수 있다. 밈 이론에 따르면, 도덕은 무엇이 옳고 그른지, 좋고 나쁜지에 대한 더 큰 믿음 체계를 형성하는 상호 연관된 밈플렉스로 생각될 수 있다. 이러한 도덕적 신념은 우리의 생각, 행동, 그리고 심지어 정서를 형성하면서 한 개인에서 다른 개인으로 전달될 수 있다. 예를 들어, 이타주의의 밈을 생각해볼 수 있다. 이 밈은 다른 사람들을 돕고 다른 사람들의 필요를 우리 자신의 것보다 우선시하는 것의 중요성을 강조한다. 이 밈은 무엇이 도덕적이고 윤리적인지에 대한 우리의 믿음을 형성할 수 있으며, 우리가 사심 없이 행동하고 도움이 필요한 다른 사람들을 돕는 것이 중요하다고 믿게 한다. 유사하게, 공정성의 밈은 무엇이 정의롭고 부당한지에 대한 우리의 믿음을 형성할 수 있다. 이 밈은 사람들을 평등하게 대하고 모든 사람에게 공평한 기회가 주어지도록 하는 것의 중요성을 강조한다. 공정성이라는 밈은 소득 불평등과 같은 문제에 대한 우리의 믿음을 형성할 수 있고, 우리가 작은 집단의 개인들이 불균형한 양의 부와 자원을 갖는 것은 잘못되었다고 믿게 한다.

밈플렉스의 일종으로서 도덕적 신념이 고정된 실체가 아니라 시간이 지남에 따라 지속적으로 진화하고 변화하고 있다는 점에 유의해야 한다.

밈은 한 개인에서 다른 개인으로 전달될 때 추가, 삭제 또는 수정될 수 있으며, 이는 더 큰 신념의 체계의 진화로 이어진다. 시대와 장소를 막론하고 보편적으로 공유되는 도덕적 신념 또한 사회와 문화적 환경 등의 요인에 따라 다른 방식으로 표현된다. 문화적 진화의 이러한 역동적인 과정은 생물학적 진화가 작동하는 방식과 유사하며, 시간이 지남에 따라 새로운 특성이 나타나고 개체군을 통해 확산되는 것이다.

도덕적 밈이 전달되는 방식은 그것들이 우리의 믿음을 형성하는 방식에도 상당한 영향을 미칠 수 있다. 예를 들어, 도덕적 밈의 전달은 미디어에 의해 영향을 받을 수 있는데, 미디어는 특정한 방식으로 문제를 틀리게 만들고 무엇이 옳은지 그른지에 대한 우리의 이해를 형성할 수 있다. 이처럼 우리의 도덕적 신념을 공유하는 집단(가정, 학교, 국가)은 옳고 그름에 대한 우리의 이해를 형성하는 일련의 도덕적 가치와 가르침을 전달하면서 우리의 도덕적 신념을 형성하는 데 중요한 역할을 할 수 있다.

밈의 견지에서 우리의 도덕적 신념의 형성 과정을 본다면, 도덕 판단에서 동기화된 추론으로 인한 문제 개선의 실마리를 찾을 수 있다. 스타노비치(2021/2022)는 밈의 관점이 신념의 확산에 대한 새로운 해석의 가능성을 제공하고 있음을 지적한다. 그는 밈 이론을 바탕으로, 진화론적 인식론이 '우리가 신념을 소유하는 것이 아니라, 신념이 우리를 소유한다는 생각(스타노비치, 2021/2022: 150)'으로의 전환을 가져올 수 있다고 평가한다. 이와 같은 관점은 신념의 형성과 확산의 새로운 해석 가능성을 제공해준다.

> *밈 이론은 신념에 관한 우리의 추론에 지대한 영향을 끼친다. 우리가 신념에 대해 생각하는 방식을 뒤집어 주기 때문이다. 사회 심리학자와 성격 심리학자는 전통적으로 '특정 개인들은 어떻게 모종의 신념을 얻게 되는가' 라고 묻는 경향이 있다. 이 인과관계 모델은 개인이 어떤 신념을 가질지를 결정하는 모델이다. 반면 밈 이론은 '특정 밈들은 어떻게 그 자신을 위해 수많은 주인들을 모을 수 있는가' 라고 묻는다. 따라서 이 질문은 '사람들은 어떻게 신념을 획득했는가가 아니라 신념은 어떻게 사람들을 획득했는가' 이다(스타노비치, 2021/2022: 152).*

신념 x가 왜 퍼져 나가느냐에 대한 상식적 관점은 그것이 그저 '사실이기 때문'이라는 생각이다. (중략) 밈 이론은 우리에게 신념이 확산하는 또 다른 이유를 제공해준다. 즉, 신념 x가 사람들 사이에 퍼져 나가는 이유는 그것이 좋은 복제자이기 때문이라는, 즉 그것이 주인을 확보하는 데 능숙하기 때문이라는 것이다. 밈 이론은 우리로 하여금 신념을 획득하는 사람들의 자질보다는 신념의 복제자로서 특성에 주목하도록 해준다. 이것이야말로 밈 개념이 기여하는 한 가지 독특한 기능으로 매우 의미심장한 것이다(스타노비치, 2021/2022: 153).

밈 이론은 우리가 특정한 도덕적 신념을 소유하고 획득하는 관점에서 벗어나, 그러한 도덕적 신념이 인간이라는 좋은 복제자를 통해 한 사회 내로 널리 확산된다. 즉, 밈플렉스를 도덕적 신념의 관점에서 본다면, 우리의 도덕 판단에서 동기화된 추론이 나타나는 까닭은 인간이 밈의 복제자로서 적합한 환경을 제공하기 때문인 것이다. 밈 이론의 견지에서 본다면, 자신의 신념을 보호하기 위한 동기화된 추론에 대해 비판적인 입장을 취할 수 있다. 내가 유지하고자 했던 나의 도덕적 신념에 대해 회의적인 시각을 갖는 것만으로도, 동기화된 추론의 문제는 개선의 여지를 확보할 수 있기 때문이다.

이러한 사실은 앞서 살펴본 동기화된 추론에서 문제가 되는 '방향성' 동기를 발견한 하트 등의 연구에서도 논의되었다(Hart et al., 2009). 해당 연구는 '신념이 변화할 수 있다'라는 생각을 갖는 것만으로도 동기화된 추론이 개선될 수 있다는 것을 발견했다. 하트 연구진은 이와 같은 발견에 대해 다음과 같은 해석을 제시하였다(Hart et al., 2009: 581).

자신의 입장을 바꿀 수 있는 능력이 입장을 바꿔야 하는 이유에 대한 고려를 촉진함으로써 인지부조화 해소를 촉진할 수 있다는 것이다. 즉, 자신의 도덕적, 정치적 신념이 '바뀔 수 있다'는 것을 인지하는 것만으로도 정보에 대한 객관적이고 정확한 해석을 촉진하는 계기를 마련해준다는 것이다. 또한 의사결정을 변경할 수 있는 지각된 능력이 불완전한 의사결정과 관련된 내용(태도, 신념, 행동 등)과 의사결정 전략(일치하는 정보의 검색, 불일치하는 정보의 회피 등)에 대한 기억을 직접적으로 개선함으로써 선천성 편향을 강화한다는 것이다. 즉, 자신의 도덕적, 정치적 신념이 '바뀔 수 있다'는 것을 인지하는 것은 최종적인 의사결정을 하기 전까지 계속해서 자신의 입장을 바꿀 수 있는

다양한 전략을 탐색할 가능성을 높일 수 있다.

해당 연구에서 발견되는 특징적인 결과 중 하나는 의사결정의 '가역성' 여부였다. 즉, 사람들은 가역적인 의사결정보다 비가역적 의사결정에서 더 큰 방향성에 의해 동기화되기 때문에 더 강력한 편향을 보인 것이다다. 동기화된 추론은 자신의 결정을 번복할 수 없거나 결정에 따른 수용 결과를 바꿀 수 없는 상황에서 더욱 활성화된다. 이처럼 동기화된 추론의 문제를 개선하기 위해서는 자신의 신념에 대한 변화 가능성을 인식하고, 열린 마음을 가지는 것에서부터 시작할 수 있다.

2. 편향의 사각지대 극복

동기화된 추론은 편향의 사각지대로 인해 비도덕적 행위를 정당화하거나 도덕적, 정치적 불일치의 양극화를 더욱 가속화시키는 문제를 내포했다. 편향의 사각지대 현상을 최초로 발견한 프로닌은 이 현상이 자기 성찰적 착각 때문이라고 주장한다(Pronin & Kugler, 2007). 이들에 따르면 사람들은 자신과 다른 사람들의 편향을 평가할 때 상대적으로 행동 정보(현재 보이는 행동에 국한된 정보)와 자기성찰적 정보(해당 행동 이전에 있었던 과거의 모든 관련 정보)에 두는 가치를 차별적으로 적용한다. 사람들은 다른 사람이 편향되어 있는지 여부를 평가할 때에는 행동 정보만을 평가의 근거로 삼는 경향을 보인다. 반면, 자신이 가진 편향에 대해 평가할 때는 표면적으로 드러나지 않은 행위의 의도, 자신의 생각과 감정을 찾아 편향된 동기 등 자기성찰적 정보를 최대한 찾아내어 평가의 기준으로 삼고자 하는 것이다. 이와 같은 맥락에서 스티븐 핑커는 편향의 사각지대를 '편향에 대한 편향(bias bias)'이라고 명명한다(Pinker, 2021: 547). 즉 편향의 사각지대는 자신은 편향에 빠지지 않으면서 자신을 제외한 다른 사람은 모두 편향에 빠지기 쉽다고 생각하는 일종의 메타 편향으로 규정될 수 있다. 따라서 우리는 자신이 편향에 빠졌을 수 있는 가능성을 인식하고, 자신의 관점만을 유지하려는 방향에서 벗어나 정보와 근거를 합리적으로 고려하고 더 정확한 결론에 도달하고자 하는

목표를 의식적으로 갖도록 노력할 때, 편향의 사각지대에서 빠져나올 수 있을 것이다.

이처럼 편향의 사각지대는 정확성 동기의 부족에서 비롯된다. 동기화된 추론의 문제가 낮은 정확성에서 비롯되었다는 것은 증거를 객관적으로 검토하거나 합리적인 근거에 기반하여 추론하지 못함을 의미한다. 따라서 도덕 판단에서 작동하는 추론이 높은 정확성을 목표한다면 편향의 사각지대는 개선될 수 있을 것이다. 인간 인지에 대해 오랫동안 널리 수용되는 관점은 추론의 기능이 적어도 어느 정도는 정확한 신념의 형성과 진리의 추구라는 목표에 기여하는 것에 있다고 주장해왔다(Kohlberg, 1969; Piaget, 1965/2000). 그러나 편향의 사각지대를 포함한 정체성 보호 인지(Kahan et al., 2012), 내 편 편향(Stanovich, 2021/2022), 확증 편향(Nickerson, 1998) 등 동기화된 추론 관련 개념의 존재는 '인지 편향'이라는 인간 추론의 심각한 결함을 시사한다(Pennycook et al., 2023). 페니쿡 등 연구진은 인간 추론이 갖는 기능에 대한 관점에 대해 광범위한 견해의 차이가 있음을 확인하였다. 그들은 인지 편향과 추론의 관계에 대한 기존의 심리학적 접근을 크게 세 가지로 구분하였다. 추론에 관한 고전적 견해는 '추론 능력의 함양은 편향 극복에 유용하다(reasoning is helpful)'는 명제로 대표되며 합리적 추론을 전제한다. 정서 혁명 및 자동성 혁명 이후의 전형적인 견해는 '추론 능력은 편향 극복에 쓸모가 없다(reasoning is helpless)'는 명제로 대표 되는데, 직관이 인간의 인지를 지배하기 때문에 추론은 대체로 비효율적이라는 견해이다. 여기에서 나아가 동기화된 추론 연구자들은 '추론은 편향을 극복하는 데 오히려 해롭다(reasoning is hultful)'는 명제로 대표는데, 이들에 따르면 추론은 일반적으로 건전한 판단력을 약화시키고 오히려 편향을 부추기는 기제가 되기 때문이다. 특히, 편향의 사각지대 문제가 추론 능력이 뛰어난 인지적 엘리트 사이에서 만연하다는 사실은 동기화된 추론의 견해를 뒷받침해준다.

이처럼 인간의 추론은 어떤 측면에서는 굉장히 합리적이고, 정확하며, 인식론적으로 유용하지만, 동시에 굉장히 비합리적이고, 부정확하며, 편

향에 빠지기 쉬워 보인다. 이러한 맥락에서 최근 인지과학 및 진화심리학의 관점에서 수행된 추론의 기원에 관한 연구는 추론의 기능이 정확한 판단을 내리고, 진리를 탐구하고, 합리적인 신념을 형성하는 것에만 있지 않다고 주장한다. 대표적으로 마음의 모듈이론의 권위자인 댄 스퍼버와 그의 지도 제자인 인지과학자 휴고 메르시에는 인간 이성과 추론에 관한 기존의 문헌들이 해결하지 못하고 있는 이 편향의 사각지대와 관련된 문제를 '이성의 수수께끼(the enigma of reason)'로 규정하고, 추론의 논증이론이 이에 대한 만족스러운 설명을 제공한다고 본다(Mercier & Sperber, 2017).

이성의 수수께끼란 추론 능력이 뛰어난 사람도 편향의 사각지대에 취약한 역설적인 상황을 지칭한다. 즉, 인간만의 고유한 특징인 이성과 추론 능력은 동물과 인간을 구분지어주며, 인간의 삶을 지금의 모습으로 발전시키는 데 크게 기여하였지만, 동시에 역설적이게도, 일상생활에서 인간의 추론은 종종 편향되고 비논리적이며 따라서 오히려 인간이 추론에 약한(poor reasoner) 모습을 보이기도 한다는 것이다.

메르시에와 스퍼버는 인간의 추론이 주로 진리를 찾기 위해 고안된 것이 아니라 논쟁에서 이기고 다른 사람들을 설득하기 위해 고안된 것이라고 제안함으로써 이성의 수수께끼에 대한 해결책을 제공한다(Mercier & Sperber, 2011; 2017). 그들은 추론이 정확성을 증진하기 위해 진화한 것이 아니라 의사소통과 사회적 상호작용을 촉진하기 위해 진화했다고 주장한다. 우리의 추론은 우리 사회 집단의 규범과 가치를 포함한 사회적, 문화적 요인에 의해 형성된다. 추론의 기능이 '진리 탐구'가 아닌 '다른 사람을 설득하기 위한 사회적 상호작용'에 있다고 본 그들은 이를 '추론의 논증 이론(the argumentative theory of reasoning)'이라고 지칭한다. 이 이론에 따르면, 인간은 추론 능력을 사회적 의사소통의 도구로 진화시켜 사회 집단의 갈등을 탐색하고 해결하는 데 도움을 주었다. 이러한 관점에서 추론은 순수한 인지적 과정이 아니라 사회적 관계를 유지하고 갈등을 피하고자 하는 욕구와 같은 사회적, 정서적 요인에 의해 형성된다. 따라서 이 이론에서 추론은 인간이 사회 집단의 갈등을 탐색하고 해

결하는 데 도움을 주기 위해 진화한 상호작용 과정에서 적응한 진화의 산물이다. 이 이론은 인간이 진리를 추구하기보다는 자신의 신념을 방어하기 위해 더 많은 동기를 부여받고, 기존의 신념을 지지하는 주장을 수용하고, 이에 모순되는 주장을 거부할 가능성이 높다는 생각에 근거를 두고 있다. 이처럼 추론의 논증 이론은 왜 사람들이 편향의 사각지대에 쉽게 빠지는지 설명해준다.

메르시에와 스퍼버가 정립한 추론의 논증 이론은 추론이 정확한 의사결정이나 문제 해결을 위한 도구가 아니라 대인 간의 관계에서 논증과 설득을 촉진하는 의사소통의 수단으로 적응한 결과로 설명한다(Mercier & Sperber, 2011; 2017). 이 이론에 따르면, 추론은 집단이나 공동체 내에서 개인 간의 사회적 교류, 협력, 경쟁을 촉진하는 방법으로 발전한 적응의 산물이다. 추론의 진화는 논증을 생성하고 평가하는 과정인 논증의 발전과 밀접한 관련이 있다. 논증을 교환하는 일은 인간의 의사소통과 사회적 상호작용에서 중심적인 역할을 해왔고, 아이디어, 믿음, 가치의 교환을 촉진했다. 추론의 논증 이론에 따르면 추론은 논증이 생성되고 평가될 수 있는 사회적 맥락을 제공함으로써 추론의 진화에 핵심적인 역할을 해왔다(Mercier & Sperber, 2011). 논증은 개인이 다른 사람들이 자신의 신념이나 행동을 채택하도록 설득하고 결함이 있거나 무효인 주장을 식별하고 거부할 수 있게 해준다. 논증은 인간의 사회적 행동의 발달에 필수적이었고, 인간이 상호작용하는 방식에 영향을 미쳤다.

추론의 논증 이론이 밝혀낸 추론의 기능에 관한 논의에 따르면 동기화된 추론으로 발생하는 도덕 판단에서의 인지 편향은 추론의 결점이 아니라, 추론의 특징인 것이다. 편향은 추론에 결함이 있거나 추론하는 사람의 인지 능력의 결핍에서 비롯된 것이 아니라, 논증을 생성하는 과정에서 추론 능력이 탁월히 발휘된 것이다. 그러나 동시에 추론의 논증 이론은 편향의 사각지대를 벗어날 수 있는 아이디어 또한 제공해준다. 자신의 논증을 반박할 사람이 아무도 없는 상황에서 스스로 합리화할 때 동기화된 추론은 문제가 된다. 편향의 사각지대에 빠지기 쉬우며, 추론을 통해 그럴듯한 근거를 만들어내어 객관성 환상이 유지되어 이로부터 벗

어나기는 더욱 어렵다. 사람들은 혼자서 추론할 때, 좋은 해결책에 도달하지 못하거나, 좋은 신념에 도달하지 못하거나, 좋은 결정을 내리지 못하는 경우가 종종 있는데, 이는 그들은 단지 도덕적 상황에서 느낀 최초의 직감만을 확인했기 때문이다(Haidt, 2001). 반면, 자신과 반대되는 사람과 의견을 나눌 수 있을 때, 다른 참여자들의 동기화된 추론으로 인한 편향된 도덕 판단이 균형을 이루게 되고, 집단은 최선의 해결책에 집중할 수 있게 될 것이다(Mercier & Sperber, 2011; 2017). 따라서 추론은 집단에서 훨씬 더 효과적이다. 사람들이 혼자서 추론할 때, 그들은 잘못된 길로 갈 가능성이 매우 높다. 그러나 실제로 함께 추론할 수 있게 되면 올바른 해결책에 도달할 가능성이 훨씬 높아진다.

한편, 발달심리학자인 모쉬맨은 이성의 수수께끼 문제에 대해서 "논리적 '추리'는 취학 전 아동들 사이에서도 일상적이지만, 논리적 '추론'은 계속해서 발달하고, 어른들에게서도 불완전하다."고 답한다(Moshman, 2004b). 그는 이성의 수수께끼 문제를 '논리적인 어린이'와 '비논리적인 성인'의 양립 가능성으로 치환하여 합리성의 진보에 대한 발달적 개념을 통해 추론의 논증 이론의 설명을 보완하며 편향의 사각지대를 극복할 수 있는 실마리를 제공해준다(Moshman, 2004b: 221-222). 그는 논리성의 발달을 핵심으로 하는 피아제의 인지발달 연구에 대해 이성의 수수께끼와 같은 맥락에서 두 가지 비판을 제시한다. 피아제의 이론과 달리 이후 발달 심리학 연구들은 어린 아이들, 심지어 취학 전 아동들 또한 근본적으로 논리적이라는 증거를 제시한다는 점과 형식적 조작 논리를 완성하는 단계인 청소년 이후의 성인들이 근본적으로 비논리적이고, 따라서 비합리적인 판단을 내린다는 연구들이 많다는 점이다(Kuhn, 2000).

이에 대해 모쉬맨은 아동의 '논리적' 사고의 발달과 '합리적' 사고의 발달이 구분된다는 사실에 주목하여 이 문제를 해소하고자 한다. 즉, 그는 합리적이라는 것이 곧 논리적임을 의미하지 않음을 발달 심리학 연구를 통해 제시한다. 가령 필로우 등(Pillow et al., 2009)의 연구는 5세에서 10세 사이의 총 112명의 아이들에게 단순 추리(inference) 과제를 제시한다. 아이들은 색깔이 다른 두 개의 장난감이 두 개의 캔에 숨겨져 있는

상황에서, 먼저 한 캔을 들여다본 후, 다른 캔에 있는 장난감의 색깔이 무엇인지 질문을 했다. 나이에 상관없이 모든 아이들은 색깔을 정확하게 추리해냈다. 이뿐만 아니라, 모쉬맨은 실제로 취학 전 아동이 일상적으로 논리적인 추리를 한다는 여러 고전적인 실험 연구 결과들에서 아동은 자신이 사용한 추리의 규칙을 이해하진 못했다고 지적한다(Moshman, 2004b: 222). 이와 유사한 맥락에서, 미취학 아동들은 올바른 결론에 도달했음에도 자신이 추리해낸 것의 본질과 목적에 대해 전혀 알지 못했고, 심지어 자신이 추리했다는 것을 알지 못한 경우도 있다고 지적한다. 즉, 아동의 합리성이 발달하는 것은 단순히 논리적 추리를 만들어내는 기본적 능력이 아니라, 논리적 추리의 본질에 대한 메타인지적 지식과 자신의 추리에 대한 자각과 지식, 통제할 수 있는 능력인 것이다. 이처럼 합리성은 자신의 사고 과정에 대한 인식론적인 자각과 그에 대한 지식과 통제력을 바탕으로 한다(Kuhn, 2000; Moshman, 1998).

한편 추론(reasoning)은 인식론적인 목적을 갖는다. 우리의 추론은 때로 자신이 원하는 것을 얻기 위해 논증을 생성하고 평가하지만, 우리의 이해를 증진시키고, 도덕적 행동을 촉진하기도 하여 사고의 발달을 촉진하기도 한다(Moshman, 2011). 특히, 도덕적 추론은 이성적 주체들 사이의 사회적 관계에 내재된 논리에 뿌리를 둔다. 이러한 의미에서 도덕적 추론은 집단 내에서 개인의 신념과 행동을 조정하는 데 도움이 되는 사회적 기능을 제공하는 것으로 간주된다. 도덕적 추론의 이러한 사회적 기능은 인간 사회의 기능에 매우 중요한 것으로 생각되는데, 이는 개인들이 공동의 목표를 향해 함께 일하고 협력적이고 평화적인 방식으로 갈등을 해결할 수 있게 하기 때문이다.

나아가 그는 도덕적 추론이 단순히 개인 인식의 문제가 아니라 사회적, 문화적 요인에 의해 형성된다고 주장한다(Moshman & Geil, 1998). 예를 들어, 특정 문화의 규범과 가치는 도덕적 문제에 대해 개인이 추론하는 방식에 영향을 미칠 수 있다. 모쉬맨이 정의한 추론의 개념이 정당한 결론에 도달하기 위한 방법으로 숙고적 노력을 통해 추리를 조정하는 것에 있기 때문에, 이것은 개인적 차원에만 국한되는 것이 아니다. 우리

는 정당한 결과를 얻기 위한 공동의 목적을 위해 둘 이상의 개인이 의도적으로 그들의 생각을 조정하게 된다. 이는 자신의 논리를 타인과 함께 주고 받으며 함께 합리성을 구성하는 과정을 필요로 하며, 이러한 과정은 유전적으로 획득된다기보다는 사회적 환경과 교육 등을 통해 후천적으로 학습되고 발달의 결과로 나타나는 것이다. 이때, 도덕적 추론은 문화적 요인에 의해 영향을 받고 개인은 사회화와 타인과의 상호작용을 통해 도덕적 개념과 추론 전략을 배운다. 이처럼 모쉬맨은 도덕적 추론이 단순히 개인의 과정이 아니라 개인이 살아가는 문화적 맥락에 의해 형성된다고 제안한다. 따라서 이러한 추론의 기능에 대한 근본적인 특징들을 고려할 때, 편향의 사각지대에서 작동하는 동기화된 추론의 문제를 개선될 수 있는 것이다.

이상의 논의를 통해 편향의 사각지대 문제는 우리가 편향에 대해 가지고 있는 편향을 인식하는 것에서부터 개선의 여지를 발견할 수 있음을 확인하였다. 특히, 도덕 판단을 도출해내거나 정당화하는 과정이 개인의 내적인 심리 과정이라는 점에서 도덕적 추론을 포함한 인간의 추론 일반의 기능 또한 개인적 차원에만 국한하여 이해되기 쉽다. 그러나 지금까지 살펴본 바와 같이 추론의 기원과 기능에 관한 최근의 연구는 인간 이성의 고차원적인 사용 방식으로서 추론 능력은 타인과 협력하고 상호작용하는 과정과 밀접하게 관련되어 있다는 사실에 주목했다. 따라서 편향의 사각지대와 동기화된 추론이 유발하는 여러 문제를 개선하기 위해 자신의 편향과 추론을 메타적으로 검토해야 하며, 이를 위해서는 개인적 차원의 성찰뿐 아니라 사회적 차원의 논증 교환이 필요할 것이다. 동기화된 추론을 개선하기 위한 구체적인 방안을 제시하기에 앞서 그 방안이 갖추어야 할 조건들을 먼저 점검해볼 필요가 있다.

제 2 절 동기화된 추론의 개선을 위한 방향

1. 도덕적 메타인지의 활용

동기화된 추론의 문제를 개선하기 위해서는 자신의 도덕적 신념을 비판적으로 검토하고, 편향의 사각지대를 인식하는 것에서 시작되어야 한다. 이와 관련하여 개선의 방향을 설정하기 위해 도덕적 메타인지의 개념을 참고해볼 수 있다. 인간의 편협한 신념과 인지 편향의 문제를 지적하고 이를 개선하기 위해 도덕적 메타인지의 중요성을 강조한 정창우(2022)는 편견과 무지로 인한 오류를 바로 잡기 위해 도덕적 메타인지를 교육적으로 활용해야할 것을 주지한다. 그는 메타인지에 관한 기존의 연구들을 종합적으로 분석한 후, 도덕적 메타인지의 개념을 정의했다. 도덕적 메타인지란 '도덕적 사고나 행동과 같은 목표 달성을 위해 자신이 알고 있는 것과 모르는 것, 인지적 과정이나 사고과정에 대해 제대로 인식하고 이를 조절해나갈 수 있는 능력'이다(정창우, 2022: 139). 특히 도덕적 추론은 올바른 도덕 판단이나 의사결정을 위한 주요 변인으로 설정될 수 있지만 그 자체만으로는 충분한 조건이 아니다(정창우, 2022: 148). 지금까지 상세하게 살펴본 바와 같이, 도덕 판단에서 작동하는 동기화된 추론의 기제와 그로 인한 문제는 도덕적 추론이 메타인지적 성찰의 대상이 되어야 한다는 결론으로 귀결될 수 있다. 따라서 도덕 판단에서 동기화된 추론의 문제를 개선하기 위한 방안으로 인지 편향을 극복하고 메타인지 능력을 함양하기 위한 기존의 방안들을 참고해볼 수 있다. 인지 편향을 극복하기 위해 제시되었던 기존의 다양한 접근들은 동기화된 추론의 문제를 인간이 가지고 있는 인지 능력과 사고 과정이 갖는 선천적인 오류에서 기인한다고 진단한다. 따라서 이러한 방법들은 메타인지를 함양하는 데 초점을 둔다. 그 중 동기화된 추론의 개선 및 도덕적 메타인지와 관련하여 구체적으로 ①인지행동치료기법과 ②비판적 사고 함양을 위한 교육적 논의를 참고해 볼 수 있다.

도덕 판단에서 동기화된 추론의 문제를 지적한 하이트는 이를 개선하기 위한 방안으로 인지행동치료의 관점을 폭넓게 참고하고 있다. 그는 도덕적 직관에 내재한 본능적인 신념이 도덕 판단을 이끌어내며, 따라서 우리의 추론은 진리를 찾기 위해서가 아니라 그러한 신념을 지지하는 이론적 근거를 개발하기 위한 방향으로 동기화되기 때문에(Haidt, 2006/2010: 78), 이를 개선하기 위해서는 추론 자체를 교정하기보다는 이를 촉발하는 자동적 사고에 주목해야 한다고 주장한다. 이는 마치 우울증 환자들의 왜곡된 생각이 부정적인 정서를 촉발하고, 이것이 다시 그 생각을 왜곡시키는 악순환을 일으키는 것과 유사하다. 우리는 자동적으로 활성화되는 도덕적 신념을 유지하고자 하는 추론이 편향된 방식으로 다시 그 신념을 강화하게 되므로 이를 개선하기 위해서는 자동적인 사고를 교정하는 것이 효과적일 것이다.

인지행동 치료는 '인지적 요인'이 부적응적 정서와 행동을 일으키는 주된 요인이라는 전제를 바탕으로, 심리적 기능 장애를 줄이기 위한 치료적 노력으로 관련 인지 변수의 변화 초점을 맞추고 있다(Beck, 1976). 인지행동치료를 제시한 심리학자이자 정신과 의사인 아론 벡은 심리적으로 실패, 상실, 좌절과 같은 부정적인 사고에 빠진 환자들을 관찰한 결과 그 원인에 '자동적 사고'과 '부적응적 도식(스키마)'에 있음을 발견했다.

자동적 사고란 어떤 상황 가운에 있을 때, 마음속에 빠르게 떠오르는 생각이다. 우리는 자신의 잠재의식 속에 자동적 사고의 존재를 알고 있다 하더라도, 합리적인 분석 과정을 거치지 않기 때문에 그 순간에 의식하기란 쉽지 않다. 특히, 자동적 사고는 그가 지닌 도식(스키마)에 근거하여 순간적으로 나타나는데, 정보를 처리하는 개인의 내적 참조 틀인 도식은 자신의 판단과 행동의 규준이 되는 핵심 신념들로 이루어져 있다. 도식은 환경으로부터 온 정보를 분류하고 필터링하고 부호화하며 의미를 부여하는 기능을 담당한다. 도식은 과거로부터 경험적으로 쌓아온 환경과의 상호작용의 축적물이며, 자신의 미래 행동을 예측하는 규준 틀인 것이다. 이때, 자신이 참조하는 도식(스키마)이 '부적응적'인 형태로 형성되어 있을 경우, 자동적 사고는 부정적인 방식으로 표출될 가능성이

높다.

인지행동치료는 자동적 사고를 교정하고, 그것이 근거하는 자기 도식을 수정하는 방식으로 부정적 사고에서 벗어나도록 돕는다. 이를 위해 치료자는 내담자가 자신의 의식에 대한 자각 상태에서 합리적 기초에 근거하여 결정을 내리도록 돕기 위해 의식적 주의를 이끌어야 한다. 의식적 주의는 ①환경과의 상호작용을 모니터링하고 평가하며 ②과거의 기억과 현재의 경험을 연결하고, ③미래의 행동을 통제하고 계획한다. 따라서 인지행동치료를 수행하는 치료사는 내담자의 합리적 사고와 문제해결과 같은 적응적인 의식적 사고를 개발하고 적용하는 것을 목적으로 한다.

최근 인지행동치료는 이중처리이론을 적용함으로써 자동적 사고로 인한 인지적 왜곡과 이를 개선하기 위한 아이디어를 참고하고 있다(Jurchiş & Opre 2018). 이 관점에 따르면 부적응적 정서를 유발하는 자동적 사고는 시스템 1의 인지과정으로 특징지을 수 있으며, 이를 개선하기 위해서는 시스템 2의 개입을 통해 자동적 사고를 모니터링하고 교정할 수 있는 가능성을 확보한다. 즉, 시스템 1의 자동적 사고를 개선하기 위해 시스템 2를 통한 개인의 반성적 성찰 능력을 포함한 자기 인식 및 조절 능력의 함양이 필요한 것이다. 이와 관련하여 대니얼 시겔은 부정적인 자동적 사고를 개선하여 인지 편향을 극복하는 데 있어 "마인드사이트(mindsight)"의 중요성을 강조한다(Siegel, 2010). 마인드사이트는 정서, 생각, 믿음을 포함한 자신의 정신적 과정을 인식하는 능력이다. 그것은 자신의 정신 상태에 대한 성찰적 인식을 발전시키고 그것들이 행동에 어떻게 영향을 미치는지 인식하는 것을 포함한다. 마인드사이트를 개발함으로써, 개인은 그들 자신의 편견과 신념을 인식하고 이러한 편견과 신념이 자신의 도덕적 추론에 어떻게 영향을 미치고 있는지를 반성할 수 있다.

다음으로 비판적 사고 교육은 도덕적 메타인지를 함양하기 위한 방안으로 제시되어 왔으며, 이는 동기화된 추론 개선 논의에도 참고해볼 수 있다. 인류의 도덕적 진보를 이성의 발전에서 찾고자 한 스티브 핑커는 인지 편향을 개선하기 위한 이성 능력의 계발로서 비판적 사고의 적용에

주목한다(Pinker, 2021: 571). 이성 능력을 옹호하는 교육자들은 '탈 편향'이라고 부르는 비판적 사고를 위한 교육 프로그램을 통해 학생들이 하나의 사안에 대해 여러 관점에서 비교, 분석, 평가해보고 자신의 입장을 수립하고 그에 대한 적절한 근거를 설정하는 연습 기회를 제공한다. 이를 통해 교육자들은 학생들에게 정체성 보호인지로 대표되는 동기화된 추론으로 인한 편향을 극복하고 객관적이고 논리적인 추론의 사용을 독려할 수 있다고 보았다.

이처럼 비판적 사고는 정보와 증거를 적극적이고 객관적으로 분석하여 합리적인 판단과 결정을 형성하는 과정으로 논증을 평가하고 그 논증의 강도와 타당성을 결정하기 위해 추론, 논리, 증거를 사용하는 것을 포함한다(Pinker, 2021). 비판적 사고는 올바른 결정을 내리고 문제를 해결하는 데 필수적인 기술이며, 광범위한 직업적, 개인적 맥락에서 높게 평가된다. 비판적 사고의 주요 특징 중 하나는 객관성과 회의성을 강조하는 것이다. 비판적인 사고를 사용하면, 당연시하는 가정에도 의문을 제기하고, 자신이 가지고 있는 기존의 도덕적 신념에 이의를 제기하며, 새로운 정보를 중립적으로 처리하고, 사실적 정보에 근거하여 증거를 평가하기 때문이다. 또한 비판적 사고는 편향된 사고와 잘못된 추론으로 이어질 수 있는 동기화된 추론을 인식하고 피하도록 돕는다.

이중 과정 이론의 관점에서 비판적 사고 또한 시스템 2에 의한 반성적 인지과정을 통해 작동한다(Kahneman, 2011/2012). 비판적 사고는 정보를 분석하고 평가하여 타당한 결론을 도출하는 체계적이고 객관적인 과정을 포함하기 때문에 의도적이고 의식적인 노력을 수반하는 시스템 2 사고를 통해 작동된다. 개인이 비판적인 사고에 참여할 때, 그들은 자동적이거나 직관적인 반응으로 나타나는 동기화된 추론을 억제할 수 있고, 이를 통해, 보다 객관적이고 공정한 방법으로 증거와 주장을 고려할 수 있다. 이처럼 비판적 사고는 자동적 사고에 비해 더 많은 인지적 노력을 요한다.

대표적으로, 도덕 판단에 대한 이중 처리적 접근을 시도한 그린 등의 연구(2008)를 통해 인지 부하와 도덕적 추론의 관계를 참고해볼 수 있다. 공리주의적 도덕적 추론은 전체적인 행복을 극대화하는 원칙에 기초

하는 반면, 의무론적 도덕적 추론은 엄격한 규칙과 의무에 기초한다. 이 연구는 고전적인 트롤리 문제의 수정된 버전을 사용했는데, 이는 두 가지 옵션 중 하나를 선택하는 도덕적 딜레마이며, 그 중 하나는 더 많은 해를 초래한다. 참가자들은 두 가지 버전의 트롤리 문제에 대해 도덕 판단을 하도록 요청받았는데, 그 중 하나는 해결하기 위해 더 많은 인지적 노력이 필요했다. 그 결과 참가자들은 인지부하가 있을 때 공리적 판단을 할 가능성이 높았던 반면, 인지부하가 없을 때는 의무적 판단을 할 가능성이 높았다. 이 결과는 사람들의 도덕적 판단이 그들의 인지능력에 의해 영향을 받을 수 있음을 시사한다는 점에서 동기화된 추론을 개선하는 데 있어 비판적 사고의 중요성을 강조한다. 나아가 해당 연구는 비판적 사고는 개인의 도덕 추론에 대한 인지 부하의 영향을 극복하는 데 도움이 된다고 제안한다. 비판적 사고에 참여함으로써, 개인은 자신의 편견과 한계를 더 잘 인식할 수 있고, 정보에 입각하고 잘 추론된 도덕적 결정을 내릴 수 있는 더 나은 준비를 할 수 있기 때문이다.

이와 같은 맥락에서 페니쿡 등의 연구진(Pennycook et al., 2015)은 동기화된 추론이 만들어내는 심오하면서도 그럴듯하게 들리지만, 실제로는 의미가 없는 일련의 진술, 소위 '그럴듯한 헛소리(pseudo-profound bullshit)'를 사용했다. 참가자들은 그 진술들을 깊이 있는 척도로 평가하도록 요청받았고, 또한 표준 측정을 사용하여 그들의 비판적 사고 능력에 대해 테스트를 받았다. 그 결과 비판적 사고력에서 더 높은 점수를 받은 사람들은 동기화된 추론이 만들어낸 그럴듯한 헛소리를 거부할 가능성이 높은 반면, 비판적 사고력에서 더 낮은 점수를 받은 사람들은 그것을 더 많이 받아들이는 것으로 나타났다. 이 연구는 또한 직관적인 사고와 영성(spirituality)의 척도에서 높은 점수를 받은 사람들이 '그럴듯한 헛소리'를 받아들일 가능성이 더 높다는 것을 발견했다. 비판적 사고 기술을 촉진함으로써, 개인은 의미 있는 진술과 무의미한 진술을 구별하고, 정보에 입각한 결정을 내릴 수 있는 능력을 더 잘 갖출 수 있다. 이것은 증거 없이 주장이나 주장을 받아들이는 것을 조심하고, 정보의 출처와 이를 뒷받침하기 위해 사용되는 주장에 기꺼이 의문을 제기하는 것을 포

함할 수 있다.

인지 편향 개선을 위해 제안되는 기존의 접근방식들은 인지행동치료나 비판적 사고 교육 이외에도 모두 이전에 가지고 있던 지식이나 사고 과정을 되돌아보며 깊이 성찰하는 과정, 자기중심적, 내집단 중심적, 인간 중심적 사고와 편견/선입견으로부터 벗어나려는 태도, 잘못된 배움을 없애가는 과정 등을 중시하는 공통점이 있다(정창우, 2020b: 88-89; 2022: 170-181).

인지 편향 개선 관련 기존 논의를 참고해 볼 때, 이와 같은 방법들은 공통적으로 동기화된 추론의 원인이 '잘못된 도덕적 신념' 또는 '합리적 추론을 방해하는 정서적 동요'에 있음에 주목하도록 한다. 도덕적 추론을 포함한 사회적 인지의 자동성에 대한 연구는 도덕적 판단의 직관주의 모델에 대한 추가적인 지지를 주었고, 많은 자동 과정이 의식적인 성찰이나 인식의 부재에서 인지적, 행동적 변화를 일으킬 수 있다는 점을 지지해준다(Haidt, 2007). 따라서 주로 개인의 숙고와 반성적 성찰을 통해 자신의 도덕 판단에서 나타난 추론이 자동적 사고로 인해 촉발된 특정한 방향성에 의해 동기화되었음을 자각하고, 합리적이고 객관적인 도덕 판단을 방해하는 자기 통제를 통해 정서를 적절한 수준으로 낮추거나 조절하며, 타당한 도덕 판단을 도출하기 위해 비판적 사고의 필요성을 강조하게 된다. 이와 같은 개선방안은 동기화된 추론에서 '동기부여'의 원인을 제거하거나 조절하기 위한 노력의 일환으로 이해될 수 있다. 즉, 기존의 개선방안들은 공통적으로 합리적인 도덕적 추론에 입각한 타당한 도덕 판단을 방해하는 요소들에 집중함으로써 동기화된 추론의 개선에 참고할 수 있다.

즉, 기존의 인지 편향 개선을 위한 여러 방안들은 동기화된 추론을 개선하기 위해 개인적 차원의 노력의 구체적인 방법과 시사점을 제공해준다. 그러나 이와 같은 접근은 동기화된 추론이 발생하게 되는 사회적, 문화적 요인을 종합적으로 고려하지 못한다는 한계가 있다. 도덕 판단에서 인지와 결합된 강한 정서적 반응이 동기화된 추론을 유발한다는 점에서 이를 개선하기 위한 방안으로 기존의 인지 편향 개선을 위한 접근을 참

고하는 것은 일면 타당하지만, 그 정서의 기원이 도덕 판단의 사회적 성격에서 기인한다는 점을 상기할 필요가 있다.

동기화된 추론 개념을 정립한 디토 연구진(Ditto et al., 2009)은 도덕 판단에서 동기화된 추론의 문제를 개선하기 위해 메타인지적인 측면을 강조하면서도, 도덕 판단의 사회적 성격에 주목한다. 도덕 판단의 평가적 의미는 필연적으로 사회적 성격을 지닌다. '새로운 종합'의 관점에서 수행된 최근의 도덕 심리학 연구들은 공통적으로 집단 생활의 진화적 압력에서 도덕 판단의 기원과 발달을 규명한다(Haidt, 2001; Mercier & Sperber, 2011; 2017, Haidt & Graham, 2007). 집단이 효과적으로 협력하기 위해서는 구성원들이 서로에 대한 신뢰를 가질 필요가 있는데, 여기서 이 신뢰는 다른 사람들이 피해를 주지 않을 것이라는 믿음을 포함하고 서로를 공평하게 대하며 집단 전체에 이익이 되는 방식으로 행동한다. 이것은 신뢰가 집단의 선(善, good)을 확립하고 구성원들이 함께 목표를 향해 활동할 있도록 하는 데 중요한 역할을 한다는 것을 의미한다. 즉, 신뢰가 없으면 협력, 협업, 정보 공유가 어려워져 집단에서의 성공적인 활동을 저해할 수 있다(Haidt & Graham, 2007). 도덕적 추론은 집단 구성원들에게 자신의 도덕적 선함을 설득하여 신뢰를 얻고, 반대로 타인의 선함을 평가하여 신뢰하는 과정에서 적응한 측면이 있다(Mercier & Sperber, 2011). 도덕적 추론과 도덕 판단의 이러한 사회적 요소는 사람들이 자신의 도덕성을 긍정적으로 평가하고, 도덕적 행위자로서 인식하기를 바라는 욕구가 반영된다(Ditto et al., 2009: 312). 따라서 동기화된 추론의 개선은 도덕적 메타인지를 함양하면서도 동시에 도덕 판단 및 추론이 갖는 사회적인 기능들이 고려될 때 더욱 효과적일 것으로 기대할 수 있다.

2. 추론의 사회적 기능의 고려

도덕 판단은 주로 개인의 동기, 생각, 의도에 대한 평가에 기초하는데, 이는 사회적 행위자의 의도적인 욕망에 기인할 수 있다는 것을 의미한

다. 즉, 동기화된 추론의 기초한 관점은 도덕 판단이 다른 사람들에 대한 우리의 정서, 우리와 다른 사람이 속한 사회 집단 등과 같은 다양한 사회적 요인에 의해 영향을 받을 수 있다는 것을 강조한다. 이는 최근 도덕 심리학 연구가 추론의 사회적 기능에 주목하고 있는 것과 맥락을 같이한다(Mercier & Sperber, 2011; 2017; Robinson, 2020). 전통적인 인지심리학적 접근이 추론의 사회적 성격을 무시해왔다고 비판한 로빈슨(2020)은 더 넓은 사회적, 문화적 환경에 의해 추론이 형성되는 방식을 이해할 때, 비로소 추론의 본질에 다가갈 수 있다고 보았다. 그는 추론이 단순히 개인의 인식의 문제가 아니라 사회적, 문화적 관행, 제도, 권력 관계에 의해 근본적으로 형성된다고 주장한다. 이러한 맥락에서 로빈슨은 사회적 구조와 문화적 가치가 개인의 추론 방식을 형성하는 방식과 추론이 이러한 사회적 구조와 문화적 가치를 어떻게 강화하고 재생산하는지 탐구한다. 따라서 도덕 판단에서의 추론은 동기화될 수밖에 없으며, 따라서 이를 개선하기 위해서는 개인적 차원에서의 메타인지 함양과 이를 통한 윤리적 성찰과 함께 사회적 차원의 노력이 고려되어야 한다.

추론의 논증 이론을 정립한 메르시에와 스퍼버는 추론의 기능이 사적인 추론을 통한 신념과 판단의 개선이 아니며, 추론의 기능은 사회적이며, 구체적으로는 논증적이라고 주장한다. 즉, 추론은 개인들이 서로 논증을 교환할 수 있도록 하는 기능을 수행한다는 것이다(Mercier & Landemore, 2012; Sperber, 2001; Mercier & Sperber, 2011; 2017). 한 종으로서, 인간은 전달된 정보에 의해 강한 영향을 받는다. 전달된 정보는 전부는 아니더라도 많은 결정에 영향을 미치며, 이러한 결정은 전반적으로 더 나은 결정이기 때문이다. 그러나 이와 같은 의사소통이 성공적으로 수행되기 위해서는 조작과 왜곡의 가능성을 고려해야 하며 서로에게 이익이 되어야 한다(Krebs & Dawkins, 1984). 의사소통의 발신자는 자신의 메시지를 수신자가 수용할 수 있도록 제시해야 하고, 수신자는 발신자가 제공한 메시지의 신뢰 수준을 판단해야한다. 따라서 인간은 다른 사람들이 말하는 것을 맹목적으로 신뢰하지 않는다. 우리는 의사소통된 정보를 평가하기 위해 일련의 인지 메커니즘을 사용한다. 스퍼버는

이러한 의사소통 능력을 인식적 경계(epistemic vigilance)의 메커니즘으로 설명한다(Sperber et al., 2010). 인식적 경계의 메커니즘은 신뢰의 교정 과정과 일관성 검사를 통해 구체화된다. 의사소통의 수신자는 발신자가 제시하는 것을 신뢰할 수 있는 것인지 의심해야하며, 그것이 자신이 가지고 있던 기존의 정보와 논리적으로 일관성을 유지하는지 파악해야 한다. 발신자의 입장에서 재진술해 본다면, 자신이 제시하는 바를 수신자가 신뢰할 수 있도록 설득력 있는 논증을 구성해야하며, 자신의 견해가 수용되기 위해서는 자신의 논증이 수신자가 가지고 있던 기존의 정보나 신념과 모순 없이 일관되어야 한다. 인식적 경계의 메커니즘에서 발휘되는 이러한 능력은 더 효율적인 의사소통을 가능하게 한다. 발신자는 설득력 있는 논증을 생성해야하고, 수신자는 전달된 논증의 질을 평가하는 의사소통 능력이 필요한 것이다. 이는 자신의 관점을 지지하거나 다른 사람들의 관점을 비판하는 데 사용될 수 있는 주장을 생성하고 평가하는 능력을 포함한다. 시간이 지남에 따라, 논증을 생성하고 평가하는 능력은 생존을 위해 점점 더 중요해졌는데, 그것은 개인들이 다른 사람들이 협력하도록 설득하고 효과적으로 경쟁하고, 그들의 생존 가능성을 향상시킬 수 있는 동맹을 형성하도록 하기 때문이다. 이처럼 논증을 생성해내는 것과 생성된 논증을 이해하고 평가하는 능력이 진화적 접근을 통해 알 수 있는 추론의 사회적 기능의 핵심 내용이다.

나아가 추론의 논증 이론의 관점에서 도덕적 추론은 도덕 판단에 도달하기 위해 사용될 뿐만 아니라, 다른 사람들이 특정한 도덕적 입장을 채택하도록 설득하기 위해 사용된다. 따라서 논증을 생성하는 것은 개인이 다른 사람들에게 더 설득력 있는 방식으로 자신의 도덕적 입장을 제시하게 해주고, 논증을 이해하고 평가하는 것은 다른 사람들의 주장을 비판적으로 평가하고 보다 객관적이고 잘 알고 있는 도덕 판단에 도달할 수 있게 하기 때문에 도덕적 추론의 핵심적인 사회적 기능으로 작동한다.

우리는 논증을 생성할 때는 상대방을 설득하기 위한 목표라는 방향성에 의해 강하게 동기부여 받는다. 따라서 자신의 주장을 구조적이고 따르기 쉬운 방식으로 제시한다. 또한 논증을 이해하고 평가할 때는 정확

성 동기에 기반하여 타인이 제시한 증거와 추론을 비판적으로 분석하고, 객관적 기준에 따라 주장이 설득력이 있는지 판단한다. 이것은 개인들이 그들의 기존 신념과 태도에 도전할 수 있는 증거를 고려하도록 강요하기 때문에 당파적 추론으로 인한 편향된 판단에 대응하는 데 도움이 될 수 있다.

더욱이, 논증을 생성하고 이해, 평가하는 것은 또한 사람들이 그들의 기존의 신념과 태도를 확인하는 방식으로 정보를 찾고 해석하려는 확증 편향을 개선하는 데 도움이 될 수 있다. 다른 사람들에게 자신의 주장을 제시하고 대화와 토론에 참여함으로써, 개인들은 다른 관점과 그들이 다른 방식으로 마주치지 않았을 수도 있는 증거에 노출될 수 있기 때문이다. 이것은 그들의 관점을 넓히고 더 객관적이고 잘 알려진 도덕적 판단으로 이끄는 데 도움이 될 수 있다.

동기화된 추론과 관련된 편향에 대응하는 것 외에도, 논증을 생성하는 것은 도덕적 추론에 대한 더 생산적이고 협력적인 접근을 촉진하는 데 도움이 될 수 있다. 우리는 신뢰하는 타인에게 자신의 논증을 제시하고, 서로의 논증을 이해하고 평가하는 과정을 바탕으로 도덕적 문제에 대해 함께 토론할 때 건설적으로 아이디어를 교환하며 자신의 잘못된 신념이나 편향을 검토할 수 있는 기회가 될 수 있다. 이처럼 추론이 가지고 있는 본질적인 기능으로서 논증의 생성과 평가라는 사회적 기능을 고려한다면 동기화된 추론으로 인한 문제는 더욱 효과적으로 개선될 수 있다.

진화심리학자들은 정신적 과정과 행동의 '기능'을 인류 조상들이 직면한 특정한 문제들을 해결하기 위해 진화한 적응의 결과로 본다. 그들은 인간의 마음이 주어진 환경에서 개인과 집단의 성공을 극대화하는 방식으로 정보를 처리하고 결정을 내리도록 진화했다고 주장한다. 이러한 관점은 자연 선택이 변화하는 환경에서 생존과 재생산의 도전에 대처하기 위해 수천 년에 걸쳐 인간의 마음을 형성했다는 생각에 근거를 두고 있다. 가령, 추론의 논증 이론이 인간의 마음이 갈등을 해결하고 공유된 이해에 도달하는 수단으로서 논쟁적으로 진화했다고 주장한 것은 '기능'에 대한 이러한 해석에서 비롯된다. 즉 추론의 사회적 기능으로서 논증의

생성과 평가 기능은 갈등을 해결하고 공유된 이해에 도달하는 수단 역할을 하므로 인간의 마음이 환경에 적응한 산물이다. 이들은 추론의 논증 이론을 통해 개인이 자신의 도덕적 신념을 시험하고 다듬고, 오류를 식별하고 수정하며, 중요한 문제에 대한 합의에 도달할 수 있다고 설명한다.

한편, 발달 심리학자들은 인간의 정신적 과정과 행동의 기능에 대해 개인의 '개체 발생적' 발달의 과정에 주목한다. 그들은 정신적 과정과 행동이 생물학, 문화, 경험의 결합에 의해 형성되며, 개인의 삶의 과정을 통해 변화하고 발달한다고 주장한다. 이러한 관점은 인간의 마음이 고정된 실체가 아니라, 일생에 걸쳐 다양한 요소에 의해 형성되는 역동적이고 변화하는 체계라는 생각에 근거를 두고 있다. 가령, 레스트(Rest, 1986)는 도덕적 추론은 개인의 인지 발달의 산물이며, 그것은 유년기와 청소년기의 과정을 거치면서 성숙해지고 더 정교해진다고 주장한다.

발달심리학자인 모쉬맨은 추론의 논증 이론이 추론에 대한 진화적 관점이 인간 행동에 대한 우리의 이해의 근본적인 측면을 반영한다는 점에서 가치가 있지만, '기능'의 의미를 특정한 목적을 수행하는 진화된 특성의 집합으로 봄으로써, 인간 마음이 갖는 다양한 층위의 역할을 무시할 가능성이 있음을 지적한다. 한 사람이라는 개체의 '발달'은 그 사람의 행동을 이해하는데 결정적이며, 이것은 단지 유전자나 진화로 환원할 수 없다. 따라서 추론의 논증 이론이 간과한 논리적 추론의 발달에 대한 기존의 광범위한 연구를 통해 추론의 사회적 기능에 대한 발달 심리학적 관점의 보완이 필요함을 제기한다(Moshman, 2004b: 221-222).

모쉬맨에 따르면 추론은 일종의 인식론적으로 자기제약적 사고(epistemologically self-constrained thinking), 즉 정당한 결론에 도달하는 것을 목표로 하는 사고로 정의될 수 있다. 이는 문제 해결, 의사결정 및 판단, 미래에 대한 계획에 있어 타당한 결론을 도출하기 위해 '논리'가 결합되어야 하며, 따라서 인지가 발달함에 따라 논리의 이해와 활용의 범위 또한 발달하게 되며, 추론의 기능 또한 더욱 정교해지고, 발달하게 된다. 추론의 논증 이론이 분석하고 있는 추론의 사회적 기능은 이

지점에서 보완될 수 있다.

또한 그는 협력적인 추론의 사용이 개인적인 추론의 사용보다 질적으로 우수하다는 것을 실험을 통해 주장한다(Moshman & Geil, 1998). 그는 웨이슨의 카드 선택 과제[42]에서, 개인이 단독적으로 참여했을 때 정답률이 9.4%인데 반해, 집단을 이뤄 참여했을 때 정답률이 70%로 극명하게 차이가 나는 것을 확인하였다. 그리고 이어 집단 참가자들이 정답에 이르는 토론의 과정을 분석하였다. 카드 실험에서 작업의 논리를 가장 잘 이해한 사람은 자신의 견해를 다른 사람들에게 납득시킬 수 있었고, 논리적으로 설득된 다른 구성원들은 그 의견에 기꺼이 따르고자 하였다. 또한, 집단 내의 모든 구성원이 처음에 올바른 선택을 하지 않았던 세 개의 집단도 결과적으로 정답을 맞추는 확률이 높았다. 이처럼 추론의 협력적 사용은 통찰력, 다양한 관점의 채택, 다양한 근거들, 의심과 가능성에 대한 검토 과정을 공유함으로써 진정한 합의에 도달하여 문제해결에 기여한다.

나아가 도덕 심리학 연구에서 소위 신-콜버그주의적 접근을 표방한 레스트와 나바에즈 등의 발달심리학자들은 도덕의 영역에서도 이와 같은 추론의 협력적 사용이 나타난다고 보았다. 이들의 관점에서 도덕적 추론의 사회적 기능을 분석하기에 앞서, 신-콜버그주의 관점에서 동기화된 추론의 문제에 대한 개선 방향을 분석해 볼 필요가 있다. 맹목적인 신념의 보호나 편향의 사각지대에 빠지는 것은 도덕 판단에서 동기화된 추론의 문제는 그것이 낮은 수준의 도덕 스키마[43]에 근거하기 때문에 자기

---

42) 피험자들은 카드의 한 면에는 낱자가, 다른 면에는 숫자가 나타난다는 말을 들었다. 과제는 다음과 같은 규칙의 타당성을 판단하는 것이었으며, 이 규칙은 네 장의 카드(E, K, 4, 7)에만 적용된다. “만일 카드 한 면에 모음이 있으면, 다른 면에는 짝수가 있다.” 피험자들의 과제는 판단해야 할 규칙의 정확성을 확인하는 데 꼭 필요한 카드들만 뒤집는 것이었다. 정답은 E와 7이다. E(모음) 뒤에 짝수가 있는지 확인하고, 7(홀수) 뒤에 모음이 없는지 확인할 때 규칙이 참임을 밝힐 수 있기 때문이다. 여러 실험 결과들을 평균해보면, 실험 참가자의 89%는 E를 선택하지만, 25%만이 7을 선택했다. 오히려, 62%의 참가자들은 4를 선택했다.

43) 인간의 정보 처리는 종종 스키마에 의해 중재된다. 스키마는 개인이 세계에 대한 정보를 조직하고 처리하기 위해 사용하는 연관 네트워크이다. 자아에 대한 정보 처리에서 자아 스키마의 역할에 상당한 주의를 기울였다. 자기 스키마는 정보를 필터링하여 특정 자극을 부착하고 기억하도록 선택할 수 있으며, 이는 해당

이익적이고 편향적으로 발생한다는 것에 있다. 따라서 우리의 도덕적 추론은 보다 높은 수준의 도덕 스키마에 근거하여 바람직한 방향성에 의해 동기화될 필요가 있다. 연구에 따르면 감정은 특정 행동이나 상황의 옳고 그름에 대한 빠르고 직관적인 감각을 제공함으로써 도덕적 의사결정에서 중요한 정보의 원천으로 작용할 수 있다(Narvaez, 2014). 그러나 이때, 감정은 또한 자기중심적이고 이기적인 도덕 판단을 이끌어냄으로써 도덕적 추론에 부정적인 영향을 미칠 수 있다. 감정이 지나치게 지배적이어서 도덕적 문제를 비판적으로 평가하고 공정한 판단을 내리는 개인의 능력을 흐리게 할 때 특히 그렇다. 이러한 감정적 반응을 특징으로 하는 도덕적 직관은 도덕 판단에 강력하고 종종 무의식적인 영향을 미치며, 그것이 우리의 도덕적 신념과 행동을 형성하고 인도하는 역할을 할 수 있다고 주장한다. 나바에즈 등은 도덕적 직관이 개인적 경험, 문화적, 사회적 규범, 도덕적 모범에 대한 노출을 포함한 다양한 요소에 의해 형성된다고 주장한다(Rest et al., 1999). 이는 우리의 도덕적 직관이 단순히 임의적이거나 주관적인 것이 아니라, 문화적, 사회적 맥락에 의해 형성된다는 것을 의미하며, 개인의 경험과 선호의 산물이 아니라, 우리가 살고 있는 더 큰 문화적, 사회적 맥락에 의해 형성되었음을 시사한다.

또한 나바에즈는 도덕 스키마에 대한 만성적 접근이 직관의 작동에 핵심적인 역할을 한다고 주장한다. 나바에즈 등은 도덕 스키마에 대한 접근이 발달 심리학과 사회인지이론에서 제시하는 '스키마' 개념의 중간 지대에 있다고 밝힌다(Rest et al., 1999: 297). 도덕 스키마는 우리가 경험을 통해 내면화한 도덕적 규범이나 가치의 정신적 표현이다. 이러한 도식은 우리에게 일련의 기대와 행동 기준을 제공하고 도덕적 상황을 평가하고 분류하는 데 사용된다. 나바에즈에 따르면, 도덕 스키마의 만성적 접근은 그것들이 우리의 마음에서 쉽게 활성화되고 항상 존재하며, 우리가 의식적으로 그것들을 의식하지 않을 때에도 우리의 도덕적 판단과 행동을 안내한다. 따라서 도덕 스키마에 근거한 직관적인 반응은 도덕적

정보에 대한 추론이 그려지는 방식에 영향을 미친다. 이를 뒷받침하는 증거는 자기 스키마 일관성 있는 정보를 위한 메모리를 선택적으로 용이하게 하는 데 있어 자기 스키마의 중요성을 확인해 주었다(Stangor & Ruble, 1989).

추론을 감정적으로 동기화할 수 있는 것이다. 문제는 그것의 방향성이다. 동기화된 도덕적 추론의 원인으로 잘못 설정된 스키마를 지적한 연구에서(Aronson & Reilly, 2006), 사람들은 일단 자신의 스키마를 채택하면, 자신의 행동을 부정확하게 반영하는 방식으로 해당 스키마를 사용한다. 즉, 주어진 결론에 도달하도록 동기를 부여받은 사람들은 그 결론을 선호하는 지식의 구조를 선택적으로 활성화하여 그들이 도출하고자 하는 결론에 대한 합리적인 정당화를 구성하는 것이다(Aronson & Reilly, 2006: 378). 따라서 우리의 도덕적 추론은 보다 높은 수준의 도덕 스키마에 근거하여 바람직한 방향성에 의해 동기화될 필요가 있는 것이다.

신-콜버그주의적 접근에서 바람직한 방향성의 설정을 가능케 하는 도덕성의 발달은 '또래와의 상호작용'을 통한 인지 발달에서 비롯된다. 자신과 다른 사람과의 상호작용은 다양한 관점을 채택해보는 경험을 가능케하고, 이를 통해 자신이 가지고 있던 세계를 이해하는 틀인 도식은 질적, 양적으로 발달할 수 있다. 콜버그가 정의추론의 최종 단계를 단일한 도덕 원리로 파악한 데 반해 나바에즈 등의 신-콜버그주의자들은 공동의 도덕성(common morality)의 존재를 주장한다.

공동의 도덕성이란, 특정한 사회나 문화의 구성원들에 의해 일반적으로 받아들여지고 공유되는 일련의 윤리적 원칙과 가치이다(Beauchamp & Childress, 1994; Toulmin, 1983; Walzer, 1983). 이러한 원칙과 가치는 종종 사회의 규범, 관습, 관행에 깊이 뿌리내리고 있으며, 의료, 법률, 정치와 같은 삶의 다양한 영역에서 도덕적 의사결정의 기초를 제공한다. 공동의 도덕성 개념은 인간이 사회적 협력과 공동체 생활에 필수적인 도덕적 신념과 가치를 공유한다는 생각에 바탕을 두고 있다. 응용윤리학자인 비첨과 칠드레스는 『생명 의료 윤리의 원칙들(Principles of Biomedical Ethics)』에서 자율성, 악행 금지, 선행, 정의라는 원칙에 기초한 의료 윤리적 의사결정의 틀을 제안한다. 이러한 원칙들은 서구 세계의 일반적인 도덕의 일부로 여겨지며 생명윤리학에서 널리 사용되는데, 이후 네 가지 원칙은 의료 윤리뿐 아니라, 다양한 응용 윤리 분야에서 인용되고 재해석되어 윤리 원칙으로 널리 사용되었다.

한편, 미국 생명윤리학자인 툴민(1983)은 자신의 개인적 경험을 바탕으로 이성적이고 공정하며 독립적인 개인들 사이에서 근본적인 도덕 원칙의 합의에 이르지 못한다고 결론내리며, 공동의 도덕성이 형성되는 과정을 설명한다(Toulmin, 1983, Rest et al., 2000: 383에서 재인용). 그는 윤리적 이슈에 대한 전문가들이 합의에 도달하는 과정에 주목했다. 그는 전문가들이 특정한 개별 사례에서의 도덕 판단에는 서로 합의할 수 있지만, 그러한 판단을 정당화하는 추상적인 도덕적 원칙에는 근본적인 합의에 도달할 수 없음을 깨닫게 되었다. 즉, 툴민은 도덕 판단이 '추상적 원칙'을 기반으로 '특정한 사례'에 적용된다는 하향식(top-down)으로 진행한다는 개념에 반론을 제기하며, 도덕적 추론은 사회적 상호작용 속에서 협력적인 방식을 통해 상향식(bottom-up)으로 도출된다고 주장한다. 즉, 우리는 서로의 도덕적 추론을 주고받으면서 특정 사례에 대해 합의에 도달한 다음, 유사한 사례를 분석함으로써 사례 간의 논리적 상호 연결 관계를 확립함에 따라 '공동의 도덕성'을 구축해가는 것이다.

정치철학자 왈처는 툴민이 경험적으로 발견한 공동의 도덕성의 구성을 이론적으로 타당화한다. 그는 도덕성이 근본적인 도덕 원칙에 기반하기보다는, '공동의 도덕성'이 각각의 개별적 사례를 다루는 공동체의 구체적인 경험으로부터 점진적으로 형성된다는 것이다(Walzer, 1983, Rest et al., 2000: 384에서 재인용). 그는 서로 다른 사회적 재화가 서로 다른 기준에 따라 분배되어야 한다는 원칙을 가리키는 '복합적 평등'의 개념을 논한다. 즉, 왈처는 도덕성의 발전은 개인의 기본 원칙을 적용하는 것으로 이해되기보다는 상황에 맞는 공유 가능한 도덕적 이상을 달성하는 방법에 대한 공동체의 합의에 도달하는 것이라고 주장한다.

신-콜버그 학파는 공동의 도덕성의 관점에 따라, 도덕성을 '사회적 구성물'로 보며, 공동체의 다양한 경험 속에서 진화를 거듭하며, 특정한 제도 간의 조정과 숙의를 해야 하는 개념으로 이해한다. 개별 사례에 대한 집단적 숙고에 의해 형성된 공동의 도덕성은 콜버그가 경계했던 무분별한 도덕적 상대주의나 도덕적 회의주의가 아니다. 공동의 도덕성은 공동체마다 다를 수 있지만 공동체 구성원들에 의해 논의되고 면밀히 조사되

며 공동체의 이상과 도덕적 직관 사이의 균형을 반영하기 때문이다. 이처럼 공동의 도덕성 개념은 우리의 도덕적 추론의 양상이, 추상적인 도덕적 원칙을 개별 사례에 적용하는 방식으로 나타나기보다는 개별적 사례들에서 구성원 간의 합의를 하는 과정에서 발생하는 것이며, 한 사회가 공유하는 도덕적 원칙이나 신념은 추론의 사회적 사용의 구성물로 이해될 수 있다. 즉, 추론은 자신이 속한 집단의 구성원들과의 사회적 상호작용 상황에서 협력을 가능케 한다. 따라서 성숙한 도덕적 추론 능력을 지닌 사람들은 구성원들과의 협력적 상호작용을 바탕으로 도덕 문제를 해결할 수 있다.

이상의 논의를 통해 분석된 추론의 사회적 기능은 동기화된 추론의 문제를 보다 근본적이고 효과적으로 개선하기 위한 준거가 된다. 다음 절에서는 이를 고려한 개선방안으로 협력적 논증 교환 활동을 검토할 것이다.

제 3 절 동기화된 추론의 개선방안으로써 협력적 논증 교환

1. 협력적 논증 교환의 의미

동기화된 추론을 개선하기 위해서는 추론의 사회적 기능을 고려하여 도덕적 메타인지를 촉진함으로써 가능하다. 특히, 추론의 사회적 기능은 추론이 논증 교환의 맥락에서 최상의 상태일 때 발휘될 수 있으며, 이는 함께 논증을 교환하고 평가할 의사가 있는 집단 내 구성원들 사이에 의견이 불일치할 때 그 효과가 극대화된다(Mercier, 2011b: 183). 또한 도덕적 이슈에 대해 자신과 입장이 다른 타인과 논증을 교환하는 경험은 자신의 기존 도덕적 신념을 되돌아보게 하고 수정의 기회를 마련해준다. 도덕의 기능은 우리에게 '서로를 대하는 법'을 알려주며, 공동의 도덕적 신념을 공유하는 공동체의 사회적 정체성이 구성원들을 연결해주기 때문이다(Moshman, 2012: 320). 자신의 신념을 보호하려는 기본적인 동기와

편향의 사각지대라는 동기화된 추론의 발생 맥락을 고려했을 때, 신뢰할 만한 집단에서 자신과 의견이 다른 구성원들과 논증을 교환하는 경험은 효과적인 개선방안으로 작용할 수 있다.

협력적 논증 교환은 대화의 맥락에서 도덕적 문제에 대한 공유된 이해를 구축하기 위해 서로 다른 관점에서 도덕적 논증을 주고받는 사회적 상호작용이다. 논증 활동은 하나의 관점을 정당화하는 명제들을 제시하여 그 관점의 수용 가능성을 확신 또는 설득하거나, 또는 그 관점에서 표현된 명제들을 (다른 관점에서) 반박하고자 하는 언어적, 사회적, 이성적인 활동이다(van Eemeren & Grootendorst, 2003). 논증에 대한 관점은 수사학, 논리학, 의사소통 이론과 같이 그것을 다루는 학문마다 다르지만, 공통적으로 논증이 청중이나 대화 상대를 설득하기 위한 목적으로 논리적으로 타당한 근거를 제공하고 이를 평가하는 과정이라는 것에 동의한다(Breton & Gauthier, 2000). 논증 활동을 분석하기 위해 널리 사용되는 툴민(1983)의 논증 모델은 주장이 논리적으로 건전한지 확인하기 위해 논증의 구조와 요소에 초점을 맞추지만, 대화 과정으로서 논증을 분석하는 데는 한계가 있고, 논증이 일어나는 맥락이나 사회 환경을 고려하지 않는다. 또한 도덕적 신념이나 정치적 쟁점과 같이 특수한 맥락이나 사회적 환경에서 일어나는 협력적인 논의의 과정과 현상을 고려한 모델이 아니라는 점에서 일정한 한계를 지닌다(van Eemeren & Grootendorst, 2003).

이와 달리 협력적 논증 교환은 특정한 주제에 대해 개개인의 입장과 관점을 공개적으로 표현하고, 그에 대해 제기된 반론과 질문에 대한 정당화와 설명을 교환하는 논의의 과정이다(Andriessen et al., 2003). 이는 개인 학습자의 일회적 논증 작성과 달리, 참여자들이 논증의 생성과 평가의 과정을 반복 수행하는 상호작용적 논증 활동의 양상을 띤다. 특히, 협력적 논증 교환은 다른 사람과 자신의 의견을 '서로 통합하여 구축하는 것'에 목적을 두는 것이 아니라, 사회적 과정에 기반하여 타인과의 지적인 협력에 참여함으로써 자신의 관점뿐만 아니라 다른 관점도 동시에 검토하여 지식을 구성하는 것에 목적이 있다(Kuhn, 2015: 46-51). 결론

적으로 협력적 논증 교환의 상황은 이질적인 견해를 가진 두 명 이상의 참여자가 같은 목표(더 나은 판단에 도달하고자 하는 목표 또는 윤리적 문제를 해결하고자 하는 목표)를 가지고 협력하는 상황을 말하며, 이때 협력적 논증은 자신의 입장을 관철시키거나 논리적으로 상대를 설복시키는 승패가 있는 논쟁이 아닌, 구성원이 해당 윤리적 이슈에 대한 공동의 이해, 합의, 동의에 도달하는 것이다(Andriessen et al., 2003). 따라서 협력적 논증 교환은 우리에게 자신의 관점이 틀리기 쉽다는 것을 확인하고, 자신의 기존 지식과 신념을 되돌아보고 의견을 스스로 수정할 수 있는 기회를 제공한다.

나아가 협력(collaboration)의 의미는 단순 협동(cooperation) 활동과 비교하여 이해될 때, 보다 명료화될 수 있다[44]. 협동과 협력은 과제를 개인이 아닌 여러 사람이 공동으로 수행한다는 점에서 유사하지만, 세부적인 활동 양상에서 차이를 지닌다. 일반적으로 협동은 참여자들이 하나의 일을 여러 부분으로 나누어, 각자 맡은 역할을 수행하고 최종적으로 각 부분을 모두 합쳐 공동의 목표를 달성하는 것을 의미한다. 반면, 협력은 공동의 목표, 결과 또는 하나의 과제를 위해 여러 사람들이 함께 동시에 논의해가며 해결해가는 과정을 의미한다[45].

협력적 논증 교환은 논증의 생성 및 평가 기능과 협력적 문제 해결 기능이라는 추론의 사회적 기능을 반영한 동기화된 추론의 개선 방법이다.

---

44) 협동과 협력의 개념 구분은 효율적인 과업 수행 요인을 주된 관심 연구 분야로 삼고 있는 경영학의 조직 행동과 관리 연구에서 수행되었다. 최근 한 연구는 (Castaner & Oliveira, 2020)그동안 경영학 연구에서 협동과 협력, 조정(coordination)의 구분이 모호하게 사용되고 있음을 지적하며, 약 60년(1949-2017)간 협동과 협력 개념이 사용된 2220편의 경영학 문헌을 대상으로, 두 개념에 대한 메타 분석을 통해 이 둘 개념의 차이를 확인하였다. 협동은 구성원 공통의 목표를 달성하기 위해 구성원 모두가 기꺼이 자신에게 주어진 과제를 수행하고, 수행한 결과를 함께 공유함으로써 목표에 도달하고자 하는 집단적 업무 처리의 방식이지만 반면, 협력은 집단 구성원들이 목표에 도달할 수 있도록 서로를 돕는 행위로 정의된다(Castaner & Oliveira, 2020: 978-979).

45) 이외에도 사회과학 연구에서는 협력과 협동의 차이를 '동시적'으로 함께 작업하는지 여부로 구분하기도 한다(Paternotte, 2014). 협력은 하나의 과제를 두고 협력 참가자들 모두가 함께 수행하는(working together) 반면, 협동은 단지 협동 참가자들 모두가 각자의 역할을 수행한다는(carrying out their part) 점에서 엄밀한 의미에서 협동은 '함께 일한다'는 의미를 포함하진 않는다고 지적한다.

추론의 논증 이론에 따르면 논증의 생성과 평가는 협력적 문제 해결의 수행에 있어 근본적으로 비대칭적인 양상을 보인다. 논증을 생성할 때는 추론이 방향성에 의해 강하게 동기화되기 때문에 편향되고 느슨하게 작동하지만, 논증을 평가하는 과정에서는 추론이 높은 정확성에 의해 동기화되기 때문에 객관적이고 엄격하게 작동하기 때문이다(Mercier & Sperber, 2011; 2017). 추론의 사회적 기능에 주목한 메르시에는 협력적 논증 교환의 개념을 규명하기 위해 대화의 범위와 갈등의 범위를 기준으로 추론이 나타나는 양상을 다음과 같이 구분한다(Mercier, 2016: 692).

[표 4-1] 대화와 갈등의 범위를 기준으로 구분되는 추론의 양상[46)]

	갈등의 양	
대화의 양	거의 없거나 발생하지 않음	갈등이 발생함
대화가 없음 (개인적 추론)	**A. 자신과 같은 의견을 가진 정보에 대해 혼자서 추론하는 것.** 우리는 대부분 우리의 의견을 지지하는 주장을 발견하고 그것이 매우 강하지 않더라도 그것을 받아들인다.	**B. 자신의 직관과 다른 정보에 대해 혼자서 추론하는 것.** 우리는 서로 다른 직관을 지지하는 주장을 발견하기 때문에 가장 많은 주장을 가진 직관이 승리한다.
대화 없이 논증만 받아들이는 경우 (단방향 논증)	**C. 모두가 동의하는 주제에 대해 일방적으로 논증을 제공 받는 것.** 우리는 우리가 받은 주장에 대한 거부에 동기화되기 때문에 반론을 생성한다. 그 논증은 효과적이지 않거나 역효과를 낳는다.	**D. 자신과 다른 의견을 가진 사람으로부터 일방적으로 논증을 제공 받는 것.** 우리는 논증이 뒷받침하는 방향으로 생각을 바꾼다.
대화 (쌍방향	**E. 모두가 동의하는 주제에 대해 다른 사람과 논증을 주고받는 것.**	**F. 자신과 다른 의견을 가진 사람과 논증을 주고받는 것.**

46) 동기화된 추론으로 인해 발생하는 문제는 주로 표의 A칸에 해당하며, 추론의 사회적 기능을 바탕으로 협력적 논증 교환의 양상은 F칸에 해당하므로 본 논문에서는 A칸과 F칸을 중심으로 서술하였다. 추론의 논증 이론가들은 단방향 논증(C, D칸)의 특징도 조사하였는데, 이는 논증의 교환이 포함되어 있지 않기 때문에 추론의 사회적 기능을 고려했을 때 동기화된 추론의 개선을 안정적으로 기대하기는 어렵다. 가독성을 위해 각 칸에 해당하는 알파벳과 밑줄을 추가하였다.

논증 교환)	합의된 의견에 대한 논쟁은 산적해 있다. 그것들은 잘 검토되지 않는다. 토론자들은 기존 의견을 더 극단적으로 지지할 수 있다.	서로 다른 의견에 대한 주장이 생성되고 평가된다. 일반적으로 사용 가능한 최고의 의견인 최선의 방어 의견이 더 영향력 있는 것으로 입증된다.

표에서 대화의 양(amount of dialog)은 추론을 사용할 때 참여하는 사람의 수를 의미한다. '대화가 없는' 경우는 추론을 단독적으로 사용하는 개인적 성찰과 일방향적인 논증 생성만을 포함하는 '대화 없이 논증만을 받아들이는 경우'로 구분된다. '대화가 있는 경우'는 추론의 참가자들이 논증을 생성하고 상대가 생성한 논증을 평가하는 과정이 포함되는 '쌍방향 논증 교환'에 해당한다. 갈등의 양(amount of conflict)은 의견의 대립 여부를 의미한다. 갈등이 발생했다는 것은 추론을 사용할 때, 반대되는 의견의 대립이 고려되고 있다는 의미이며, 갈등이 거의 없거나 발생하지 않는 경우는 반대되는 의견의 대립이 없는 경우를 지칭한다. 이를 기준으로 구분된 추론의 양상에서 동기화된 추론 개선을 위한 협력적 논증 교환의 양상은 '쌍방향 논증 교환'에서 '갈등이 발생하는 경우'에 해당한다. 이는 둘 이상의 사람이 동의하지 않는 주제에 대해 서로 다른 의견을 논증의 교환을 통해 주고받는 것이다.

추론의 사회적 기능의 관점에서 보았을 때, 대화와 갈등이 나타나지 않는 개인의 고독한 추론(solitary reasoning, 표의 A칸에 해당)은 동기화된 추론으로 인해 편향된 판단을 도출해내는 최적의 조건에 해당한다. 특히, 개인 내적으로 자신의 기존 신념에 대립하는 입장을 고려하지 않는 경우에는 문제 상황에 의해 촉발된 강한 직관이 도덕 판단을 좌우하며 도덕적 추론은 그러한 직관에 부합하는 방향으로 강하게 동기화된다. 메르시에는 이 경우에 동기화된 추론으로 인해 일반적으로 잘못된 직관을 바로잡지 못하고(Trouche et al., 2015), 비도덕적 행동을 정당화하기도 하며(Paharia et al., 2013), 자신의 입장에 대한 과한 확신(Windschitl et al., 2013)과 그로 인한 도덕적, 정치적 양극화의 심화(Wayne et al., 2015)로 이어질 수 있음을 지적한다(Mercier, 2016: 693). 특히, 추론의

단독적 사용으로 인해 발생하는 동기화된 추론의 문제는 추론하는 사람이 추론 능력이 부족하거나 무지해서 생기는 것이 아니다. 오히려 추론 능력이 뛰어나거나 지식이 풍부한 사람들이 가장 편향된 도덕 판단을 도출해낼 수 있는 것이다. 이는 앞서 살펴본 편향의 사각지대와 인지 엘리트의 편향 문제에 해당한다.

이와 달리 협력적 논증 교환(표의 E칸)은 사람들 사이의 대립된 의견 갈등이 논증 교환의 대화 상황으로 묘사된다. 여기에서 사람들은 논증을 생성하고 다른 사람들의 논증을 평가할 수 있는 기회를 갖는다. 이는 동기화된 추론을 극복하고 문제 해결을 위한 좋은 아이디어와 건전한 신념이 확산되기 위한 추론의 사회적 기능이 극대화되는 이상적인 조건에 해당한다(Mercier, 2016: 695). 특히, 논증의 평가는 자신과 의견이 같은 사람들 사이에서 이루어질 때는 상대적으로 잘 검토되지 않은 상태로 남아 있을 가능성이 높다. 동기화된 추론으로 인한 내 편 편향의 양상과 집단의 양극화는 이로 인해 심화될 수 있음을 3장에서 살펴보기도 했다. 즉, 논증의 평가는 자신의 신념에 도전하는 논증일 때 가장 잘 기능한다. 따라서 사람들은 어떤 문제에 대해 의견이 일치하지 않을 때 논증을 교환할 가능성이 높고 이때, 논증의 평가는 최상의 결과를 낳는 것이다.

추론의 논증 이론은 의사소통 과정의 인식론적 효과를 극대화하기 위해 논증 교환의 적절한 조건으로 세 가지를 제안한다. 먼저, 토론 중인 주제에 관한 한, 상반된 의견 교환의 틀에서 추론이 만들어진다는 것이다(Mercier & Landemore, 2012: 248). 진정한 논증의 교환이 이루어지려면 적어도 두 가지 상반된 관점에서 추론하는 사이에서 피드백이 순환하는 구조가 구축되어야 한다. 두 사람이 별개의 의견을 갖는다는 전제하에 일반적으로 아래와 같은 순서를 따른다. ①논증 참가자 A는 자신의 관점(의견 A)에 근거한 논증을 제시하기 위해 추론을 사용한다. ②그런 다음 논증 참가자 B는 추론을 사용하여 이 논증을 자신의 관점(의견 B)에서 분석한다. ③참가자 B는 의견 B에 기초하여 반론의 형태로 논증을 구성하기 위해 추론을 활용한다. ④참가자 A는 다시 의견 A에 기초하여 이 논증을 분석하기 위해 추론을 사용한다. 이와 같은 정의에 따르면, 진

정한 숙고는 우리가 타인의 논증을 인식하고 그것들을 검토할 때만 이루어질 수 있다. 즉, 자신의 논증을 생성하는 사람은 다시 타인의 논증을 이해하고 평가하는 역할을 수행해야 한다.

다음으로, 대화 상대방의 논증에 대한 평가가 이루어진다는 것이다(Mercier, 2011a: 140; Landmore & Mercier, 2012: 916). 논증의 수용자로서, 우리는 결론을 받아들여야 하는지를 결정하기 위해 논증을 평가해야 한다. 이러한 추론의 기능은, 우리가 입장을 정당화할 때 이로 인해 타인의 생각이 바뀔 수 있다는 사실을 알기 때문에 더 설득력 있는 논증을 구성하도록 동기부여 된다. 만약 자신의 도덕 판단이 다른 사람들에게 거의 영향을 미치지 않고, 다른 사람을 설득하기 어렵다고 느낀다면 논증을 구성하는 동기는 제한적일 것이다. 반면, 논증의 상대방이 무조건적인 지지자인 경우, 즉 자신의 입장을 너무 쉽게 받아들이고 납득할 경우 또한 더 나은 논증을 구성하고자 하는 동기는 제한될 것이다. 여러 실험 연구에서 참가자들은 자신의 논증이 평가된다는 사실을 인식할 때, 더욱 설득력있고 합리적인 논증을 구성했다(Mercier & Sperber, 2017/2018: 330-336). 한편 스스로를 정당화하기 위해 논증을 만들어내는 과정은 종종 편향에 빠지기 쉽다. 왜냐하면 자신이 이미 도달한 도덕 판단을 뒷받침해주고 강화하는 논거들을 더 높이 평가하고 더 쉽게 수용하기 때문이다. 직관적으로 도달한 판단에 저항하거나 자신의 직관을 수정해야하는 일은 노력이 필요하며 인지적인 부담을 느끼기 때문에 우리의 추론은 기존의 도덕 판단을 지지하는 방향으로 동기화된다. 그러나 이는 평가 단계가 편향적이었다는 것을 의미하는 것은 아니며, 다만 참여자들도 논증 생산에 참여했다는 것을 의미한다. 따라서 생성된 논증은 항상 상대방에 의해 평가되어야 한다.

마지막으로, 참가자는 적절한 동기를 부여받아야 한다(Merier & Sperber, 2011: 61). 적절한 동기란, 자신의 입장만을 고수하거나 타인의 입장만을 무비판적으로 따르는 것이 아니라 '진실 지향적이어야'함을 의미한다. 진정성이 있는 논증의 교환은 논증의 참가자가 자신의 도덕 판단이 다른 사람의 삶에 영향을 미친다는 것을 인식하고 있음을 시사한

다. 이러한 조건에서, 논증의 교환은 정보 평가를 통해 더 나은 결정을 내리는 결과를 낳을 수 있다. 여러 실험 연구들은 추론의 긍정적인 영향은 집단이 사실적이고 논리적인 답을 가진 과제를 다룰 때 자주 관찰했다(Laughlin & Ellis, 1986; Moshman & Geil, 1998). 진실이 승리하고 누군가가 올바른 해결책을 주장할 때 그들은 다른 모든 그룹 구성원들을 설득할 수 있기 때문이다. 그러나 본질적으로 사실보다는 가치의 영역에서 객관적인 평가가 어려운 도덕의 문제는 이러한 기준을 적용하는 것이 쉽지 않다(Uhlmann et al., 2009). 개인의 도덕 판단은 본질적으로 주관적인 것으로 보일 수 있지만 객관적인 기준을 적용한 평가 대상이 될 수 있다. 또한 타인과 함께 추론하는 것은 정확성을 증진시킴으로써 도덕 판단을 향상시키는데 도움이 될 수 있다(Blatt & Kohlberg 1975). 개인적 추론의 경우보다, 논증 교환을 통한 집단적 숙고의 과정에서 정확한 도덕 판단을 내리는 것으로 예측할 수 있다. 나아가 의사결정의 맥락에서 볼 때 도덕 판단은 '더 많은' 것을 결정하는 것이 아니라 '더 좋은' 것을 결정하는 것에 초점을 둔다. 진화적 관점에서 도덕적 행동을 유지하는 것은 좋은 협력자나 집단 구성원으로서 자신의 명성을 향상시키기 때문에 유용한데, 이때 논증은 사람들에게 그들의 도덕적 원칙에 반하는 행동의 잠재적인 결과를 상기시키는 데 사용될 수 있다. 이러한 기준은 도덕 판단과 의사결정에서도 진실성에 도달하고자 하는 동기를 부여하게 해준다.

2. 협력적 논증 교환의 교육적 타당성

도덕 판단에서 동기화된 추론의 역할에 주목한 하이트는 추론의 주된 기능이 직관에 의해 도출된 판단을 사후적으로 정당화하는 것으로 보기 때문에 개인이 스스로 자발적인 추론을 통해 초기 직관에 기반한 판단을 수정하거나 포기하는 일은 매우 드물다고 보았다. 하이트는 사회적 직관주의의 3번 링크에 해당하는 추론된 설득의 기제가 도덕 판단의 과정에 있어 추론이 직접적으로 기능하는 유일한 경우에 해당한다고 보았다. 추

론된 설득이란 자신의 도덕 판단을 정당화하기 위해 언어적으로 표현된 논증에 의해 타인의 직관에 영향을 미치는 기제이다. 그러나 그는 논증을 활용한 추론된 설득에 대해서는 회의적인 입장을 취한다. 특히, 상대의 정서적 태도 변화를 위해서는 추론에 의한 인지적 설득보다는 정서적 설득이 더욱 효과적이라는 연구(Edward & von Hippel, 1995)를 토대로 (Haidt, 2001: 819), 추론된 설득이 성공적으로 기능하기 위해서는 논리적 논증 교환이 아닌 내러티브 등을 활용을 제안했다.

> *사람들 사이에서는 추론 능력도 중요하며, 특히 추론이 새로운 직관을 불러일으키기도 한다는 점에서는 그 중요성을 무시할 수 없다. 둘의 관계에서 지배하는 쪽은 코끼리로(직관), 코끼리는 말 못하는 벙어리도 포악한 전제 군주도 아니다. 때로는 합리적 추론을 통해 직관이 형성되기도 한다. 특히 호의적 대화나 감정에 깊이 와닿는 소설, 영화, 기사가 합리성을 감싸 안을 때 추론을 통해 직관은 잘 형성될 수 있다(Haidt, 2012/2014: 146).*

이처럼 하이트는, 직관에 의해 도출된 최초의 판단을 수정하도록 작용하는 추론의 유일한 방식은 추론이 갖는 논리적 완결성에 의해서가 아니라, 정서적 영향력에 의한 설득에서 비롯된다고 보았다. 이처럼 초기 직관으로 도출된 판단을 바꿀 수 있는 과정은 논증 그 자체보다는 내러티브('나에게 꿈이 있다'로 시작하는 마틴 루터 킹의 연설)나 정서적 호소(죽어가는 아이들의 사진을 보여주는 것)에 있다(Haidt, 2001: 819). 그러나 추론의 사회적 기능에 주목한 추론의 논증 이론 연구자들은 도덕적 변화에 있어 논증 교환을 통한 설득의 영향력에 주목한다(Mercier, 2011a). 논증은 인식론적으로나, 실제적 영역에서 더 효율적인 의사결정에 기여하며, 다른 설득 수단 보다 도덕의 영역에서 우세한다. 메르시에는 내러티브의 영향력을 강조한 기존의 연구들을 검토하며, 해당 사례들에서 사용된 내러티브가 종종 논증의 일환으로 사용되었음을 지적한다 (Mercier, 2011a: 143). 예를 들어, 정보 통신 기술의 발달로 인해 작은 오류가 증폭될 수 있다는 점을 강조하기 위해 공무원의 사소한 도덕적 해이로 인한 작은 오류가 증폭되고 광범위하게 확산된다는 이야기를 사

용하는 경우이다(Ryfe, 2006: 77). 즉, 내러티브가 없는 논증과 논증의 일부가 아닌 내러티브는 구분이 쉽지 않다.

또한 도덕적 변화에서 논증적이지 않은 설득 수단의 명백한 역할이 대중 연설에서 부풀려질 수 있다는 점을 지적할 수 있다. 대중 연설가가 많은 청중들을 설득하고자 하는 양상은 협력적 논증 교환과는 두 측면에서 다르다. 하나는 청중들이 다양한 이유로 연설자의 입장에 동의하지 않을 수 있다는 것이며, 다른 하나는 청중들은 연설자와 대화하는 것이 아니기 때문에 어떠한 반론도 제기할 수 없다는 점이다. 이러한 특징은 대중 연설이 익명의 다수 청중들의 도덕적 태도를 변화시키고자 할 때 그 효과를 확인할 수 있다. 이와 달리 논증은 일상에서의 사적인 상황에서 특정한 상대방의 도덕적 변화에 적합하다. 윤리적 상황에 대해서 서로 다른 초기 직관으로 인해 다른 입장을 보이는 두 사람은 정서적 호소나 내러티브보다는 논증을 생성하고 평가하는 과정을 거치면서 자신의 견해를 진지하게 고민해보고, 검토하여 수정할 기회를 얻는다. 특히 협력적 논증 교환의 참가자는 상호 간의 신뢰를 기반으로 하고, 더 나은 도덕 판단과 문제 해결에 초점을 둔다. 이들은 서로의 논증을 진정성 있게 평가하고, 그 평가에 대해서도 겸허히 받아들일 준비가 되어있다. 협력적 논증 교환 참가자들은 자신의 신념을 위협하는 경우에도 정보를 왜곡하거나 편향되게 처리하기보다는 자신의 기존 입장에 대해 진지하게 고려할 것으로 기대할 수 있다.

사람들은 논증을 생성할 때, 종종 그들의 주장을 뒷받침하기 위해 증거와 추론에 의존하고, 이러한 주장을 구조적이고 따르기 쉬운 방식으로 제시한다. 또한 논증을 이해하고 평가할 때 타인이 제시한 증거와 추론을 비판적으로 분석하고, 객관적 기준에 따라 주장이 설득력이 있는지 판단한다. 이것은 개인들이 그들의 기존 신념과 태도에 도전할 수 있는 증거를 고려하도록 강요하기 때문에 동기화된 추론과 관련된 편향에 대응하는 데 도움이 될 수 있다. 나아가 논증을 생성하고 이해, 평가하는 것은 또한 사람들이 그들의 기존의 신념과 태도를 확인하는 방식으로 정보를 찾고 해석하는 경향인 확증 편향과 관련된 편견을 상쇄하는 데 도

움이 될 수 있다. 다른 사람들에게 자신의 주장을 제시하고 대화와 토론에 참여함으로써, 자신이 생각해보지 못한 관점과 새로운 시선에서 증거를 바라볼 수 있는 것이다. 이것은 그들의 관점을 넓히고, 객관적이며 잘 알려진 도덕적 판단으로 이끄는 데 도움이 될 수 있다.

동기화된 추론과 관련된 편향에 대응하는 것 외에도, 논증을 생성하는 것은 도덕적 추론에 대한 더 생산적이고 협력적인 접근을 촉진하는 데 도움이 될 수 있다. 개인들이 그들의 주장을 체계적이고 증거에 기반한 방식으로 제시할 때, 그것은 다른 사람들이 그들의 추론에 참여하고 건설적이고 존경스러운 방식으로 그들의 주장을 비판할 수 있게 한다. 또한 논증을 이해하고 평가함으로써 개인이 도덕적 문제에 대해 대화와 토론을 할 때, 그것은 건설적이고 존경스러운 방식으로 아이디어를 교환하고 주장을 비판할 수 있는 기회를 만든다. 이는 도덕적 담론을 위한, 보다 개방적이고 협력적인 분위기를 조성하는 데 도움이 될 수 있으며, 도덕적 원칙과 도덕적 딜레마에 대한 공유된 이해에 도달하는 데 더 도움이 될 수 있다.

한편, 교육적 맥락[47]에서 협력학습은 협동학습과의 비교를 통해 그 교육적 의미를 명료화할 수 있다. 협동학습은 잘 구조화된 하나의 과제에 대해 소집단으로 나누어 각자가 담당한 부분의 과제를 수행하는 것이다. 구성원 각각은 자신이 맡은 과제에 대해 책임이 있으며, 개별 과제를 합쳐 공동의 전체 과제를 해결하게 된다. 한편, 협력학습은 학생들이 하나의 과제를 상호작용하며 공동의 과제를 함께 해결하도록 설계된 학습 방식이다(Kirschner et al., 1996: 206). 협동학습과 협력학습의 차이를 살펴본 제이콥스는 일반적으로 이 두 학습을 구별하는 연구자들은 집단활동을 촉진하는 교사 중심의 방법으로 협동학습을 정의하는 한편, 학생 중심의 방법으로 협력학습을 정의하는 경향을 발견했다(Jacobs, 2015). 그에 따르면 용어 상의 차이는 있지만 일반적으로 협력학습이 협동학습에

47) 교육 연구에서도 학생과 학생의 상호작용에 대한 강조는 오래전부터 이어져 왔다(Dewey, 1929; Johnson & Johnson, 1975; Piaget, 1985; Vygotsky, 1978). 학생 간 상호작용을 촉진하기 위한 교육적 접근법으로 역시 협동학습(Gillies, 2007; Johnson, Johnson, & Holubec, 2008)과 협력학습(Bruffee, 1972; Kirschner et al., 1996; Alkis, 2022)을 강조한다.

비해 학생들 간 실시간 상호작용이 더 광범위하고 깊이 있게 나타난다.

협력학습의 심리학적 기원은 피아제의 인지발달론과 비고츠키의 사회적 구성주의에서 찾을 수 있다. 일찍이 피아제는 아동의 인지 발달은 또래와의 사회적 상호작용에 의해 촉진된다는 사실을 발견한다. 피아제에 따르면 우리가 자신의 생각을 타인에게 말하고, 타인의 의견을 듣고자 하는 이유는 자신의 입장을 검증받고자하는 근본적인 욕구에 기반한다(Piaget, 1928: 204). 인지 발달의 주요 요인 중 하나인 사회적 상호작용은 아이들의 경험과 타인을 포함한 그들의 환경과의 관계 속에서 세상에 대한 아동의 이해를 형성하는 데 결정적인 역할을 수행한다. 피아제는 사회적 상호작용을 통해 아이들이 새로운 생각, 관점, 사고방식에 노출될 수 있고, 이는 정신의 구조가 발달하는 기회가 된다. 나아가 사회적 상호작용은 아이들이 세상에 대한 이해를 시험하고 수정할 수 있는 기회를 제공한다. 예를 들어, 아이들이 협동적인 놀이나 다른 사람들과의 대화에 참여할 때, 다른 관점과 생각에 노출되고, 이것은 아이들이 기존에 가지고 있던 신념과 가정에 도전할 수 있다. 이러한 협력과 상호 적응의 과정을 통해, 아이들은 다른 사람들의 관점을 고려하고 세상에 대한 자신의 이해를 수정하는 것을 배우게 되는 것이다.

한편, 비고츠키의 사회적 구성주의는 인간의 발달과 학습에 대한 이론으로, 고등 정신 기능의 발달에서 사회적 상호작용의 역할을 강조한다(Vygotsky, 1978). 추론의 논증 이론가들과 마찬가지로 그는 학습이 고립된 과정이 아니라, 다른 사람들과의 목적적이고 의미 있는 상호작용을 통해 일어난다고 주장한다. 비고츠키는 아이들이 자신보다 더 지식이 많거나 숙련된 다른 사람들과의 상호작용을 통해 가장 잘 배우고, 문제 해결, 추론, 언어와 같은 고등 정신 기능의 발달이 사회적 상호작용과 밀접하게 연관되어 있다고 믿었다. 또한 비고츠키는 신념과 지식의 구성은 사회적 맥락에서 발생하며, 이는 각 개인의 문화에 의해 형성된 각 개인의 언어, 기술 및 경험을 기반으로 하는 실제 문제 또는 작업에 대한 학생-학생 협력을 포함한다고 보며(Vygotsky, 1978: 102), 협력 학습이 사회 구성주의의 핵심적인 측면이라고 믿었다. 그는 다른 사람들과의 협력

은 개인들이 새로운 정보와 기술을 내면화할 수 있게 하며, 이러한 내면화는 더 높은 정신 기능의 발달에 필수적이라고 주장했다. 협력을 통해 개인이 공동으로 문제 해결에 참여하고 스스로 이룰 수 없었던 새로운 이해와 기술을 개발할 수 있다. 이와 같이 협력의 중요성은 비고츠키의 사회적 구성주의를 이루는 큰 두 개의 축인 근접 발달 영역(Zone of Proximal Development, ZPD)과 비계 설정[Scaffolding]의 관계 속에서 더욱 구체화 된다. 근접발달영역이란, 현재의 인지 발달 수준과 잠재적인 인지 발달 수준의 차이로 이해된다. 비계란 학습자의 ZPD를 확인하여 학습과 발달을 촉진할 수 있도록 적절히 돕는 행위이다. 비계를 통해 개인은 공동의 문제 해결에 참여할 수 있고, 스스로 달성할 수 없었던 새로운 이해와 기술을 개발할 수 있다. 비계는 모델링, 피드백, 직접적 교수 뿐만 아니라 건전한 논증의 교환을 통해서도 이루어진다. 즉, 이와 같은 견지에서 협력적 논증 교환은 단순히 함께 일하는 문제가 아니라, 논증 참여자들이 서로의 비계를 설정해줌으로써 함께 학습해나가고 발달하는 과정인 것이다.

추론의 사회적 기능에 주목한 발달 심리학자 모쉬맨은 협력적 추론의 사용을 개인이 특정한 문제에 대한 공유된 이해나 해결에 도달하기 위해 근거와 주장을 교환할 수 있는 구조화된 대화에 참여하는 과정으로 본다(Moshman, 1995). 모쉬맨에 따르면, 협력적 논증 교환에서 제시된 논증들은 그 자체로 비계로 작용한다. 다른 사람들과 이유와 주장을 교환하는 과정은 그들이 대안적 관점을 고려하고 반론에 비추어 그들 자신의 주장을 방어하도록 강요받기 때문에 개인들이 보다 객관적인 관점을 채택하는 데 도움이 될 수 있다. 가령 한 연구(Moshman, 1995)에서 참가자들은 결론에 도달하기 위해 다른 사람들의 관점을 고려하고 여러 증거를 통합해야 하는 협력적 추론 작업에 참여하도록 요청받았다. 결과는 참가자들이 협력적인 환경에서 일할 때 동기화된 추론의 영향을 덜 받고 체계적으로 추론에 참여할 가능성이 더 높다는 것을 보여주었다. 이처럼 협력적 문제 해결을 가능하게 해주는 추론의 기능은 개인의 편향과 정서의 영향에 덜 민감한 사유와 주장을 교환할 수 있는 구조를 제공하기 때

문에 동기화된 추론의 개선에 기여할 수 있다.

하나의 윤리적 이슈에 대한 서로 다른 의견을 주장과 근거의 형태로 교환하는 협력적 논증 교환은 이러한 맥락에서 '협력'적이다. 논증 교환의 참가자들은 단지 자신의 논증을 생성하고, 그것의 정당성을 평가하는 것에 그치지 않고 이를 바탕으로 타인을 설득하기도 하고, 타인이 생성한 논증을 평가하기도 한다. 그 과정은 자신의 기존 신념을 되돌아보도록 서로가 서로를 돕는 모습이다. 상대방의 논증에 대한 나의 평가는, 상대방이 자신의 기존 신념과 지식의 결함을 재고해볼 수 있는 기회를 제공해주기 때문이다.

협력적 논증 교환은 동기화된 추론의 영향력을 줄이고 체계적이고 객관적인 추론의 사용을 증가시키는 데 도움이 될 수 있다. 협력적 논증 교환에 참여하는 개인은 지지적이고 존중받는 분위기 속에서 선입견이나 가치보다는 비판적 사고에 참여하고, 다양한 관점과 주장에 노출되어 대안적 관점을 고려하며, 증거와 추론에 근거하여 결정을 내리는 경향이 더 높기 때문이다(Johnson & Johnson, 1989).

협력적 논증 교환이 동기화된 추론을 극복할 수 있는 핵심 메커니즘 중 하나는 자신의 논증이 모순임을 입증하는 근거의 제공을 통해 이루어진다. 동기화된 추론의 초기 연구 중 한 연구에 따르면(Lord et al., 1979), 사람들은 자신의 논증이 모순임을 입증하는 근거에 대한 사회적 교환은 자신의 기존 신념에 도전하고 자신의 관점에 재평가를 요구하기 때문에 초기 신념을 객관적으로 파악하는 기회로 작용한다. 유사한 또 다른 연구(Edwards & Smith, 1996)에서는 참여자들이 자신의 기존 신념과 모순되는 논증을 자세히 검토하기 위해 더 오랜 시간을 보냈으며, 자신의 신념과 일치하는 논증보다는 모순되는 논증에 반응하기 위해 더 많은 논증을 생성하였다. 특히 이러한 연구 결과들은 자신의 논증이 모순임을 입증하는 근거가 개인적으로 제시되었을 때 동기화된 추론이 오히려 강화된다는 결론과 비교된다. 이처럼 협력적 논증 교환에서, 개인은 타인의 다양한 논증에 노출되며, 이를 토대로 그들의 동기화된 추론에 도전하는 근거에 대해 진지하게 고려해볼 수 있게 된다.

또한 협력적 논증 교환은 사유와 논증의 교환을 위한 구조를 제공함으로써 개인이 보다 체계적이고 객관적인 추론에 참여하도록 장려할 수 있다. 개인들이 협력적 논증 교환에 참여할 때, 그들은 자신의 논증을 생성하여 공개적으로 표명하고, 또 표명된 다른 사람들의 논증을 평가하도록 요구 받고, 타인에 의해 평가된 자신의 논증에 대한 피드백을 제공하는 등 동기화된 추론의 사용을 검토하고 이를 개선하기 위한 구조화된 기회를 제공받는다. 이러한 규칙과 규범은 자신의 주장을 뒷받침하기 위한 증거와 이유를 제공하고, 다른 사람의 주장을 경청하며, 제시된 증거와 주장에 대한 비판적 평가와 분석에 참여하는 요구사항을 포함할 수 있다. 이와 같은 논증 교환의 구조는 일련의 규칙과 규범으로 체계화된다. 논증 교환의 참여자들은 반드시 자신의 주장을 뒷받침하기 위한 증거와 이유를 제시해야한다. 단순히 의견을 교환하는 것이 아니기 때문에, 논증 생성자는 주어진 도덕적 문제에 대한 자신의 정서적 반응으로 판단이 도출되었다 할지라도, 그것을 뒷받침하는 사실적 차원의 증거나, 논리적인 이유를 제공해야하는 의무를 지닌다. 또한 논증 교환의 참여자들은 타인을 존중하며, 다른 사람의 의견과 논증을 경청해야 한다. 협력적 논증 교환이 동기화된 추론의 개선에 기여하기 위해서는 논증 참여자들이 서로를 신뢰해야 한다. 이러한 신뢰를 바탕으로 할 때, 타인이 제공한 논증의 평가를 객관적으로 수용할 수 있으며 자신의 기존 신념을 수정할 가능성이 커지기 때문이다. 또한 논증의 평가자는 논증 생성자에게 제시된 증거와 이유에 대한 비판적 평가와 분석에 참여해야만 한다. 논증의 평가자는 단순히 타인의 논증에 동의 여부를 표현하거나, 지지해주는 역할이 아니다. 논증 생성자가 제시한 증거의 신빙성을 확인하고, 주장과 근거의 관계에 대해 논리적으로 분석할 수 있어야 하며, 그 논증에 대한 자신의 평가를 포함해야 한다. 이러한 구조는 개인이 제시된 정보와 주장에 대해 보다 체계적이고 객관적인 평가에 참여하고 대안적 관점을 고려하도록 함으로써 동기화된 추론의 영향력을 줄이는 데 도움이 된다.

제 5 장 협력적 논증 교환의 도덕교육적 활용

제 1 절 협력적 논증 교환 수업의 특징

1. 동기화된 추론 개선을 위한 수업 목표 설정

협력적 논증 교환은 도덕 판단에서 발생할 수 있는 동기화된 추론의 문제를 개선하기 위한 목적으로 도덕 수업에서 활용될 수 있다. 이 교수·학습활동의 목표는 높은 정확성의 추구와 적절한 방향성의 탐색을 통해 도달될 수 있다. 동기화된 추론의 개념을 도입하고 유형화한 대표적인 연구들(Kunda, 1990; Lodge & Taber, 2006)은 정확성 동기가 편향에 빠지지 않고 객관성을 유지하며 올바른 판단을 유도하는 것으로 보고한다. 또한 도덕 판단에서 문제가 되는 동기화된 추론을 개선하기 위해 방향성을 제거하는 것은 현실적으로 불가능하며 바람직하지도 않다.

동기화된 추론 연구는 추론의 인지 과정이 정서와 불가분한 관계에 있기 때문에 항상 일정한 방향성에 의해 동기부여된다고 전제한다. 방향성을 제거하는 것은 도덕 판단에서 나타나는 추론의 심리 과정의 현실성을 고려하지 못한 것이다. 물론 추상화되고 이상화된 도덕적 추론의 모습은 아무런 사심 없이 엄격한 도덕적 원칙의 논리적 적용을 의미할 수 있다. 그러나 현재까지의 동기화된 추론 논의에 비추어 보았을 때, 일상의 도덕 판단에서 추론은 정서적 반응으로 인해 특정한 결론에 도달하고자 하는 일정한 방향성에 의해 동기화된다. 특히, 추론이 올바른 도덕적 정체성에 기반한 방향성-도덕적으로 바람직한 결론에 도달하고자 하는 방향성-에 의해 동기화된다면 그것은 개선이나 극복의 대상이 아닐 것이다. 따라서 동기화된 추론을 개선하고자 하는 협력적 논증 교환 활동은 도덕적 상황에 대한 정서적 반응으로 나타나는 방향성을 제거하거나 그 수준을 낮추기보다는 학생들에게 방향성을 재점검하는 기회를 제공해야 할

것이다. 정리하자면, 동기화된 추론을 개선하기 위한 목적으로 수행되는 협력적 논증 교환 활동은 일차적으로 학생들이 높은 정확성을 지향하면서 동시에, 도덕적으로 적절한 수준의 방향성을 갖도록 이끌어야 한다.

1) 정확성 목표의 추구

먼저, 협력적 논증 교환 활동은 정확성 동기의 함양을 통해 '객관성의 환상'의 극복을 지향해야 한다. 자신의 신념과 판단이 객관적인 증거와 추론에 근거한다고 믿는 경향을 의미하는 객관성 환상은 자신이 (직관적으로) 내린 도덕 판단에 부합하고자 하는 강한 방향성 목표에 의해 동기화된 추론을 통해 유지된다. 객관적 환상에서 벗어나는 일은 추론이 '정확성 목표'를 추구할 때 가능해진다. 정확성 목표와 동기화된 추론으로 인해 유발되는 인지 편향 사이에는 부적 상관관계를 보이는 데, 그 이유는 세 가지로 분석된다(Kunda, 1990: 481-482). 먼저 정당화 요구 조건이다. 실험 참가자들은 자신의 판단이 타인에 의해 평가될 것이라고 기대하기 때문에 더 신중하게 정보와 지식을 해석했다. 즉, 실험 참가자들이 정확성 목표로 동기화되는 경우에는 판단의 결과가 곧 평가의 대상임을 인식했기 때문에 인지적 편향에 빠지지 않도록 노력하는 것이다.

다음은 공개성의 조건이다. 실험 참가자들이 자신의 판단이 공개될 것으로 기대하는 경우이다. 특정 이슈에 대한 자신의 판단의 결과가 내적인 마음의 상태에만 영향을 미치게 될 경우에는 인지적 편향에 쉽게 노출될 수 있다. 그러나 사람들은 자신의 판단이 공개적으로 표출될 것이라고 예상하는 경우에 추론은 정확성 목표에 의해 쉽게 동기부여되며, 최대한 객관적이고 공정하게 논증을 만들어내었다.

마지막으로 진정성의 조건이다. 자신이 내린 판단이 타인의 삶에 영향을 미칠 것으로 예상되는 경우이다. 우리는 우리의 판단이 누군가의 삶에 직접적인 영향을 미친다는 것을 인식할 때, 더욱 신중하고 진정성 있는 판단을 내리고자 한다. 이처럼 정확성 목표의 존재는 동기화된 추론을 하는 경우에도 인지 편향에 덜 의존하게 되는 가능성을 시사한다. 정

확하도록 동기부여되는 경우 더 적절하다고 여겨지는 정보를 처리하기 위해 보편적인 규칙을 적용하고자 하고, 보다 복잡한 인지적 전략에 접근할 가능성이 높기 때문이다. 즉, 판단을 내림에 있어 ①정당화 요구 조건, ②공개성의 조건, ③진정성의 조건과 같은 적절한 정확성 목표의 조건이 충족된다면 동기화된 추론의 과정을 통해서도 객관적이고 공정하게 판단을 도출해낼 수 있을 것으로 기대할 수 있다.

협력적 논증의 교환 활동은 이와 같은 세 가지 조건을 고려하여 설계될 때 정확성 목표를 함양할 수 있다. 먼저, 수업에서 협력적 논증 교환 활동을 할 때 교사는 학생들에게 자신의 판단에 대한 정당화를 요구해야 한다. 학생들은 논증을 생성하는 과정을 통해 제시된 윤리적 문제에 대한 자신의 최초 도덕 판단이 정당한 것인지 확인할 수 있다. 이는 자신의 논증이 타인을 설득하기 위한 것임을 염두에 두지만 동시에 그것이 정당한 것인지를 스스로 확인함으로써 '객관성 환상'으로부터 벗어날 계기가 된다. 동기화된 추론 연구는 사람들이 그들이 원하는 것을 단순히 믿거나 행하는 것이 아니라 그렇게 해야 할 이유를 찾아 정당화가 될 때 비로소 따른다는 사실을 보여준다(Kunda, 1990). 반대로 우리가 내부적으로 이유를 찾고 스스로 충분한 설득력을 얻을 수 있는 이유를 찾지 못해 정당화에 실패한다면, 우리의 마음을 바꿀 수 있다는 것을 의미한다(Haidt 2001). 우리의 최초 판단을 적절히 방어하지 못한다면, 우리는 일관되지 않고 불합리하다고 생각되는 고통을 피하기 위해 기존의 입장을 정확하게 수정해야하기 때문이다(Mercier, 2011a). 한 사람에 의해 일방적으로 제공되는 단방향 논증은 일반적으로 실제 토론과 논증 교환보다 사람들의 마음을 바꾸는 데 덜 효과적이며(Mercier et al., 2017: 695), 서로에게 정당화를 요구하고 이를 평가하는 과정을 포함하여 협력적으로 논증을 교환하는 과정을 통해 기존 신념을 기꺼이 수정할 수 있다(Mercier & Sperber, 2017/2018).

둘째, 협력적 논증 교환 활동에서는 참가하는 학생 모두 자신의 입장을 공개적으로 표현해야 한다(Mercier & Sperber, 2011). 추론의 단독적인 사용과 달리 추론의 사회적 기능을 고려한 공개적 논증 교환의 상황

에서는 자신이 생성한 논증은 반드시 타인의 평가의 대상이 되어야 하기 때문이다. 논증을 생성한 학생들은 동시에 다른 학생이 생성한 논증을 평가하는 역할을 수행한다. 추론의 논증 이론에 따르면 사람들은 자신의 논증을 생성할 때는 편향적인 판단을 도출하는 경향이 있지만, 공개적으로 자신의 견해에 도전하는 논증을 평가할 때는 객관성을 유지한다. 그들은 형편없는 논증(poor arguments)에 의해 자신의 신념이 흔들리는 것을 막기 위해 공개적인 논증 교환 활동에서 엄격하게 논증을 평가하기도 하고, 상대방에게 양질의 논증을 요구하기도 한다. 이처럼 추론의 사회적 기능으로서 논증의 생성과 평가는 정확성에 의한 동기화의 측면에서 보았을 때 근본적으로 비대칭적이다. 그러나 협력적 문제 해결을 위한 목표를 공유하고 이를 위해 공개적인 논증 교환에 참여하게 될 때, 우리는 편향된 논증을 생성해내기도 하지만, 생성된 논증의 객관적 평가에 참여하고자 하는 정확성에 의해 동기화되며 이를 통해 편향된 판단을 개선할 수 있다.

셋째, 협력적 논증 교환은 서로의 삶에 의미 있는 영향을 주는 진정성 있는 의견의 교환을 전제할 때 정확성 목표에 의해 동기화될 수 있다. 여기에서 '서로의 삶에 의미 있는 영향'이란 협력적 논증 교환 활동의 참여자들이 자신의 기존 도덕적 신념을 수정할 기회를 갖게 된다는 것을 의미한다. 협력적 논증 교환에서 진정성이 있는 논증의 교환은 논증의 참가자가 자신의 도덕 판단이 다른 사람의 삶에 영향을 미친다는 것을 인식하고 있음을 시사한다. 추론의 기능이 사회적인 것에 있음을 밝힌 메르시에는 자신과 입장이 다른 상대방이 제시한 논증에 동의하지 않더라도 그것이 자신의 논증에 대해 진정성 있는 반대 의견으로 수용될 수 있으며, 이때, 개인은 도덕적 신념을 성찰하고, 더 나은 결정에 도달하는 데 도움이 된다고 보았다(Mercier et al., 2017: 111). 이처럼 논증 교환을 통해 정확성 목표를 높이고 동기화된 추론을 개선하기 위해서는 대립적인 의견을 포함한 다양한 의견이 존재해야 한다. 논증 교환에서 의도적인 대립적 견해를 제시하는 것의 효과에 대한 연구들은 인위적으로 만들어낸 반대 의견이 동기화된 추론을 개선하는 데 효과적이지 않았음을 지

적한다(Nemeth et al., 2001). 즉, 자신의 도덕 판단이 단지 순간의 토론을 위해 만들어낸 것이거나 일시적인 의견에 의해 반박되거나 위협을 받게 되면, 오히려 자신을 보호하려는 방향성 동기를 강화하고 이는 정확성 동기 함양에 부정적인 영향으로 이어진다. 수업 시간에 이루어지는 토론 활동이 서로의 의견 차이를 확인하는 것에서 그치거나, 토론 이전에 비해 오히려 태도의 양극화를 심화시키는 경우도 있다. 따라서 교사는 협력적 논증 교환의 참여자들이 자신의 도덕 판단이 본인의 삶에 적용되고 타인의 삶에 영향을 미친다는 사실을 인지할 수 있도록 동기유발하고, 진정성 있는 논증 생성과 평가가 이루어질 수 있도록 논증 교환 집단을 적절하게 구성해야 할 것이다.

위 세 가지 조건에 더해, 협력적 논증 교환 활동은 의도적 무지(willful/deliberate ignorance)의 경계를 하위 목표로 설정해야 한다. 의도적 무지란, 도덕적 상황과 관련하여 개인적으로 관련성이 있고 쉽게 접근할 수 있지만 잠재적으로 불리한 비윤리적 속성에 대한 지식을 의도적으로 알지 못하도록 선택하는 것을 의미한다(Moody-Adams, 1999). 세계 곳곳에 여전히 존재하는 노예 제도를 비판한 무디-아담스는 '사회적 실천에서 도덕적 진보의 주요 장애물은 이미 알 수 있고, 알아야 할 것에 대한 의도적 무지가 만연하기 때문(Moody-Adams, 1999: 180)'이라고 지적하며 우리의 도덕적 삶을 방해하는 의도적 무지의 경계를 강조했다. 의도적 무지는 자신의 기존 신념, 태도 또는 가치에 도전하거나 모순될 수 있는 정보를 의도적으로 회피하도록 한다. 특히, 의도적 무지는 올바른 정보를 접하기 어렵거나, 높은 지식수준을 요구하는 등의 외부적 장벽보다는 자신의 의지로 인한 무지에서 비롯된다. 이러한 형태의 무지는 진실을 추구하려는 정확성 동기보다는 자신의 현재 신념과 태도를 유지하려는 욕구에 의해 동기부여 된다. 동기화된 추론 연구자들 또한 의도적인 무지가 사람들이 자신의 신념이나 태도에 도전할 수 있는 정보를 찾거나 처리하는 것을 피하는 동기화된 추론의 원인이라고 주장했다(Ehrich & Irwin, 2005; Mata et al., 2022; Paharia et al., 2013).

의도적 무지는 협력적 논증 교환 수업을 통해 적절히 개선될 수 있다.

협력적 논증 교환 수업에 참여하는 학생들은 서로의 관점을 이해하고 공통점을 찾는 것을 목표로 존중하고 열린 마음의 교류를 하는 소통의 과정을 경험하기 때문이다. 특히 이 과정은 건설적 대화에 참여하는 과정과 유사한데, 이를 통해 학생들은 기존의 편견과 선입견의 영향을 줄이고, 다른 관점에 대한 이해를 높일 수 있다. 한 연구는 사람들이 다른 정치적 견해를 가진 다른 사람들과 건설적인 대화를 할 때 양극화된 사고를 할 가능성이 적었고, 정보에 대해 보다 수용적인 태도를 지니게 되며, 다양한 의견들이 갖는 공통점을 찾을 가능성이 더 높다는 것을 발견했다(Iwry, 2021). 이처럼 협력적 논증 교환 수업에 나타나는 건설적 대화의 양상은 학생들에게 자신의 신념에 맞서고 도전할 수 있는 기회를 제공하고, 그러한 근거를 객관적이고 폭넓게 이해하도록 독려하기 때문에, 의도적인 무지의 문제를 극복하는 데 특히 효과적일 수 있다.

한편, 방향성에 의해 동기화된 추론의 폐해를 밝히고 이를 극복의 대상으로 보고 그 대안으로 정확성에 의해 동기화된 추론의 중요성을 강조한 갈렙(2021)은 자신이 정확성에 의해 동기화된 추론을 하는 사람인지 확인할 수 있는 다섯 가지 지표를 제시하였다(Galef, 2021: 83-90). 이는 협력적 논증 교환 수업의 목표 설정을 함에 있어 적절히 수정하여 참고할만하다.

첫째, 상대가 옳았음을 깨달은 다음 그 사람에게 이 사실을 말할 수 있는가? 동기화된 추론의 개선은 원칙적으로, 도덕적 주체 스스로 자신의 판단이 틀렸음을 인정한다면 해결될 수 있다. 그러나 앞서 살펴본 것처럼, 개인적 차원의 성찰은 그 자체로 여러 어려움을 내포하고 있으며 특히 추론의 사회적 기능을 고려하였을 때 보완될 수 있다. 특히 우리의 추론은 타인의 논증을 평가할 때 편향에 빠지지 않고 객관적이고 사심 없는 판단을 도출한다. 따라서 협력적 논증 교환 수업 상황에서 학생들은 자신의 논증을 생성하고 그것을 검증하는 일과 더불어, 자신의 동료가 제시한 논증의 진위를 평가하는 일에도 매진해야 한다.

둘째, 나를 비판하는 사람에게 어떻게 반응하는가? 갈렙은 비판을 기꺼이 받아들이는지 측정하기 위해 ‘나는 비판에 너그러운가?’라고 자문하

는 것으로는 부족하다고 지적한다(Galef, 2021: 85). 추론의 사회적 기능을 고려한다면, 우리의 동기화된 추론은 타인에 의해 논증이 평가될 때 근본적으로 개선될 수 있다. 따라서 협력적 논증 교환의 상황에서 우리는 타인이 자신의 판단을 전면적으로 비판하고 부정하는 경우에도 기꺼이 경청할 준비가 되어 있어야 한다. (나아가 실제로 자신이 타인의 비판에 대해 어떻게 반응했는지를 추적해볼 필요가 있다. 타인의 비판을 받아들여 자신의 판단이나 행동을 바꾼 경험이 있는지, 비판한 사람에게 적절한 반응(감사, 보상 등)을 했는지 등이다.

셋째, 자신이 틀렸다는 것을 입증할 수 있는가? 동기화된 추론이 문제가 되는 이유 중 하나는 편향된 자신의 도덕 판단이 자동적이고 직관적인 사고를 통해 도출된 것이 아니라, 나름의 합리적인 논리적 추론을 통해 도출되었다는 점에 있다. 따라서 자신의 판단이 '옳다는' 근거를 갖춘 상태이기 때문에 자신이 틀렸다는 것을 입증하기란 쉬운 작업이 아닐 수 있다. 이를 위해서는 자신의 도덕 판단을 다른 사람들에게 공언하고, 만약 그것이 틀렸을 때, 무엇이, 왜 틀렸는지 제시할 수 있어야 한다. 이렇게 자신의 오류를 다른 사람 앞에서 인정한다는 것은 자기 보호의 목표보다 진실을 추구하는 지표로 사용될 수 있다.

넷째, 자기 기만을 피하는 예방적인 조치가 있는가? 동기화된 추론은 그로 인한 문제가 발생하기 이전에 예방하는 것이 더욱 중요할 것이다. 한 번 동기화된 추론에 빠지게 되면, 그로 인해 도출된 신념이 옳다고 스스로 진심으로 여기기 때문에 그것을 개선하기란 쉽지 않다. 따라서 동기화된 추론이 발생하지 않도록 사전적인 조치가 필요할 것이다.

마지막으로 다섯째, 좋은 비평가를 곁에 두었는가? 정확성 목표에 의해 동기화된 추론을 하는 사람은 결코 자신의 주장에 대해 과신하지 않는다. 또한 자신의 생각이나 의견을 가감 없이 비평해줄 동료를 곁에 두어야 한다. 협력적 논증 교환 수업은 교실 속 동료 학생들이 진정한 비평가가 되는 것을 목표로 한다. 사형 제도, 낙태 문제와 같이 윤리적, 정치적인 태도의 양극화가 일어나는 윤리적 사안에 대해 의견을 달리하거나, 기후 변화와 환경 문제 등과 같이 과학적 사실을 다르게 이해하는

사람들과 함께 논증을 교환함으로써 자신의 판단이 동기화된 추론에 의해 편향되었을 수 있음을 점검할 수 있다.

2) 도덕적으로 바람직한 방향성 목표의 설정

정확성 목표의 추구와 더불어 동기화된 추론의 개선을 위해 협력적 논증 교환 활동은 학생들이 자신의 방향성 목표를 재점검하고, 적절한 방향성을 탐색하는 기회가 되어야 한다. 특히, 방향성에 의해 동기화된 추론은 주로 개인의 기존 신념과 사회적 정체성의 보호라는 자기 이익에 근거하기 때문에 사심 없는 도덕적 추론을 방해한다. 이를 개선하기 위해 제시된 기존의 개인적 차원의 인지 편향을 극복방안들은 공통적으로 '감정의 동요'로부터 벗어나 추론에서 방향성을 제거하고자 하였다. 그러나 추론의 심리 과정에 대한 최근 연구에서 살펴본 것처럼, 실제 우리의 도덕 판단의 과정에서 추론의 인지 과정과 정서는 분리 불가능하다. 특히, 추론의 과정에 결합된 정서의 역할이 추론의 방향에 영향을 미친다는 점에서 도덕적 추론은 도덕적 상황에 대한 감정이 이끄는 방향에 의해 동기화될 것이다. 특히, 추론의 사회적 기능에 관한 앞선 논의를 참고했을 때, 타인과의 상호작용 속에서 작동하는 추론은 자신의 논증을 통해 타인을 설득하고자 하는 방향성 목표에 의해 동기화될 것이다. 따라서 동기화된 추론을 개선하기 위한 협력적 논증 교환 활동은 방향성을 제거하기보다는, 적절한 방향성에 의한 동기화된 추론을 추구해야 한다.

이때 적절한 방향성 수준은 방향성 목표가 기반하는 '정체성'과 협력적 논증 교환이 지향하는 '협력'의 모습에서 찾을 수 있다. 편향된 도덕 판단을 유발하는 동기화된 추론은 자신의 정체성을 보호하고자 하는 방향성을 내포한다. 이것이 문제가 되는 까닭은 자신의 정체성을 보호하고자 하는 방향성 동기가 적절한 수준을 넘어서 자기본위적인 것으로 흐르기 때문이다. 따라서 사람들은 동기화된 추론을 통해 자신의 도덕적 위선 행위를 정당화하기도 하고, 자신과 입장이 다른 사람의 견해를 과소평가하고 자신의 입장에 부합하는 증거나 새로운 정보는 편향적으로 해석한

다. 이처럼 동기화된 추론은 도덕적, 정치적 양극화를 가속화하는 기제로 작동하기도 한다. 나아가 적절한 수준을 넘어서는 방향성 동기는 추론 기술의 부재나 인지 능력의 결핍이 아니라 오히려 그 반대로 성숙한 인지 능력을 지닌 엘리트 사이에서도 흔히 발견되었다. 이때 성공적인 협력적 논증 교환은 한 사람의 추론과 그의 정체성을 분리시켜줌으로써 (Kuhn, 1991; Mercier & Sperber, 2011, 2017; Pinker, 2021: 573) 방향성 동기의 적절한 수준을 탐색하게 해준다. 협력적 논증 교환의 활동은 논증을 생성하여 공개적으로 제시하고, 공개된 논증은 타인에 의해 객관적이고 엄격한 잣대로 평가받는 과정을 반드시 거친다. 이 과정에서 지나친 자기 방어적인 방향에 의해 도출된 논증은 거부될 것이다. 협력적 논증 교환에 참여하는 참가자들은 서로에 대한 신뢰를 바탕으로 진정성 있는 반대 의견을 제시하기 때문에 자신의 정체성을 보호하고자 하는 동기는 자신의 도덕적 논증으로부터 분리될 수 있는 것이다. 이 과정을 좀 더 자세히 살펴보자.

동기화된 추론의 근원은 자신이 가지고 있던 기존의 신념에 있으며, 그 신념은 강한 직관적 반응을 산출하는 도덕적 감정으로 이루어져 있다. 자신의 기존 신념이 지시하는 방향에 의해 동기화된 추론은, 자신의 가치를 지지하는 확증적인 근거는 과대평가하고, 자신과 반대되는 가치 혹은 자신의 가치를 위협하는 입장에 대한 근거는 과소평가한다. 동기화된 추론이 근거하는 도덕적 가치는 하이트가 제안한 도덕 기반의 모습과 유사하다. 도덕적 기반 이론은 인간의 도덕적 가치의 기원과 다양성을 설명하고자 하는 이론이다(Graham et al., 2013; Haidt, 2007). 하이트 연구진이 고안한 이 이론은 인간의 도덕적 가치가 개인과 공동체가 생존하고 번영하는 것을 돕기 위해 진화한, 제한된 직관적 기반 또는 선천적인 심리적 시스템에 기초한다고 가정한다. 이러한 기반에는 보살핌/해로움, 공정/부정, 충성/배신, 권위/전복, 신성/변절이 포함되며, 문화와 역사적 시기에 걸쳐 보편적인 것으로 받아들여진다.

도덕 기반 이론에 따르면, 우리가 가지고 있는 기존의 도덕적 신념과 지키고자 하는 도덕적 가치는 이성과 보편적 원칙에 기초하는 것이 아니

라, 우리의 진화 역사와 문화적 경험에 의해 형성된 정서와 직관에 기초한다. 이것은 사람들의 도덕적 가치가 반드시 의식적인 선택이나 추론의 결과가 아니라 생물학과 문화의 복잡한 상호작용에서 나온다는 것을 의미한다. 한 문화권에서는 충성과 권위에 더 큰 강조를 둘 수 있는 반면, 다른 문화권에서는 보살핌과 공정성을 우선시할 수 있다. 서로 다른 도덕적 기반을 가진 사람들이 서로의 관점을 이해하고 이해하는 데 어려움을 겪을 수 있기 때문에 이러한 도덕적 가치의 다양성은 때때로 개인과 집단 간의 갈등과 오해를 초래할 수 있다. 동기화된 추론은 이러한 갈등과 오해를 더욱 부추기며, 도덕적 위선의 문제나, 도덕적, 정치적 양극화의 심화 문제를 낳기도 하였다.

그러나 이러한 다양성에도 불구하고 도덕 기반 이론은 모든 인간이 공통의 도덕 문법, 즉 문화와 개인에 걸쳐 보편적인 기본적인 도덕 개념과 정서를 공유한다고 제안한다. 이 일반적인 도덕 문법은 사람들이 올바른 행동 방침에 대해 다른 결론에 도달할지라도 유사한 방식으로 도덕적 문제를 이해하고 대응할 수 있게 해준다. 이 지점이 협력적 논증 교환 활동을 통해 동기화된 추론이 개선될 수 있도록 하는 기제이다. 나와 입장이 다르고, 중요시하는 도덕적 가치가 다른 타인이지만, 여전히 우리는 동료 인간으로서 이해 가능한 범위 내에서 도덕 기반을 공유하고 있다. 따라서 우리는 자신의 기존 도덕적 신념에 위협을 받은 상황일지라도, 협력적 논증 교환의 테두리 내에서는 기꺼이 자신의 신념을 수정할 수도 있고, 추론된 설득과 사회적 설득 등의 양태로 타인의 도덕적 직관에 영향을 줄 수도 있다.

따라서 이러한 공통적인 도덕 기반에 대한 이해는 동기화된 추론을 개선하기 위해 협력적 논증 교환 활동을 활용할 때, 교사와 학생이 반드시 유념해야 할 수업의 원칙 중 하나이다. 극심한 도덕적 불일치 상황에서 우리는 협력적 논증 교환에 앞서, 도덕적 가치의 다양성에 대한 인식과 이해를 공유해야 한다. 도덕 기반 이론이 가정하는 것처럼 인간의 도덕적 가치가 개인과 공동체가 생존하고 번영하는 것을 돕기 위해 진화한 직관적 기초 또는 선천적인 심리적 체계의 제한된 세트에 기초한다는 것

을 이해할 필요가 있는 것이다. 우리 각자가 발 딛고 있는 도덕 기반이 단지 개인적인 차원의 것이 아니라 오랜 진화적 축적으로 인한 생물학적, 문화적 토대에 의해 형성된 것이며, 이로 인해 나와 타인, 우리 집단과 타 집단 사이에 도덕적 가치와 행동의 차이가 발생했음을 인식해야 한다. 하이트는 자신의 저서 '바른 마음'의 마지막 장을 다음과 같이 서술하며 마무리한다(Hadit, 2012/2014: 560-561).

> *... 어떤 사람은 선하고, 어떤 사람은 악해서가 아니다. 그보다는 우리의 마음이 집단적 바름을 추구하도록 만들어졌기 때문이라는 것이 이 책의 답이었다. 우리 인간은 지극히 직관적인 생물체로서, 우리의 전략적 추론 능력도 사실은 직감에 따라 움직인다. 그러다 보니 나와 다른 매트릭스의 사람들을 만나면, 더구나 그런 이들의 도덕 매트릭스는 우리의 것과는 다른 식으로 배열된 도덕성 기반에 의지하고 있는 때가 많기 때문에, 그들과 연결된다는 것은 쉬운 일이 아니다(그러나 불가능한 일도 아니다). ... 그러니만큼 이다음에 옆자리에 나와 다른 매트릭스의 사람이 앉게 된다면, 그때는 한번 연결을 시도해보자. 하지만 그 사람의 매트릭스로 곧장 뛰어들려고 해서는 안 된다. 또 몇 가지 공통점이 발견되거나 어떤 식으로든 약간의 신뢰가 생기기 전까지는 도덕성의 문제를 꺼내 들어서도 안 된다. 그러다가 시의적절하게 도덕성과 관련된 이슈를 무사히 꺼냈다 싶으면, 그때는 다른 것보다 먼저 그 사람의 입장을 얼마간 추어주고 그에 대해 진정한 관심을 표한다. ... 우리는 어차피 한동안은 이 땅에 다 같이 발붙이고 살아가야 한다. 그러니 서로 잘 지낼 수 있게 함께 노력해보자.*

나아가 우리의 '도덕적 신념'을 소유물로 보는 것에서 벗어난 밈 이론의 통찰을 고려한다면, 협력적 논증 교환 활동은 신념에 기반한 동기화된 추론과 자신의 정체성을 효과적으로 분리시켜 적절한 수준의 방향성 동기를 탐색할 수 있는 계기를 효과적으로 마련해줄 수 있다. 앞서 살펴본 밈 이론의 분석에 따르면, 자신이 가지고 있던 기존의 도덕적 신념이 과연 '자신의 성찰'을 통해 획득된 것인지 비판적으로 검토해볼 필요가 있다. 특히, 동기화된 추론을 유발하는 자신의 기존 신념은 밈으로서 그 형태가 유지되어 왔을 가능성을 인식할 필요가 있다. 밈 이론의 핵심 통찰 중 하나는 밈이 강력하고 때로는 예상치 못한 방식으로 인간의 믿음

과 행동을 형성할 수 있다는 것이다. 밈은 단순히 문화적 정보의 수동적 운반자일 뿐만 아니라 인간의 행동과 사고를 형성하는 데 적극적인 참여자이기 때문이다. 예를 들어, 신념 체계나 실천은 비록 그것이 반드시 이성적이거나 유익하지는 않더라도 반복과 모방을 통해 문화 내에 널리 퍼질 수 있다. 밈 이론의 관점에서 도덕적 신념은 문화를 통해 확산되고 개인의 행동과 태도를 형성하는 밈의 일종으로 볼 수 있다. 다른 밈처럼, 도덕적 믿음은 사람들이 옳고 그름, 선과 악, 그리고 다른 도덕적 질문에 대해 생각하는 방식에 영향을 미치면서 인간의 행동과 생각에 강력한 영향을 미칠 수 있다. 따라서 '우리의 신념은 소유물이 아니다'라는 생각은 협력적 논증 교환에서 학생들이 공유해야 할 기본적인 원칙이다.

밈은 집단적 신념의 일종이다. 즉, 동기화된 추론을 촉발하는 아동의 기존 신념은 주로 개인의 반성과 성찰이 아니라 집단적 특성에서 비롯되는 경우가 많다(Stanovich, 2021/2022: 162). 특히, 아이들은 사회적 환경, 다른 사람에게서 배우는 방식, 특정 신념에 대한 타고난 성향에 대한 통제력이 제한적이다. 그들은 타고난 성향에 대해 의식적인 선택을 하지 않으며, 사회적 학습을 통해 배우는 것은 의도적인 생각의 결과가 아니다. 우리의 믿음이 대부분 반성적이거나 성찰적이지 않은 수단을 통해 획득된다는 이 생각은 정확한 정보 제공을 통해 정치적으로 잘못된 정보를 수정하는 것이 얼마나 어려운지를 보여주는 연구들과도 일치한다(Hibbing et al., 2014; Flynn et al., 2017; Stanovich, 2021/2022: 163에서 재인용). 따라서 동기화된 추론 개선을 목적으로 수행하는 협력적 논증 교환 활동 참가자들은 자신의 신념이 나의 소유물이 아님을 인식할 필요가 있다.

자신의 신념이 소유물이라는 인식에 따르면, 우리는 자신의 신념이 의식적으로 선택되었으며 이것은 곧 자기 자신의 이익에 부합한다고 믿게 된다. 반면에 앞서 살펴본 밈의 관점에 따르면, 우리의 신념은 의식적으로 선택되지 않았을 수도 있고 우리의 개인적인 이익에 도움이 되지 않을 수도 있다. 오히려 밈은 우리에게 유용성 여부와 관계없이 밈 자체의 복제를 위해서만 존재한다. 또한 우리는 의식적인 생각을 통해서든 타고

난 심리적 성향과 일치하는 무의식적 요인의 결과로 특정 신념을 어떻게 획득했는지조차 알지 못한다. 밈 이론은 신념을 가진 사람들의 심리적 특성보다는 신념 자체의 특성에 초점을 맞춘 것이다(Stanovich, 2021/2022: 171). 우리의 신념이 자신을 복제하려는 이기적인 밈 복합체라는 것을 인식하면 동기화된 추론의 기반이자, 자신이 가지고 있는 기존 신념에 대해 더 회의적인 관점을 취할 수 있다. 우리 마음 속에 있는 밈 복합체는 자신과 반대되는 생각을 거부하는 경향이 있는데, 이는 그 생각이 현재의 신념을 위협할 수 있기 때문이다(Stanovich, 2021/2022: 234). 따라서 밈은 우리로부터 벗어나지 않기 위해 반대되는 생각을 거부하고, 동기화된 추론에 입각하여 논리적이고, 정교하게 우리 자신으로부터 그 생각을 차단시킨다. 이 과정을 통해 지지 관계에 있는 밈은 밈플렉스를 형성하고, 충돌하는 다른 밈이 우리 마음의 공간을 차지하지 못하도록 한다. 흡수되고 기존의 밈플렉스를 강화하는 새로운 신념이나 정보는 쉽게 받아들여지는 것이다. 이런 식으로 우리가 축적한 모든 밈은 우리의 신념을 강화하고 이를 확신으로 바꾸게 되는 것이다.

따라서 '신념은 우리의 소유물이 아니다'의 원칙을 기반으로 협력적 논증 교환에 참여한다면 학생들은 저마다 적절한 수준의 방향성을 탐색할 수 있을 것이다. 이처럼 적절한 수준의 방향성은 고정되거나 교사에 의해 사전에 인위적으로 정해질 수 있는 것이 아니다. 신뢰할 수 있는 동료와 진정성 있는 공개적인 협력적 논증 교환의 과정을 통해 조정을 거치며 학생들은 각자 적절한 수준의 방향성에 의해 동기화 수준을 획득할 것으로 기대할 수 있다. 특히, 협력적 논증 교환 수업은 학생들로 하여금 자신의 도덕적 신념과 자신의 정체성이 분리될 수 있다는 사실을 확인시켜줌으로써 자신을 보호하고자 하는 방향으로 동기화된 추론을 효과적으로 개선하는 데 기여할 수 있다.

이처럼 협력적 논증 교환 활동은 근본적으로는 학생들로 하여금 자신이 가지고 있는 기존 신념과 자신의 정체성을 메타적으로 성찰하도록 하여 도덕적 신념을 수정하고 재설정할 수 있는 기회를 제공한다. 또한 신뢰할 수 있는 동료로부터 자신의 논증에 대한 객관적 평가를 기반으로

한 비판을 통해 편향의 사각지대으로부터 빠져나오는 데에도 효과적으로 작용할 것이라고 기대해볼 수 있다.

2. 인지적 작업의 대립적 분업으로서 수업 양상

협력적 논증 교환 활동은 추론의 사회적 기능을 고려한 동기화된 추론의 개선을 목표로 하는 수업에서 활용될 수 있는 교수-학습 방법이다. 일반적으로 협력적 논증 교환은 교실 수업 상황에서 토론 활동의 모습을 띤다. 교실에서의 토론은 학생의 참여, 비판적 사고 및 학습을 촉진하는 데 도움이 될 수 있는 중요한 교육학적 도구이다(Brooks & Brooks, 1993). 학습에 대한 구성주의적 관점에서 토론 수업의 의의를 강조한 브룩스 등은 토론 수업의 주요 목적을 학습자의 능동적 학습과 비판적 사고를 촉진하는 것으로 보았다(Brooks & Brooks, 1993: 12). 토론을 통해 학생들은 학습내용을 적극적으로 이해하고, 지식을 재구성하여 자신의 생각과 의견을 공유하며, 주제에 대한 공동의 탐구에 참여할 수 있기 때문이다. 이는 비판적 사고와 의사소통 능력의 개발뿐만 아니라 해당 학습 문제에 대한 더 깊은 이해로 이어질 수 있다. 또한 일반적으로 토론은 학생들이 다른 사람들의 관점뿐만 아니라 자신의 신념과 가치를 성찰하도록 장려함으로써 비판적 사고를 촉진한다(Brooks & Brooks, 1993: 42).

추론의 사회적 기능에 주목한 연구자들 또한 교육에서의 협력적 토론 수업의 중요성을 강조한다. 추론의 논증 이론을 개발한 메르시에와 스퍼버는 인간 추론의 주된 기능이 다른 사람들과 논증을 교환하는 것에 있기 때문에, 추론은 개인이 단독적으로 사용할 때보다, 대화의 맥락에서 사용될 때 더 나은 결과를 가져올 것이라고 예측한다(Mercier, 2016; Mercier & Sperber, 2011; 2017). 이러한 예측은 함께 토론에 참여하는 소규모 참가자 그룹이 문제 해결 과제에서 개별 참가자를 크게 능가하는 것으로 나타났다는 경험적 연구들에 의해 지지되었다(Laughlin, 2011; Moshman & Geil, 1998). 특히, 인지 편향 연구에서 널리 사용되는 왓슨

의 카드 선택 과제(1966)와 결합 오류를 입증하는 데 사용된 문제(Tversky & Kaneman, 1983), 인지적 성찰 테스트(Frederick, 2005)와 같은 판단 및 의사 결정에 사용된 많은 과제를 해결함에 있어 개인의 과제 수행에 비해 집단 토론을 통한 과제 수행이 효과적이었다. 그러한 과제에 대해 적어도 한 명의 구성원이 정답을 찾았을 경우, 그는 일관되게 모든 그룹 구성원이 그것을 받아들이도록 설득했으며 대부분의 경우 설득이 되었고 답을 찾았다(Trouche et al., 2014). 나아가, 집단 토론에 따라 생각이 바뀌는 참가자들은 토론 전보다 전이 문제에 더 잘 대처하는 경향이 있었다(Laughlin, 2011; Trouchie et al., 2014).

이와 같은 토론의 특징은 도덕적, 정치적 이슈와 같이 대립적이고 논쟁적인 주제를 다루는 수업에서 널리 활용된다. 특히, 도덕 수업에서 토론은 학생들에게 도덕적 추론을 촉진시키고, 윤리적 의사결정에 참여할 수 있는 기회를 제공한다(Nucci, 2001). 토론은 학생들에게 서로 다른 도덕적 견해가 존중되는 분위기 속에서 토론에 참여함으로써 다양한 도덕적 관점에 노출되어 도덕적 문제에 대한 이해를 넓히고 도덕적 의사결정에서 다양한 관점을 고려하는 능력을 개발하는 데 도움을 준다. 뿐만 아니라, 토론은 학생들에게 자신의 가치와 신념을 성찰하고 표현할 수 있는 기회를 제공한다. 학생들은 자신의 도덕적 견해를 표현하고 이에 대한 다른 사람들의 피드백을 받는 과정을 거치며 자신의 기존 도덕적 신념과 도덕적 정체성에 대해 성찰할 수 있다. 나아가 교실에서의 토론은 또한 학생들이 도덕적인 추론과 판단에 필수적인 비판적인 사고력을 발달시키는 데 도움을 줄 수 있다. 예를 들어, 학생들은 다른 사람들의 주장과 관점을 비판적으로 평가하고, 윤리적 딜레마를 식별하고 분석하고, 다른 행동 과정의 결과를 탐구하도록 요청받을 수 있다. 이러한 유형의 비판적 사고는 학생들이 정보에 입각한 윤리적 결정을 내리는 능력을 개발하는 것을 돕기 때문에 도덕 교육의 핵심 요소이다.

이와 같은 도덕 수업은 일반적인 토론 수업의 양상과 함께, 협력적 논증 교환만의 특징이 반영되어야 한다. 추론이 사용되는 사회적 맥락을 고려하였을 때, 협력적 논증 교환의 참여자는 크게 두 축, '논증을 생성

하는 자'와 '논증을 평가하는 자'의 역할을 수행해야 한다. 논증 생성에서 발생하는 동기화된 추론의 문제는 협력적 논증 교환의 과정에서 논증 평가를 통해 교정될 수 있다. 이와 같은 동기화된 추론 문제의 교정 과정은 인지적 작업의 대립적 분업(the adversarial division of cognitive labor)이라는 개념으로 이해될 수 있다. 이는 지식을 추구하고, 신념을 구축하는 과정에서 사람들 간의 의견 불일치와 반대되는 관점의 중요성을 강조하는 맥락에서 고안된 탐구의 방법이다[48](Callard, 2018). 이 방법은 소크라테스의 철학적 방법론의 인식론적 이점에 대해 연구한 캘러드가 그의 방법론을 인식론적 작업의 대립적 분업으로 개념화한 데에서 유래한다. 일상생활에서 대부분의 사람들은 문맥에 따라 진실의 축적을 우선시하는 것과 거짓을 피하는 것을 우선시하는 것 사이에서 행동에 필요한 신념이 무엇인지를 자연스럽게 전환한다. 그러나 캘러드는 그러한 상황판단이 궁극적으로 올바른 지식을 추구하는 작업에는 부적절하다고 지적한다. 지식의 습득은 항상 진리의 추구와 거짓의 회피를 모두 필요로 하기 때문에 맥락에 따라 우선순위를 결정하는 것은 적절하지 않기 때문이다. 소크라테스의 방법이 진정한 천재성을 보여주는 것은 이 문제와 관련이 있다. 개인들은 진실의 축적을 우선시하는 것과 거짓의 회피 사이에서 실용적인 선택을 할 수밖에 없다. 그리고 그들은 두 가지를 동시

48) 고대 그리스의 인식론 및 도덕철학을 주 연구 분야로 삼는 캘러드(Agnes Callard)는 자신의 향후 연구에 관한 여러 인터뷰에서 소크라테스식 탐구 방법을 '인식론적 작업의 대립적 분업(the adversarial division of epistemic labor)'으로 개념화하였다. 이 논의가 포함된 그녀의 저서 "The World Socrates Made"는 현재 출간 준비 중인데, 해당 연구를 소개하는 여러 인터뷰에서 이 개념을 소개하였다. 또한 책의 저술을 지원하는 구겐하임 펠로십에서는 해당 개념에 대해 "논증에 참여함으로써 기대되는 이상적인 효과와 현실적 어려움 사이에서 소크라테스식 방법은 우리가 원하고 배워야 하는 반박 문화를 만드는 데 기여할 것이다"라고 간접적으로 평가하였다. 한편, 인식론적 작업의 대립적 분업이라는 개념을 직접적으로 언급하진 않지만, 소크라테스식 논박술을 분석하고 그것의 인식론적 가치를 재조명한 그녀의 작업물로는 Aspiration: The agency of becoming(NY: Oxford University press, 2018)이 있다. 그리고 그녀가 제안한 이 개념은 소크라테스식 논박술이 '대립적 대화'의 가치와 기능을 재확인했다는 평가를 받으며, '철학적 작업의 대립적 분업', '신학적 작업의 대립적 분업' 등으로 파생되어 사용되기도 한다(R. Duerr, 2022). 협력적 논증 교환 또한 대립적 대화의 인식론적 가치에 근거하여 논증의 생성과 평가라는 인지적 작업의 분업을 토대로 한다는 점에서 해당 개념을 빌려 그 양상을 설명하고자 한다.

에 우선적으로 고려할 수 없기 때문에, 그들 스스로 지식을 충분히 추구할 수 없다. 그러나 소크라테스는 어떤 대가를 치르더라도 진리의 축적을 우선시 여기며 거짓을 피하고자 하는 두 사람이 협력한다면 두 가지 목표를 한꺼번에 우선시할 수 있다는 것을 깨달았다. 그것은 현대의 형사사법제도와 비슷하다고 캘러드는 지적한다. 일반적으로 정의를 추구하기 위해서는 한 사건 내에서 판사는 유죄 또는 무죄 판결 중 하나를 내려야 할 것이다. 두 가지 목표를 동시에 추구할 수는 없다. 만약 한 사람의 궁극적인 목표가 모든 유죄 판결을 내리는 것이라면, 모든 사람들은 유죄 판결을 받아야 한다. 만약 한 사람의 궁극적인 목표가 모든 무고한 사람들을 무죄로 만드는 것이라면, 아주 조금의 의심이라도 주어진 피고인을 무죄로 만들기에 충분해야 한다. 유죄를 선고하고 무죄를 선고하려는 욕망은 서로 다른 목적을 가지고 있으며, 따라서 현대의 사법 제도는 이 작업을 분열시킨다. 검사는 유죄 평결을 추구하고 피고는 무죄 평결을 추구한다. 그리고 협력적인 사법적 대화가 이루어짐에 따라, 두 가지 규칙(죄인을 유죄로 인정하고 무죄를 무죄로 인정하는 것)이 모두 정의를 추구할 수 있는 방식으로 지켜질 수 있다.

이처럼 인지적 작업의 대립적 분업은 지식을 탐구하고 문제 상황에서 보다 적절한 해결책을 찾기 위한 작업의 양상이 적대적일 수 있음을 시사하는데, 이는 개인이나 집단이 특정 문제에 대한 보다 정확한 이해에 도달하기 위해 서로에 대해 적극적으로 노력한다는 것을 의미한다. 동기화된 추론을 개선하고자 하는 우리의 인지적 노력이 개인의 내적 성찰과 같은 독립적인 방법으로 이루어지기보다는, 자신과 대립하는 의견을 지닌 상대방과의 사회적 상호작용의 방법으로 이루어질 때 더욱 효과적이기 때문이다.

인지적 작업의 대립적 분업은 추론의 과제를 두 개의 분리된 대립적 요소로 나누어 동기화된 추론의 문제를 개선하는 방법이다. 두 개의 분리된 요소 중 한 축은 논증의 생성이고 다른 한 축은 논증의 평가이다. 인식론적 작업의 대립적 분업의 개념은 윤리나 정치 등 지식의 특정 분야에서는 반대 의견을 가진 개인 간에 분업이 일어날 수 있다는 생각을

말한다. 캘러드가 제시한 이러한 분업은 정당화와 비판의 책임을 서로 다른 두 사람 사이에 나누는 방법으로 정의된다. 한쪽에서 가설을 내놓는 역할을, 다른 사람이 그 가설을 무너뜨리는 역할을 수행하는 협력적 논증 교환의 구체적인 모습이다. 협력적 논증 교환 활동에서 가설을 내놓는 역할을 맡은 사람은 주어진 도덕적 이슈에 대한 자신의 입장을 정당화해야한다. 이를 위해 적절한 논증을 구성해서 상대방에게 제시해야 한다. 가설을 무너뜨리는 역할을 수행해야 하는 사람은 제공된 논증을 평가하는 책임이 있다. 이러한 분업은 개인이 자신의 편견과 신념을 확인하는 자연스러운 경향에 대한 점검을 제공하고 제시된 주장이 건전한 증거에 기초한다는 것을 보장하기 위한 것이다.

동기화된 추론으로 인해 발생하는 문제를 개선하기 위해 다른 의견을 교환하는 것에 중요성을 강조한 레슬리는 캘러드의 견해에 동의하며, 구체적인 방법으로 제시된 소크라테스의 대화법을 분석한다(Leslie, 2021: 76-78). 그는 깊이 사고하는 사람을 떠올릴 때, 흔히 로댕의 작품인 '생각하는 사람'을 떠올릴 수 있지만, 추론의 본질적인 측면을 고려한다면 고대 아테네의 광장에서 청년들과 논박을 하는 소크라테스가 젊은이들에게 질문을 던지고 논박하는 모습을 상상하는 것이 더 적절하다고 말한다. 소크라테스는 '논증을 생성하는 사람'들에게 적절한 질문을 던짐으로써, 자신의 논증의 불완전함을 깨닫도록 만들기 때문이다.

논박술(elenchus)이라고도 부르는 소크라테스의 대화법은 개인의 신념과 가정에 도전하기 위해 일련의 질문을 던지는 것과 관련된 대화식 접근 방식이다. 소크라테스는 지식과 지혜를 추구하려면 자신의 믿음에 대한 끊임없는 자기 성찰과 질문이 필요하며, 이 과정이 세상과 자신에 대한 더 깊은 이해로 이어질 수 있다고 믿었다. 논박술은 상대에게 일련의 질문을 던지는 대화 과정이다. 이 대화의 주요 기능 중 하나는 개인의 신념과 의견의 모순과 불일치를 드러내는 것에 있다. 개인의 신념에 도전하는 질문을 함으로써 소크라테스는 이러한 신념이 충분히 뒷받침되거나 논리적으로 일관성이 없음을 입증하려고 했다. 이러한 질문은 비판적 사고를 촉진하고 개인이 새로운 정보와 통찰력에 비추어 자신의 신념을

수정하도록 장려하기 위한 것이다. 이처럼 개인 간 협력적 논증 교환의 한 형태로, 도덕 판단에 대한 기본 전제를 비판적으로 검토할 수 있도록 질문을 던지고 답하는 것에 기반하는 소크라테스의 대화법은 다음의 단계를 따른다.

0. 논증 생성자는 논증 평가자에게 논증(주장과 근거)을 요구한다.
1. 논증 생성자(아테네의 청년)가 하나의 논증을 제시한다.
2. 논증 평가자(소크라테스)는 논증의 진위 여부를 판단하고, 반박한다.
3. 논증 평가자는 논증 생성자에게 추가적인 전제에 대해 동의 여부를 확인한다.
4. 논증 평가자는 논증 생성자가 동의한 추가적 전제가 최초의 주장(혹은 근거)와 양립할 수 없음을 상기시킨다.
5. 논증 생성자는 자신의 주장(혹은 근거)를 점검하고, 수정된 논증을 다시 제시한다.
6. 논증 평가자는 2-4를 반복 수행한다.
7. 논증 생성자와 논증 평가자는 공동의 결론에 도달한다.

이러한 대화 과정의 양상은 대립적 분업을 통해 복잡한 문제에 대해 대화 참가자들 사이의 지식과 이해를 증진시키고, 문제를 해결하기 위한 공동의 노력을 포함한다. 소크라테스의 대화법의 인식론적 가치에 대해서는 해당 연구 분야 내에서 여전히 논쟁 중이지만(Salkever 2014), 우리는 이러한 대화 방식을 통해 동기화된 추론의 문제가 개선되는 구체적인 양상을 확인할 수 있다.

먼저 협력적 논증 교환의 참여자들은 특정 윤리적 이슈에 대한 자신의 입장을 정한다. 이 입장은 해당 이슈에 대한 자신의 기존 신념과 도덕적 감정에 기반하여 촉발된 판단이 포함된다. 참가자들은 자신의 입장을 타인에게 설득하기 위한 목적으로 논증의 형태로 구체화한 후, 자신과 입장이 다른 상대방에게 논증을 제시한다. 생성된 논증은 논리적으로나 윤리적으로 타당할 수도 있지만 그렇지 않을 수도 있다. 동기화된 추론의

견지에서 본다면, 논증의 생성 과정에서 추론은 처음부터 가장 좋은 주장을 구성하지 않기 때문이다(Mercier, 2011b: 182). 추론의 논증 이론에 따르면, 추론은 논쟁적인 맥락에서 우리가 이유를 찾고 평가하는 것을 돕기 위해 진화했다. 이러한 맥락은 참가자들이 그들의 주장을 만들기 위해 몇 가지 주장을 시도할 수 있도록 하는 특징을 지닌다. 한 번의 실패는 거의 비용이 들지 않는다. 두 번째 주장은 언제나 제기될 수 있다. 첫 번째 시도에서 확실한 주장을 찾는 것으로 인지적인 노력을 너무 많이 들일 필요는 없는 것이다. 따라서 추론은 매우 높은 만족도를 나타내야 한다. 그것은 최선의 주장을 추구하는 것이 아니라 충분히 좋은 주장을 추구해야 한다(Mercier, 2011b: 181-183). 이는 관찰자들이 종종 순진한 성인들의 주장 수준에 의해 낙담하는 이유 중 하나이다(Kuhn, 1991; Perkins, 1985). 사람들에게 더 나은 주장을 구성하도록 강요하는 가장 자연스러운 방법은 반론을 제공하는 것이며, 이는 집단에서 자발적으로 발생하고 부분적으로 그들의 더 나은 추론 성과를 설명하는 것이다. 이것이 협력적 논증 교환의 양상이다.

동기화된 추론의 개선은 학습자들이 논증을 상호 교환하며 평가하는 과정에서 기대할 수 있다. 추론의 사회적 기능을 고려했을 때, 상대방의 논증을 이해하고 평가할 때 우리는 좋은 주장과 나쁜 주장을 구분하는 능력을 십분 발휘하기 때문이다. 학생들은 논증을 완벽하게 이해하고, 거부할 수 있다(Mercier, 2011b: 181). 특히, 도덕적 행동과 추론의 영역은 서로 다른 주장이 교환될 때 학생들이 논증을 더 잘 분석하고, 평가하는 능력이 가장 잘 발휘된다(Nucci, 1984; Killen, 19991). 누찌와 킬렌이 각각 실시한 연구에서, 아동은 도덕의 영역과 사회인습의 영역과 관련된 주장의 적절성에 대해 구분하여 평가하는 능력이 있음을 확인했다. 미취학 아동부터 청소년에 이르는 참가자들은 도덕적 영역과 관습적 영역에서의 위반에 직면했을 때 교사들의 반응을 다르게 평가했다. 다른 연구들은 또한 장난감이나 사탕을 나누는 것과 같은 도덕적인 행동을 수행하라는 명령에 대한 반응으로 아이들의 행동을 분석했다. 결과적으로, 아이들은 어떤 주장도 무비판적이고 맹목적으로 받아들이지 않았다. 아이들

은 좋은 주장과 나쁜 주장을 구분할 수 있으며, 이는 협력적 논증 교환 수업에서 논증을 상호 교환하고 평가하는 과정에서 잘 발휘된다.

이처럼 협력적 논증 교환 수업은 기본적으로 인지적 작업의 대립적 분업의 양상으로 나타난다. 그 양상은 개인이 혼자서 성찰적 능력을 발휘하는 모습과 달리, 둘 이상의 동료가 함께 토론하고 논박하는 과정을 포함한다. 서로 다른 입장을 지닌 집단의 구성원들이 하나의 윤리적 이슈에 대해 자신의 의견을 논증의 형태로 생성하는 과정에서 동기화된 추론이 촉발되고, 대립적 의견의 교환을 통해 서로의 논증을 평가하는 과정은 이를 개선해주고 자신의 기존 신념을 수정할 기회를 제공해준다. 동기화된 추론의 개선을 목적으로 하는 인지적 작업은 대립적 의견을 가진 둘 이상의 학습자가 함께 수행하는 것인데, 이 모습은 소크라테스의 대화법의 양상과 닮아있다.

도덕 교육에서 소크라테스의 대화법은 이미 널리 활용되고 있다. 콜버그는 도덕교육의 한 방법으로서 소크라테스의 대화법이 지니는 함축적 의미를 바탕으로 도덕 발달을 위한 딜레마 토론 수업을 개발했다. 그는 학생들에게 도덕을 가르치기 위해서는 권위에 복종하도록 하는 방법이 아니라 도덕적 의견의 자유로운 교환 속에서 이루어져야 한다는 것이 소크라테스의 대화법이 도덕 교육에 갖는 함의라고 분석한다(Kolhberg, 1970/2000: 68-89) 이러한 관점을 바탕으로 도덕 교육에서 소크라테스의 대화법을 적용하는 첫 단계는 학생 스스로 자신의 선의 지식에 불만을 갖도록 부추기는 소크라테스적 질문으로 시작한다. 교사는 학생에게 학생이 가지고 있는 기존의 도덕적 원리로는 해결할 수 없는 도덕적 갈등 상황을 제시해주어야 한다. 두 번째 단계는 주어진 갈등 상황에 대해서 동료 학생들과 함께 의견을 주고 받고 논박할 기회를 제공하는 것이다. 이 과정을 통해 학생들은 이전에는 고려하지 않았던 도덕 원리를 확인하게 되고, 자신이 가지고 있던 기존의 신념을 반성적으로 성찰하게 되는 것이다. 이 과정에서 교사는 마치 소크라테스가 메논, 에우티프론, 크세노폰 등 아테네의 청년들에게 무지를 자각시키고 진리로 나아가는 방향을 안내한 것과 같은 역할을 수행하는 것이다. 아테네 청년들과 마찬가

지로 학생들이 자신의 의견을 논증의 형태로 생성해내면, 마치 소크라테스가 그랬던 것처럼 교사는 학생들에게 적절한 질문을 던짐으로써 논증을 평가하고, 그들의 최초 논증을 수정하도록 요구한다. 이처럼 소크라테스의 대화법과 그것을 기반으로 한 콜버그식 도덕 딜레마 토론은 주로 소크라테스/교사는 논증 평가자의 역할을, 아네테 청년/학생들이 논증 생성자의 역할을 부여 받는다.

그러나 협력적 논증 교환 수업은 질문을 던지는 소크라테스와 그에 대한 적절한 답을 제공해야하는 아테네 청년들의 역할처럼 고정되지 않는다. 학생들은 논증의 생성자이자, 동시에 논증의 평가자 역할을 수행하고 이 두 역할은 끊임없이 순환하며 상호작용한다. 논증을 생성한 학습자는 타인의 논증을 평가하기 위해 다른 학생이 제시한 논증에 대해 반론을 제기하고 질문을 던져야 하며, 타인이 제공한 논증을 이해하고 평가하는 역할을 부여받은 사람 또한 자신만의 논증을 생성하고 제시해야 하기 때문이다. 이 지점이 협력적 논증 교환의 활동과 콜버그 딜레마 토론의 차이점이다. 협력적 논증 교환 활동을 포함한 수업에서 교사는 학생들에게 윤리적 이슈를 제기하며 그와 관련한 도덕적 감정을 불러일으키고, 그 감정에 의해 동기화된 추론을 바탕으로 판단을 내려보는 기회도 제공해야 한다. 또한 대립적 의견이 존재하지 않거나 더 나은 반론을 계발하지 못하는 경우 종종 교사는 학생들이 논증 교환을 이어가도록 적절한 발문을 던져야 할 것이다. 이처럼 교사는 의견이 다른 학생들 사이에 논증이 원활히 교환될 수 있도록 중재의 역할을 수행할 수 있다. 그러나 본질적으로 협력적 논증 교환, 즉 인지적 작업의 대립적 분업의 주체는 학생 개개인인 것이다. 이것이 협력적 논증 교환 활동을 통해 나타날 수 있는 수업의 양상이다.

3. 교사의 역할 및 실행 전략

추론의 기능이 사회적 상호작용에 있음을 고려했을 때, 협력적 논증 교환 수업은 동기화된 추론의 개선에 효과적인 방법으로 제시될 수 있

다. 그러나 때로는 자신과 입장이 다른 사람과 함께 수행하는 토론의 결과가 오히려 자신의 입장만을 공고히하고, 도덕적, 정치적 태도의 양극화를 심화시키는 방향으로 결론지어기도 한다(Mercier, 2016: 693). 추론의 논증 이론을 통해 확인된 추론의 사회적 기능은 개인이 자신과 입장이 다른 사람과의 논증 교환을 통해 자신의 신념을 형성하고, 방어하는 과정을 거쳐 적응한 결과이다. 이에 따르면 추론은 본질적으로 개인 내적이기 보다는 사회적 상호작용에서의 논증의 한 형태이다. 즉, 다른 사람들이 자신의 신념을 채택하도록 설득하기 위해 논증을 생성하고, 타인의 논증을 평가하는 과정에서 추론을 사용하는 것이다. 또한 발달 심리학적 관점에서 추론 능력의 발달은 개인의 인지발달과 공변하지만, 또래와의 상호작용 속에서 발달이 촉진된다. 추론의 사회적 기능에 주목한 연구들은 공통적으로 추론이 잘 발휘되기 위한 환경을 강조한다. 이처럼 도덕 수업에서 협력적 논증 교환이 동기화된 추론을 개선하기 위해 효과적으로 작동하기 위해서는 함께 논증을 교환하는 구성원들이 누구이며, 어떻게 집단이 형성되는지가 중요한 요소로 작용할 수 있다. 추론의 사회적 기능의 관점에서 본다면, 논증 교환이 동기화된 추론 문제를 개선하지 못하는 이유는 추론의 사회적 기능이 발휘되기 위한 적절한 맥락이 고려되지 않았기 때문이다. 즉, 협력적 논증 교환이 효과적으로 이루어지기 위해서는 ①구성원 간의 신뢰를 바탕으로, ②적정 인원으로 구성된 소규모 집단에서 ③이질적 의견교환이 이루어져야 한다. 교사는 이러한 특징을 고려하여 수업을 설계하고 운영해야할 것이다.

먼저, 구성원 간의 신뢰이다. 협력적 논증 교환 수업은 추론이 본질적 사회적 활동이라는 점에 주목한다. 다른 사람을 효과적으로 설득하기 위해서는 상대방의 반론과 반대 의견을 잘 파악하고, 대응할 수 있어야 한다. 또한 타인의 논증이 타당한지를 객관적으로 판단하여 그 의견의 수용 여부를 결정해야 한다. 이를 위해서는 협력적 논증 교환 활동에 참여하는 구성원들 사이에는 기본적인 신뢰가 전제되어야 한다. 방향성에 동기화된 추론에서 벗어나 정확성에 의해 동기화된 추론으로 전환하기 위해 반대 의견의 경청을 강조한 갈렙은 단순히 반대의견을 가진 사람과의

토론이 동기화된 추론을 개선해주지 못한다는 점을 지적한다(Galef, 2021). 의견이 불일치한 상황에서 반대 의견에 노출되는 것이 동기화된 추론을 개선하지 못하는 가장 큰 이유는 이를 너무 간단한 문제로 여기기 때문이다. 흔히 이성적이고, 진실하게 논증 교환에 참여할 때, 반대 의견을 통해 자신의 신념이 수정되거나 문제의 해결책에 도달하는 것이 간단한 과정일 것이라고 예상한다. 각자의 주장을 제시하고, 한 사람이 논리적이고 증거에 기반한 주장을 제시하면 상대방은 그 타당성을 인정하고 시각을 바꿀 것이라는 낙관적인 기대를 하는 것이다. 그러나 오히려 더 많은 경우에 반대 입장을 지닌 사람이 설득력 있는 주장을 제시했음에도 불구하고 기존 신념이 공고히 유지하기도 한다. 심지어 양측 모두 사실적 정보에 대해 공유하고 있고, 논리적으로 논증을 구성할 능력을 지니며, 진실되게 자신의 입장을 설명하고, 타인의 입장을 이해하는 이상적인 조건에서도 반대 의견을 통해 동기화된 추론이 개선되는 것은 쉽지 않다(Galef, 2021: 240). 많은 연구들은 토론의 결과로 인해 의견의 불일치와 태도의 양극화가 완화되기보다는 오히려 각자 기존의 관점을 공고히한다고 주장한다. 가령, 한 실험 연구에서 사람들에게 자신의 기존 신념과 반대되는 전문가 집단의 견해를 계속해서 제공해주었을 때 사람들은 오히려 자신의 기존 신념을 더욱 옹호하는 모습을 보였다(C. A. Bail et al., 2018). 실험에서는 미국의 진보주의적 신념을 옹호하는 민주당원과 보수주의적 신념을 옹호하는 공화당원을 대상으로 다양한 사회정책 문제에 대해 종합적인 설문을 마친 후 임의로 통제 집단과 실험 집단으로 배정하였다. 이후, SNS를 활용해 민주당원들에게는 보수주의적 관점에 부합하는 전문가 견해를, 공화당원들에게는 진보주의적 관점에 부합하는 전문가 견해를 계속해서 제공하였다. 연구 결과, 진보주의적 견해에 노출된 보수당원들은 실험 후에 이전에 비해 보수적인 견해를 더 강하게 지지하였다(민주당원들 또한 더 진보적인 양상을 보였지만, 통계적으로는 유의미하지 않았다). 이 실험은 SNS의 반향실(Echo-chamber) 효과가 동기화된 추론을 촉진하여 정치적 양극화를 심화시킴을 시사한다. 즉, 실험 결과는 반대 의견에 대한 노출이 동기화된 추론을 개선하기

보다는 오히려 더 강하게 방향성에 의해 동기화된 추론을 부추기는 결과를 초래했다고 해석된다. 물론 이처럼 집단 내에서 다른 입장을 가진 사람과의 논증 교환이 효과가 엇갈리는 이유에 대해 실험 설계의 문제를 지적할 수 있다. 반대 입장에 대한 노출이 동기화된 추론을 개선하지 못하는 것처럼 보이는 이유는 대부분의 집단 실험에서, 참가자들은 서로를 알지 못하는 경우가 많았고, 실험 이후 다시는 만날 수 없을 사람들과 상호작용하며, 실험실 밖에서 그들의 삶과 전혀 관련이 없는 작업을 수행하도록 요청받았기 때문이다(Mercier & Speber, 2011: 73).

실험 설계의 신빙성과 별개로, 갈렙은 이와 같이 협력적 논증 교환 효과에 대한 부정적 실험 결과가 반대 의견이 무의미하다는 것이 아니라, 동기화된 추론을 개선하기 위해서는 반대 의견을 듣고 배우기 위한 방식에 변화를 줄 필요가 있다는 사실에 있음을 강조하였다. 즉, 본 실험에서 반대 의견이 효과적이지 못했던 까닭은 상대방을 잘못 선택했기 때문이다. 기본적으로 우리가 반대 의견을 접하는 가장 일반적인 통로는 우리를 진심으로 신뢰하여 설득하고자 하는 사람이 아니라 적대적인 사람인 경우가 대부분이다. 일상에서 반대 의견을 접하는 경로는 상대방의 견해에 무례하게 반응하거나, 자신의 입장을 지지하는 편에게는 격려를, 상대편에게는 비난과 조롱을 하는 경우가 많다(Galef, 2021: 232). 하이트에 따르면 이 경우에는 강한 거부감으로 인해 다른 사람의 도덕 매트릭스를 이해하고자 시도조차 하지 못한다. 이것이 동기화된 추론 개선을 위한 목적으로 이루어진 협력적 논증 교환에서 집단 구성원들 사이에 신뢰가 전제되어야 하는 까닭이다.

갈렙은 의견 불일치 속에서 동기화된 추론이 개선되기 위해서는 마음을 '더 쉽게 열리도록' 만드는 사람의 말에 경청해야 하며, 이를 위해서는 비록 자신과 반대되는 의견이지만 자신이 좋아하거나 존중하는 사람들, 사안에 따라 견해는 달라도 어느 정도 공감대를 형성하는 사람들, 진실되게 토론에 참여하고 미묘한 의미 차이와 불확실성의 영역을 인정하는 합리적인 사람들(Galef, 2021: 232-233)과 집단을 이뤄 논증을 교환해야함을 강조한다.

수업 상황에서 효과적인 협력적 논증 교환이 이루어지기 위해 교사는 교실의 분위기를 지지적이고 배려적으로 조성해야 한다. 특히, 집단을 구성함에 있어 교우 관계를 고려하여 서로 신뢰 관계가 형성되어 있는 학생들을 고려해야 한다. 이때, 교우 관계에서의 '친밀도'를 신뢰로 치환해서는 안 된다. 친밀도와 신뢰는 양의 상관관계를 보이는 것이 일반적이지만, 논증 교환의 상대방으로서의 신뢰도는 친밀도와 구분되는 지표이기 때문이다. 따라서 단지 평소 친한 친구들과 함께 토론하는 것이 아니라, 서로의 발언을 비중있게 다루고 자신과 입장이 다를지라도 경청할 수 있는 최소한의 신뢰관계가 형성되어 있는 학생들로 집단을 구성해야 할 것이다. 따라서 궁극적으로는 교실 환경을 서로가 서로를 신뢰하고 지지하는 분위기를 조성해야 한다.

다음은 이질적인 구성원들로 집단을 구성해야 한다. 동기화된 추론을 개선하기 위해 협력적 논증 교환 수업은 집단 구성에 있어 숙의 민주주의의 실증적 연구 결과들을 참고할 수 있다. 추론의 사회적 기능은 시민 심의 모델에서 실천적으로 고려되어왔다. 시민 심의가 효과적이기 위해서는, 즉 협력적인 논증 교환을 통해 입장의 변화가 발생하기 위해서는 반드시 '이질적인' 집단 구성이 되어야 한다(Fishkin & Luskin, 1999; Page, 2007). 이때 '이질적'이라는 의미는 다음의 내용을 포함한다.

집단의 구성은 도덕 판단에서 동기화된 추론의 문제에 상당한 영향을 미칠 수 있다. 동기화된 추론은 증거를 객관적으로 평가하기보다는 기존의 신념과 태도에 부합하는 방식으로 정보를 해석하고 처리하는 양상을 보인다. 이는 특히 개인이 비슷한 배경과 관점을 공유하는 동질적 집단에서 두드러질 수 있다. 연구에 따르면 다양한 집단은 균일한 집단에 비해 주장에 대한 더 정확하고 균형 잡힌 평가를 하는 경향이 있다(Page, 2007: 4-6). 또한, 인구통계학적으로도 다양성 측면에서 성별, 인종, 학력 등은 집단 다양성에 기여할 수 있는 요인 중 하나이다. 그러나 중요한 것은 인구통계학적 다양성뿐만 아니라 전문성과 지식의 측면에서도 다양성이다. 전문 분야가 다른 개인으로 구성된 그룹은 유사한 배경을 가진 개인으로 구성된 그룹보다 더 나은 결과를 얻을 수 있다(Page, 2007:

368-369). 한편, 다양한 집단은 다른 배경, 경험, 관점을 가진 개인들로 구성되어 있으며, 이는 더 넓은 범위의 주장을 고려하고 평가하게 하기 때문이다. 다양한 관점, 경험, 배경을 가진 개인들로 구성된 이질적 집단에서의 협력적 논증 교환은 개인을 다양한 범위의 의견에 노출 시켜주며, 자신의 편향과 선입견에만 의존하게 될 때 상대방을 설득하기 어려워지며, 따라서 자연스럽게 대안적 관점을 고려하고 건설적인 대화에 참여할 가능성이 더 높아진다.

또한 도덕적 상황이나 이슈에 대해 이질적인 견해를 갖는 참가자들로 협력적 논증 교환의 집단을 구성해야 한다. 도덕 판단 연구에서도 집단이 더 정확한 판단을 향해 나아가는 경향이 있는 현상은 협력적 논증 교환의 목적이자, 방향성 목표에 의해 동기화된 추론을 개선하고자 하는 목표인 정확성 동기로의 향상으로 해석될 수 있다(Blatt & Kohlberg, 1975; Mercier, 2011b). 그러나 이러한 개선이 이루어지기 위해서는 집단 구성원이 처음에 동의하지 않으며, 서로 다른 도덕적 신념에 기반한 이질적인 입장에서 출발해야 한다는 점이 강조된다(Mercier, 2011b). 최초에 구성원들은 도덕적 상황이나 이슈에 대해 서로 동의하지 않지만, 해당 문제를 해결하고 공동의 도덕성에 도달하고자 하는 목표를 공유하기 때문에 집단 내의 협력적 논증 교환은 동기화된 추론을 개선하는 데 기여할 것이다. 판단에 동의하는 집단 구성원이 토론하면 자신의 판단을 뒷받침하는 주장만 찾을 가능성이 높으며 아무도 이러한 주장에 대해 비판적이지 않을 것이기 때문에 방향성에 의한 동기화된 추론이 강화될 것이다. 결과적으로 집단은 양극화되고 더 극단적인 판단을 발전시킬 가능성이 있다(Sunstein, 2002). 그러나 사람들은 어떤 문제에 대해 의견이 일치하지 않을 때 논증을 교환할 가능성이 더 높다. 이때, 논증의 평가가 가장 좋은 결과를 낳기 때문이다. 논증이나 반론의 앞뒤에서 약한 논증은 거부되고 좋은 논증은 결국 사람들의 기존 견해를 바꿀 가능성을 높인다(Mercier, 2016: 695). 가령, 레스닉 등(Resnic et al., 1993)의 연구는 추론의 사회적 기능이 발휘되는 잘 구성된 논증적 환경에서 사람들의 동기화된 추론이 효과적으로 개선될 수 있음을 시사한다. 이들은 주어진

문제에 대해 동의하지 않는 세 명의 참가자로 구성된 집단을 만들었다. 참가자들은 합의에 도달하는 것을 목표로 약 20분 간 도덕적 이슈에 대해 논증을 교환하는 토론을 수행하도록 요청 받았다. 연구진은 이들의 논증 교환 과정을 모두 비디오로 촬영하여 논증의 사회적 분배의 양상을 분석하였다. 논증 교환 시작 후 첫 2분은 자신과 다른 입장에 주목하기보다는 자신의 기존 입장을 고수하는 데 집중했다. 그러나 소수의 집단으로 구성되어 있었기 때문에 참가자들은 서로의 주장에 더 민감하게 반응하였다. 시간이 흐를수록 이들은 반론 및 재반론 요구에 대해 서로를 설득하기 위한 더 객관적인 근거를 바탕으로 논증을 구성하였다. 나아가 연구진들은 이들이 사용한 문장의 구조를 정량적으로 분석하였다. 그 결과 두드러진 특징은, 양보의 언어(합의에 도달하는 과정에서 자신의 기존 입장에 대한 변화)가 도전의 언어(설득을 통해 타인의 입장에 대한 변화 요구)에 비해 두 배 가까이 더 많은 어휘들과 연결되었다. 양보의 언어 전후에는 상대방을 설득하기 위해 공격하는 논증과 자신의 입장을 방어하기 위해 사용한 논증이 구축되었기 때문이다. 협력적 논증 교환의 참가자들은 양보적 진술을 한 후에, 자신의 기존 신념을 약화시키거나 이전의 주장을 수정하는 모습을 보였다. 이러한 패턴은 이후 후속 연구들에서도 유사하게 나타났다(Blum-Kulka et al. 2002). 이처럼 추론의 사회적 기능이 적응한 진화적 맥락을 고려했을 때, 협력적 논증 교환이 효과적으로 동기화된 추론을 개선하기 위해서는 집단 구성원들이 도덕 문제를 해결하고자 하는 공동의 목표를 공유하지만 동시에 이질적인 입장에서 시작해야 한다.

수업 상황에서 집단 내 이질적 입장이 자연스럽게 형성되지 않을 수 있다. 이 경우, 악마의 옹호자 역할을 고려해볼 수 있다. 악마의 옹호자의 개념은 가톨릭 교회의 전통에 뿌리를 두고 있는데, 악마의 옹호자는 성인의 시복(諡福)이나 시성(諡聖)[49]을 선언하기에 앞서 검증 절차상 그

49) 시복(諡福)이란 로마 가톨릭 교회에서 그 성스러움(聖性, Holiness)이나 순교로 인해 이름 높은 자에게 순교한 사람에게 교황이 '복자(福者)'라는 칭호를 주는 선언을 의미하며, 시성(諡聖)이란 교황이 이미 시복이 된자를 성인의 명우에 올리고 전 세계 교회로 하여금 그를 공경할 수 있도록 선언하는 것을 의미한다. (가톨릭사전, url: https://maria.catholic.or.kr/dictionary/term/term_search.asp 검색일

인물을 반대하기 위한 목적으로 임명된 사람을 의미한다. 가톨릭 교회는 교황이 새로운 성인을 추대하는 작업에 있어 가장 정교한 검증 작업 절차를 고안하고자 했고, 1587년, 바티칸은 악마의 옹호자라는 장치를 마련했다[50].

악마의 옹호자로 지정된 사람은 교회의 결정이 모든 관련 정보의 철저한 평가에 기초할 수 있도록 제안된 대상 인물이 그럴만한 자격이 없음을 다양한 방식으로 주장함으로써 반대 의견을 제시하고 잠재적인 문제를 제기하는 역할이다. 가톨릭 교회는 이들을 추천하는 신의 옹호자(god's advocate)에 반대하는 악마의 옹호자를 통해 집단의 결정에 반대하는 역할을 명시적으로 제공함으로써, 집단 사고로 인한 동기화된 추론에 빠지는 것을 예방하고, 반증을 제시함으로써 더 나은 해결방안에 도달하고자 한다. 이는 편향의 사각지대를 연구한 행동 윤리학자들(Bazerman & Tenbrunsel, 2011/2014)에 의해서도 유사한 맥락으로 제안되었다[51].

악마의 옹호자는 협력적 논증 교환 수업에서 이질 의견이 존재하지 않을 때 의도적인 대립 의견을 제시하는 장치로 활용될 수 있다. 협력적 논증 교환 수업에서 교사는 참가자 일부에게 악마의 옹호자 역할을 부여하여 대립적인 의견이 교환될 수 있도록 수업을 운영할 수 있다. 참가자들은 악마의 옹호자가 제시한 반대 논증을 평가함으로써 자신의 신념을 검토해보고, 대안을 고려하고, 더 나은 결론에 도달할 수 있게 된다. 악마의 옹호자의 개입이 동기화된 추론으로 인한 문제를 개선해줄 가능성

자, 2023. 1. 14)

50) R. Burtsel, "Adovatud Diabloi" in <The Cathiolic Encyclopedia> Vol 1. 1907 (url: www.newadvent.org/cathen/01168b.htm. 검색일자, 2023. 1. 14.)

51) 악마의 옹호자 개념은 도덕 심리학 연구에서도 종종 사용된다. 대표적으로 하이트와 그 연구진이 수행한 연구에서 활용되었다(Haidt et al., 2000; Haidt, 2001). 연구자들은 실험 참가자들에게 일련의 도덕적 이야기를 제시하고, 그에 대한 즉각적인 도덕 판단을 요청한다. 이후, 그 판단에 대한 근거를 묻는 질문을 하는데, 이 때 해당 근거에 대해 반대되는 '이야기 속 단서'를 제시함으로써 악마의 옹호자로서의 질문을 하게 된다. 해당 연구에서 악마의 옹호자의 역할은 도덕 판단을 내린 실험 참가자들이 자신의 주장을 '논리적'이고 '합리적'으로 뒷받침하는 근거가 없음에도 여전히 자신의 도덕적 신념과 판단을 유지하는 모습을 확인하는 것에 있었다.

은 여러 연구를 통해 입증되었다(Lord et al., 1984). 대표적으로 한 연구(Lord et al., 1984)는 동기화된 추론으로 인한 도덕 판단의 편향이 자신의 신념이나 인식에 역행하는 대안을 무시하는 것에서 비롯된다는 가정을 바탕으로, 동기화된 추론의 문제를 완화하기 위해 반대되는 관점을 고려하도록 장려되어야 한다는 가설을 세웠다. 이들은 가설을 검증하기 위해 150명의 학부생을 대상으로 두 가지 실험을 실시하여 윤리적 이슈에 대한 새로운 증거의 편향된 동화(assimilation)와 성격 인상[52]에 대한 편향된 가설 검증라는 두 가지 영역에서 사회적 판단의 편향을 탐색하였다. 연구자들은 실험 참가자들에게 대안 가능성을 강조하는 명시적인 지침과 자료를 통해 반대되는 관점을 고려하도록 요청했다. 결과는 "반대를 고려하는" 전략의 유도가 공정성을 위해 노력하라는 지시보다 편견을 줄이는 데 더 효과적임을 나타냈다. 악마의 옹호자로서 연구자의 개입이 동기화된 추론을 약화시킨 것이다.

이와 같은 악마의 옹호자 효과는 학교 교육에서 비판적 사고 교육을 위한 장치로 활용되기도 한다(Davis, 2013; Gose, 2009). 수업에서 악마의 옹호자 역할은 일반적으로 문제에 대한 대안적인 관점과 해결책을 제공하는 개인에 의해 수행되며, 종종 집단적으로 전제하고 있는 것에 도전하는 역할이다. 악마의 옹호자를 수업에 적용하는 것은 학생들이 논리와 가정을 분석하고, 주장의 타당성과 건전성을 비판하고, 정확한 결론에 도달하도록 돕는다(Gose, 2009: 48). 악마의 옹호자로서의 교사는 학생들이 자신의 도덕적 신념과 그것의 기초가 되는 가정에 반대하는 다양한 논증을 제시하는 것을 목표로 한다. 특히, 학생들의 비판적 사고를 발달시키기 위한 학습을 위해 적절한 발문을 던져야 하는 교사의 역할은 마치 악마의 옹호자가 의도적으로 반대 의견을 제시하는 것과 유사하다

52) 성격 인상은 사람들이 다른 사람들의 성격 특성과 특성에 대해 형성하는 믿음, 태도, 판단을 말한다. 이러한 인상은 종종 외모, 행동 또는 누군가의 말하는 방식과 같은 제한된 정보에 기초하여 형성되며, 다양한 편견과 고정관념에 의해 영향을 받을 수 있다. 성격 인상은 사람들이 다른 사람들과 상호작용하고 인식하는 방식에 상당한 영향을 미칠 수 있으며, 사회적 판단의 요소가 될 수 있다. 그러나, 성격 인상이 항상 정확한 것은 아니며 편향과 판단의 오류에 노출될 수 있다는 것을 인식하는 것이 중요하다(Asch, 1961; Greenwald, 2002; Nauts et al., 2014).

(Davis, 2013: 3). 교사는 학생의 논증에 대해 반대 의견을 제시하거나, 의도적으로 비윤리적인 입장을 제시함으로써 학생들로 하여금 자신의 신념을 검토하고, 자신의 행동의 결과를 고려하도록 이끈다. 이 과정은 학생들이 복잡한 도덕 문제에 대해 보다 미묘한 이해를 발전시키고, 자신의 신념을 비판적으로 성찰함으로써 동기화된 추론으로 인한 문제를 확인시켜준다. 대표적으로 콜버그의 도덕 딜레마 토론에서 교사의 역할이다. 콜버그의 도덕교육론에 따르면 도덕 딜레마는 개인의 도덕적 신념과 가치에 도전하는 상황이며, 도덕적 성장과 발전의 계기가 될 수 있다 (Reimer et al., 1990). 도덕적 딜레마에 대한 논의에 참여함으로써, 개인은 다른 사람들의 관점을 고려할 뿐만 아니라, 자신의 신념과 가치를 성찰할 수 있다. 악마의 옹호자 접근법은 개인이 대안적 관점을 고려하고 윤리적 문제의 복잡성에 대해 깊이 생각하도록 장려하기 때문에 이러한 논의에서 특히 유용할 수 있다.

그러나 논증 교환에서 악마의 옹호자 역할에 대한 실질적인 효과가 미미하다는 연구들도 있다. 즉, 고의적으로 '반대 의견'을 제시하는 것이 주장의 신뢰성을 약화시키기 때문에 논증 참가자들이 반대 논증 평가에 대한 참여의 동기를 감소시킨다는 것이다. 대표적으로 네메스는 여러 연구를 통해 악마의 옹호자의 개입 여부와 동기화된 추론의 관계를 조사했다 (Nemeth et al., 2001a; 2001b). 네메스 연구진은 토론 집단 내 진정으로 반대 의견이 있는 경우와 '악마의 옹호자'가 있는 경우를 비교하였다. 네 명이 하나의 논증 교환 집단으로 구성하였고, 반대 의견이 없는 경우와 진정성이 있는 반대 의견이 있는 경우, 악마의 옹호자의 반대 의견이 있는 경우로 나누어 연구를 설계했다. 네 명의 집단에서 3명은 연구진들과 사전에 협의가 된 연구 협력자였고, 실제 실험 참가자는 한 명이었다. 실험 참가자들은 타인에게 피해를 주는 윤리적 문제를 제시받았고, 해당 이야기 속 피해자들은 이미 의료 비용과 휴업 손해금을 지급 받았다. 연구자들을 실험 참가자들에게 피해자의 신체적, 정신적 피해에 대한 보상금에 대해 정하고 토론하도록 요청했다. 보상금 선택은 1달러부터 52만 5000달러까지 총 8개의 선택지 중 하나로 정하도록 하였다. 거의 대부분

의 참가자들이 낮은 보상금을 선택하였다. 이때 첫 번째 실험 집단에서는 실험 참가자를 포함하여 세 명의 연구 협력자 모두 같은 의견을 제시하였고, 두 번째와 세 번째 실험 집단에서는 연구 협력자 중 한 명이 반대 의견을 제시하였다. 두 번째 실험 집단에서의 반대 의견은 실제 자신의 의견이었고, 세 번째 실험 집단에서의 반대 의견은 악마의 옹호자 역할이었다. 논증 교환을 마친 후, 참가자들은 사건에 대한 자신의 생각을 시간 순으로 보고하도록 했다. 이 방법은 개인의 사고가 변화하는 양상을 질적으로 평가하기에 적합한 연구 방법이다(Cacioppo & Petty, 1981). 진정성 있는 반대 의견(연구 협력자의 실제 의견)을 제공받은 참가자들은 집단 내 다른 사람의 논증을 따라하기 보다는, 자신의 의견을 스스로 수정하거나 성찰하는 모습을 드러냈다. 나아가 이들은 자신의 초기 입장에 관계없이 상해 사고에 대한 찬반의 입장을 모두 고려하는 모습을 보였다. 반면 악마의 옹호자와 함께 토론에 참여한 참가자들의 논증 교환은 다른 양상이었다. 이들은 자신의 관점으로만 상황을 해석하고 있었으며 오히려 동기화된 추론의 경향을 더 많이 보였다. 악마의 옹호자의 관점에 반박하려는 입장을 더 많이 수집하였고, 자신의 입장을 마지막까지 고수하는 경향을 보였다. 네메스는 해당 연구에 대해 다음과 같이 평가했다.

> *악마의 변호인의 본래 취지는 사람들이 자신이 선호하는 의견의 긍정적인 측면 외에도 부정적인 측면도 고려하도록 만들자는 것이다. 그러나 오히려 그 반대 결과를 이끄는 것으로 나타났다. 악마의 변호인과 같은 집단이었던 참가자들은 자신들의 의견이 옳다는 것을 다른 사람에게 설득하려는 모습을 보였다. 이와 대조적으로, 진정성 있는 반대 의견은 한 가지 입장에 대한 찬반 모두를 균형적으로 바라보도록 유도했다(Nemeth, 2018/2020: 237).*

악마의 옹호자 투입으로 인해 반대 의견을 경험한 참가자들이 오히려 자신의 입장을 고수하려는 동기화된 추론의 양상에 대해 네메스 연구진은 '인지적 강화(cognitive bolstering)'로 개념화했다. 악마의 옹호자에 의해 반대 의견에 노출됨으로써 자신의 기존 신념과 태도가 더욱 견고해

졌기 때문이다. 후속 연구는 악마의 옹호자의 효과를 더 구체적으로 확인하기 위해 실험을 다양하게 설계했다. 이전 연구와 유사한 상황에서 악마의 옹호자 관련 설정값에 변화를 주었다. 대조군에는 악마의 옹호자 없이 진정성 있는 반대 의견이 포함된 집단으로 구성하였고, 실험 집단은 세 개로 구분하였다. 첫 번째 집단은 악마의 옹호자 역할을 맡은 사람의 실제 의견을 구성원 모두가 모르는 상황이다. 두 번째 집단은 악마의 옹호자 역할을 맡은 사람의 실제 의견이 자신의 의견과는 다르지만 다수의 의견과 반대하도록 역할을 부여한 상황이다. 마지막 세 번째 집단은 악마의 옹호자에게 주어진 의견이 그 역할을 맡은 사람의 실제 의견과 같은 경우이다. 세 집단 모두 악마의 옹호자 역할을 맡은 참가자에게 주어진 대본에 따라 동일한 반대 의견을 제시하도록 요청했다. 연구를 계획함에 있어 악마의 옹호자의 양상에 따라 다른 결과가 나올 것이라 예측했으나, 실험 결과 악마의 옹호자 유형은 토론의 결과에 유의미한 차이를 만들어내지 못했다. 유일하게 논증에 필요한 근거를 양적으로 더 많이 산출한 것은, '진정성 있는' 반대 의견이 있는 경우였다(Nemeth, 2018/2020: 271). 특히, 자신의 실제 의견과 악마의 옹호자 입장이 같은 경우에도 논증 교환은 양적으로나 질적으로 큰 차이가 없었다. 이처럼 상대방의 반대 의견이 단지 자신의 의견을 반대하기 위해 인위적으로 제시된 것이란 걸 인지했을 때, 실험 참가자들은 동기화된 추론의 모습을 보였다. 따라서 협력적 논증 교환에 있어 가장 중요한 전제는 반대 의견을 진정성 있게 표현하는 것에 있음을 알 수 있다.

악마의 옹호자 역할은 대립적인 입장을 의도적으로 취하기 때문에 진정성 있는 반대 의견에 비해 효과가 덜하지만, 그럼에도 반대 의견이 없는 경우에 비해 효과적이다. 특히, 악마의 옹호자가 좀 더 고차원적이고 정교한 반대 논증을 제시했을 때 그 효과가 극대화될 수 있다(Sunstein & Hastie, 2015). 또한 협력적 논증 교환 수업에서 이질적 집단 구성이 중요하지만, 이질적이라는 것이 반드시 '반대 입장'을 의미하는 것은 아니다. 동기화된 추론을 개선하기 위해서는 서로 다른 의견을 가진 집단 내 구성원들 사이에서의 논증 평가가 활성화되어야 한다는 점에 주목할

필요가 있다. 적절한 논증의 평가가 결여된 논증 교환의 참가자들은 동기화된 추론으로 인해 초기 의견을 재확인하려는 경향이 강하고, 그 결과 양극화가 심화될 수 있기 때문이다. 그러한 상황에서 논증 참여자는 이성적인 합의에 도달하려고 하지 않는다. 오히려 그가 틀릴지도 모른다는 생각을 하지 않고 완강하게 자신의 기존 신념을 옹호하기 때문이다. 자신이 가지고 있는 이러한 비합리적인 태도와 상대방의 주장에 대한 비판적 평가의 결여는 합의 대신 입장의 양극화를 만들어내고 동기화된 추론을 이끌어낸다.

이와 관련하여 최근 연구는 논증 교환의 효과를 위해서 적대적인 반대 의견의 역할이 필수불가결한 것은 아니지만, 대립적 의견에 대해 이의를 제기하거나 다양한 대안을 검토할 수 있도록 문제를 제기하는 착한 악마의 옹호자(Angelic Devil's Advocate) 개념을 제안한다(Stevens & Cohen, 2020). 이는 대립적 논증 교환의 모델과 대립적 의견이 없는 완전히 협력적 논증 모델이라는 논증 교환의 양극단의 모델 사이의 절충적인 논증 교환의 양상을 설명한다. 즉, 논증 교환의 성공에 있어 악마의 옹호자와 같은 인위적인 '대립적 의견'은 오히려 인지적 강화와 같은 동기화된 추론을 유발하고 고착화하기 때문에, 대립된 의견보다는 다양한 근거들을 포괄적으로 검토해볼 수 있도록 적절한 질문을 던지는 착한 악마의 옹호자 역할이 필요하다는 것이다.

앞서 살펴본 것처럼, 동기화된 추론을 개선하기 위해서는 자신의 입장과 다른 입장에 대해 진정성 있게 검토하는 기회가 필요하다. 추론을 잘하기 위해서는 우리의 기존 신념에 대한 이의 제기가 필요하고, 우리의 논증에서 오류가 있다는 것을 평가해 줄 상대방이 필요하다. 스티븐슨과 코헨은 천사의 옹호자의 역할을 '스파링 파트너'로 비유한다(Stevens & Cohen, 2020: 903). 스파링 파트너는 복싱 경기에 앞서, 실전과 동일한 상황을 가정하고 연습하지만 실제로는 선수의 훈련을 도와주는 역할을 한다. 즉, 착한 악마의 옹호자는 동기화된 추론으로부터 벗어나고자 하는 목표를 위해 때로는 반대 의견을 제시하기도, 때로는 반대는 아니지만 유사한 다른 대안을 제시하기도 한다. 착한 악마의 옹호자는 논증에서

이기고 싶은 유혹을 피하면서도 협력적인 태도를 유지해야 한다. 이를 통해 학생은 자신의 의견을 종합적으로 검토하고 오류 가능성을 인지하며 다양한 대안을 수용할 자세를 함양하는 연습을 할 수 있다. 따라서 착한 악마의 옹호자는 논증 교환에 대한 이해도가 높은 전문적 기술을 가진 사람들이어야 한다. 필요에 따라 실제 논증에 참여하기도 하고, 빠져나올 수도 있어야 하며, 상대방이 제시한 논증을 객관적으로 분석할 수 있는 논리력을 갖추어야 한다. 또한 적절하게 질문을 던질 수 있는 능력이 필요하다.

착한 악마의 옹호자 모습은 새로운 개념으로 보이지만, 기존의 논증을 강조한 이론에서 제시된 다양한 형태와 유사하다(Perelman & Olbrechts-Tyteca, 1958). 일찍이 데카르트식 독단적 추론의 사용을 비판하고 논증 교환의 교육적 효과의 인식론적 기반을 제공한 페렐만과 올브레히츠-티테카의 보편적 청중(universal audience) 개념이나, 랄프 존슨의 이상적 대화자(ideal interlocutor)는 착한 악마의 옹호자의 구체적 사례를 보여준다.

페렐만과 올브레히츠-티테카는 보편적 청중에 대한 개념은 개인적 편향으로부터 자유롭게 논증을 객관적으로 평가할 수 있는 가상의 개인 집단이다. 보편적 청중은 모든 주장이 충족시키기 위해 노력해야 하는 합리성과 공정성의 기준을 나타내기 때문에 논쟁에 이상적인 청중으로 간주된다. 페렐만과 올브레히츠-티테카는 논증의 구성이 보편적 합리성의 원칙에 기초해야 한다고 주장하는데, 이는 논증이 그들의 구체적인 배경이나 신념에 관계 없이 보편적 청중을 설득할 수 있어야 한다는 것을 의미한다. 이들은 "유능한 청중을 향하지 않고 보편적인 원칙에 근거하지 않는 어떠한 주장도 무효"라고 말한다(Perelman & Olbrechts-Tyteca, 1958: 98). 즉, 보편적 청중을 설득할 수 있는 논증이 더 타당하고 신뢰할 수 있는 것으로 보이기 때문에 보편적 청중은 논증의 질에 대한 평가자 역할을 한다. 또한 보편적 청중은 실제 청중이 아니라 주장의 타당성과 설득력을 평가하는 데 사용되는 이론적 구성이다. 즉, 악마의 옹호자가 실재하지 않더라도 논증 교환의 상황에서 논증의 생성자는 보편적 청

중을 염두에 둔 논증 생성을 연습할 필요가 있다.

이와 유사하게 랄프 존슨은 이상적인 대화자의 중요성을 강조하는 논증 이론을 제시한다. 존슨에 따르면, 이상적인 대화 상대는 생산적인 논쟁에 참여하기 위해 필요한 지적, 정서적, 윤리적 기술을 가진 사람이다. 여기에는 주장을 이해하고 평가하는 능력, 대안적 관점을 고려하려는 의지, 논쟁에 참여하면서 협조적 태도를 유지하는 능력 등이 포함된다. 이상적 대화자가 갖춰야 할 구체적인 기술들은 다음과 같다(Blair & Johnson, 1987: 50-53). 먼저, 지적 능력이다. 증거를 식별하고 평가하며, 제시된 논증의 오류를 식별하고 강한 논증과 약한 논증을 구별하는 능력을 포함하여 논증을 이해하고 평가하는 능력이 포함된다. 다음은 감정 기술이다. 이상적 대화자는 열린 자세를 바탕으로 대안적인 관점을 고려하고 변화에 열려있는 의지와 논쟁에 참여하면서 협조적인 태도를 유지하는 능력을 지녀야 한다. 끝으로 윤리적 기술이다. 이상적 대화자는 논증의 핵심과 관계없는 인신공격과 같은 비윤리적인 전술의 사용을 피하고 논증에서 공정성의 원칙을 유지하면서 윤리적이고 책임감 있는 행동을 할 수 있는 능력이 필요하다. 이상적인 대화자는 진실의 발견과 갈등의 해결로 이어지는 비판적 평가와 협력의 과정에 참여할 수 있기 때문에 효과적인 논쟁의 필요한 구성요소이다.

착한 악마의 옹호자는 자신이 가진 신념과 별개로 그러한 역할을 수행하고, 상대방의 동기화된 추론 개선이라는 목표의식을 지닌 사람이어야 한다. 그러나 착한 악마의 옹호자는 우리 일상의 논증 교환 상황에서 거의 발견되지 않는다(Stevens & Cohen, 2020: 906). 개념을 고안한 스티븐스와 코헨은 이상적으로, 착한 악마의 옹호자들이 일상적인 논쟁에서 발견될 수 있다고 가정하는 것은 희망적인 생각일 것이라고 말한다. 실제 삶에서 우리는 누군가의 스파링 파트너로서의 역할보다는 직접 복싱에 참여하는 선수이기 때문이다. 또한 우리가 논증에 이해관계가 없을때에도 편파적이고 경쟁적인 성향에 지배되기도 한다. 우리는 자신의 논증에 결함이 있고 결론이 반박될 때 비추어질 자신의 모습을 걱정하며, 중요하지 않은 논증에서도 악의적인 적대적 태도에 너무 쉽게 굴복한다

(Stevens & Cohen, 2020: 906). 따라서 대부분의 논증 교환 참가자들은 자신의 기존 신념을 바탕으로 상대방을 설득하는 것에 초점을 두고 논증을 생성해낼 것이다. 이는 그 자체로 불완전한 논증 교환의 상황이다. 그러나 착한 악마의 옹호자가 현실에서 드물게 나타난다는 것은 이러한 역할이 결코 성공적으로 수행할 수 없다는 것을 의미하는 것이 아니다. 역으로 누군가가 우리의 말을 끝까지 듣고, 우리의 논증에 참여하고, 능숙하게 비판하는 상황에 놓이는 것 자체가 큰 행운임을 인지하는 것이 중요하다(Stevens & Cohen, 2020: 907). 사회적 상호작용에서 착한 악마의 옹호자가 없다면 그것은 불완전한 논증 교환일 것이다. 이 점을 고려하여 교사는 교실 상황에서 이질적 의견이 형성되지 않을 때, 교사 자신이 착한 악마의 옹호자 역할을 수행하거나 특정 학생-가령, 주제에 대한 자신의 의견을 정하지 못한 학생-들에게 이러한 역할을 부여할 수 있어야 한다.

제 2 절 협력적 논증 교환 수업의 사례 검토

1. 협력적 논증 교환 수업의 사례1: KMDD 수업 모형

협력적 논증 교환 활동은 우리나라 도덕과 교육의 맥락에서 동기화된 추론을 개선하기 위한 목적으로 수행될 수 있는 하나의 수업 방법으로 제시할 수 있다. 이는 도덕과에서 이미 개발되어 사용되고 있는 다양한 토론 기반 수업 모형 속에서 효과적으로 구체화 될 수 있다. 이하에서는 협력적 논증 교환 활동을 활용할 수 있는 도덕과 수업 모형을 검토하여, 수업의 절차, 수업 설계를 위한 지침, 수업 운영 시 유의사항 등 학교 현장에서 적용하기 위한 실천적인 시사점을 도출해보고자 한다.

협력적 논증 교환 활동은 일반적으로 인지적 작업의 대립적 분업의 양상을 띤다는 점에서 소크라테스의 대화법을 활용한 콜버그의 딜레마 토

론과 유사한 모습을 지니지만, 활동의 참여자들이 맡은 대립적 분업의 역할이 순환된다는 점에서 그 차이를 발견할 수 있었다. 이러한 차이에 주목해 볼 때, 협력적 논증 교환 수업은 콜버그의 딜레마 토론 수업을 개선한 린트의 콘스탄츠 방식의 도덕 딜레마 토론 수업(Konstanz Method of Dilemma Discussion, 이하 KMDD)에서 구체적으로 실현될 수 있다.

독일의 심리학자이자 교육가인 게오르그 린트는 콜버그의 도덕 발달 이론과 도덕 판단 측정 도구를 비판적으로 검토하며, 이를 개선하기 위한 방안과 함께 도덕적 역량을 촉진하기 위한 딜레마 토론 수업 방식을 수정·보완하여 제안하였다(Lind, 2016). 국내 도덕과 교육 학계와 현장 또한 그의 이론을 널리 수용하고 있다. 이는 린트의 도덕적 역량 개념과 이를 측정하기 위한 검사 도구에 대한 타당화 연구(박균열, 2016; 이인태, 박균열, 2019; 석자춘, 2021)뿐 아니라 해당 내용을 바탕으로 수업에 적용하고자 하는 노력(박균열, 2017; 강민겸, 박균열, 2019)도 함께 이루어졌다. 나아가 그의 도덕적 역량 교육 이론을 역량 중심 도덕과 교육과정을 실현하기 위한 이론으로서 검토되었으며(강민겸, 2019), KMDD로 구체화된 그의 수업 방식을 도덕과의 주요 교수-학습 모형으로 개발되기도 하였다(정창우 외, 2020: 82-91).

KMDD 수업의 목표와 구체적 절차를 확인하기에 앞서, 이론적 배경을 간단히 검토해보자. 린트는 콜버그의 도덕 발달 이론과 도덕성 측정 도구, 도덕 교육 이론을 비판적으로 수용하여 자신만의 도덕적 역량 교육 이론을 구축한다(강민겸, 2019). KMDD 수업은 그가 정립한 도덕적 역량을 획득하는 것을 목표로 하므로, 이 수업의 설계 원리를 이해하기 위해서는 먼저 그의 도덕적 역량 개념을 살필 필요가 있다. 그는 도덕적 행위가 가진 이중 측면(dual-aspect)에 주목한다. 이중 측면이란, 도덕적 행위를 함에 있어 정서적 측면과 인지적 측면이 마치 하나의 공을 구성하듯이 동시에 존재하지만, 구분되는 두 개의 측면으로 존재한다는 관점이다. 정서적 측면은 자신이 가지고 있는 도덕적 이상에 대한 열망과 개인적인 도덕적 선호로 표현되는 도덕적 지향을 의미한다. 린트에 따르면

민주주의 사회에서 모든 시민은 저마다 도덕적 이상을 이미 가지고 있으며, 그것에 따라 행동하고 싶어하는 열망을 가지고 있다. 그러나 그 열망을 실제 삶에서 실천할 수 있는지는 그가 가진 도덕적 역량에 따라 다르다. 도덕적 역량은 도덕적 행위의 인지적 측면에 해당하는데, 이는 자신이 추구하는 도덕적 이상이나 기존의 도덕적 신념을 원리로 하여 일상생활에서 이를 일관되게 적용하여 판단하고 행동할 수 있는 능력을 의미한다. 린트에 따르면 도덕적 역량을 갖춘 사람은 민주주의 사회에서 폭력이나 속임수 대신 숙고와 토론을 통해 평화적으로 문제를 해결할 수 있다(Lind, 2016).

KMDD는 이러한 도덕적 역량을 측정하기 위해 린트가 개발한 도덕적 역량 검사(MCT)의 구성원리를 충실히 반영한 수업이다. MCT가 기존의 도덕 판단 검사지와 다른 가장 큰 차이점은 찬성과 반대 의견 모두의 근거를 평가하도록 설계한 점이다. 만약 높은 도덕적 역량을 가진 사람이라면 어떠한 도덕적 이슈에 대한 자신의 도덕적 지향이 찬성일지라도, 같은 논리를 가진 반대 근거 또한 설득력 있는 근거로 채택할 수 있어야 할 것이다. 반대로, 낮은 도덕적 역량을 가진 사람이라면, 자신의 도덕적 지향에 의해 동기화된 추론을 보일 가능성이 높기 때문에 그는 자신과 반대되는 입장의 근거에는 모두 낮은 점수를 줄 가능성이 높다. 린트에 따르면 바람직한 도덕적 역량을 함양한 사람은 자신의 도덕적 지향에 근거한 입장뿐만 아니라 반대 입장을 지지하는 다양한 도덕적 원리에 대한 수용 여부를 각각 평가하여 일관된 의사 결정을 내릴 수 있는 사람이다. 따라서 MCT의 이러한 원리를 반영한 KMDD 수업은 기본적으로 학생들이 저마다 가진 '도덕적 지향'에 의해 동기부여를 일으킴과 동시에, 반대 입장에 대한 적절한 평가의 과정을 포함함으로써 학생들의 '도덕적 역량'을 향상시키는 데 그 목적이 있다. 특히 후자의 특징이 이전의 콜버그식 도덕 딜레마 토론과 구분되는 지점이자 KMDD에서 가장 핵심이 되는 절차이다.

이처럼 KMDD는 학생들의 도덕적 역량을 향상하고, 그들이 추구하는 도덕 원칙과 일상에서의 행동 사이의 불일치를 줄이는 것을 목표로 개발

되었다. KMDD 프로그램의 목적은 도덕적 갈등 상황 속에서 권력과 속임수가 아닌 숙고와 토론을 통해 보편적인 도덕 원칙에 입각한 문제 해결 능력을 향상시키는 것에 있다. KMDD의 목표는 다음과 같은 하위 기술을 습득함으로써 달성되는데(Lind, 2016: 104-105), 이는 협력적 논증 교환 활동과의 밀접한 관련성을 보인다.

1. 언어를 통해 자신의 도덕적 감정을 인식하고 표현하기(협력적 논증 교환의 기초)
2. 윤리적 상황을 둘러싼 환경을 주의 깊게 관찰하고 객관적으로 인식(논증 교환을 위한 준비)
3. 원칙의 중요성과 적절성에 따라 차별화(논증 평가의 자세)
4. 원칙 간의 갈등을 해결하는 데 도움이 될 수 있는 최우선 원칙 찾기(논증 평가의 기준)
5. 반대하는 사람들에게 자신의 입장을 표명(논증 교환의 참여)
6. 정서적, 사회적, 시간적 압박 속에서 반성과 토론의 기회를 주는 것(논증 수정의 기회)
7. 다른 의견을 단순히 용인하는 것뿐만 아니라 그것을 평가하여 올바르게 판단하기(논증의 재평가)

첫 번째 하위 기술은 숙고와 토론의 기초에 해당한다. KMDD 수업은 도덕적 행위의 이중 측면을 고려하여, 도덕적 역량과 함께 학생이 가지고 있는 기존의 도덕적 지향을 촉발하는 것을 하위 목표로 한다. 도덕적 행위의 이중 측면 이론에 따르면 도덕적 과제에 직면했을 때 우리는 즉시 도덕적 지향의 차원에 대해 감정적으로 반응한다. 이때 도덕적 역량은 자신의 도덕적 지향을 실제 도덕적 행위에서 실현하는 역할을 하는데, 이를 위해 KMDD는 주로 학생들에게 숙고와 토론의 기초가 되는 언어를 통해 자신의 도덕적 감정을 인식하고 표현하는 연습을 요구하는 것이다. 이 하위 목표는 협력적 논증 교환 활동의 기본 전제에 해당한다. 협력적 논증 교환은 냉철하고 합리적인 도덕 판단을 논증의 형태로 도출

해내는 것이 아니라, 주어진 윤리적 이슈와 갈등 상황에서 느껴진 즉각적인 도덕적 감정과 직관적 반응에 기반을 둔 입장에 대해 논증을 구성하는 것에서 시작되기 때문이다.

또한 KMDD의 참가자들은 자신의 도덕적 역량을 발휘하기에 앞서 자신이 직면한 윤리적 상황을 관찰해야 한다. 해당 상황에서 쟁점이 되고 있는 것이 무엇인지 정확히 파악하고, 도덕적 행위자로서 알아야 할 도덕적 지식과 사실적 차원의 정보들을 수집하여 자신의 도덕적 지향과 일치하는 도덕적 역량을 발휘할 준비를 해야 하기 때문이다. 이와 같은 두 번째 하위 기술은 논증을 생성하고 교환하는 과정을 통해 갖추게 되는 자세와 관련된다. 동기화된 추론이 포함된 도덕 판단과 논증은 확증 편향 등의 모습으로 왜곡될 가능성이 크다. 자신의 관점을 고수하고, 상대방을 설득하고자 하는 목적에 의해 동기화되었기 때문에 자신에게 유리한 정보는 과대평가하고, 자신의 도덕적 신념에 반하는 사실은 의도적으로 외면하거나 과소평가하여 논증을 생성하기 때문이다. 협력적 논증 교환의 참여자들은 자신의 논증에 대해 평가받고, 타인의 논증을 평가하는 과정에서 윤리적 상황을 더욱 객관적으로 인식하고 정보를 더욱 정확하게 평가할 수 있도록 하는 정확성 목표에 의해 동기화된다. 이를 통해 동기화된 추론의 문제는 점진적으로 개선될 수 있다.

세 번째와 네 번째 하위 기술은 KMDD에서 지향하는 도덕적 역량의 구체적 모습인 '도덕적으로 중요한 것은 중요하게, 사소한 것은 사소하게(high-high, low-low)'의 원칙을 반영한다. 학생들은 토론의 과정을 통해 도덕적 원칙들 사이에서 중요성과 적절성에 따라 우선순위를 세워보기도 하고, 논의를 바탕으로 최우선적인 원칙을 찾기도 한다. 이는 논증의 평가를 위한 구체적인 기준이 될 수 있다. 이러한 기준을 바탕으로 상대방의 논증을 평가하기도 하고, 평가된 자신의 논증을 점검해보는 경험을 갖게 된다.

다섯 번째와 여섯 번째 기술은 KMDD 수업에서 특히 강조하는 부분으로 기존의 콜버그식 딜레마 토론 방식과 구분되는 지점이다. 학생들은 자신의 입장에 반대하는 사람들에게 공개적으로 자신의 주장을 제시하

고, 반대 의견을 명시적으로 평가해 보는 절차를 통해 정서적, 사회적, 시간적 압박 속에서 반성과 토론의 기회를 갖게 된다. 비록 콜버그식 딜레마 토론에서도 이와 같은 활동이 수행될 수 있지만 이는 수업을 진행하는 교사의 재량에 달려있다. 이와 달리 KMDD는 반대 의견을 표현하고, 이를 평가하는 절차를 반드시 포함하고 있다. 이 점이 협력적 논증 교환 활동이 KMDD에서 효과적으로 수행될 수 있는 접점이다. 대립하는 의견을 표명하고 그것에 대해 상호 간 평가를 수행하는 활동은 협력적 논증 교환 수업의 핵심 사항이기 때문이다.

끝으로 일곱 번째 하위 기술은 KMDD를 통해 최종적으로 도덕적 역량을 발휘하는 것과 관련이 있다. 토론을 통해 자신의 입장을 비합리적으로 고수하거나, 무비판적으로 자신의 의견을 쉽게 바꾸는 것이 아니라 반대 의견 교류를 통해 자신의 입장을 스스로 재평가하여 수정의 기회로 삼는 것이다. 이러한 기술은 협력적 논증 교환 활동이 궁극적으로 지향하는 바이다. 학생들은 협력적 논증 교환 활동을 통해 동기화된 추론을 촉발한 자신의 기존 도덕적 신념과 정체성에 대해 점검해 보는 기회를 얻게 되며, 정확성 목표를 추구하고 적절한 방향성에 동기부여 되는 것을 목표로 한다.

이와 같은 KMDD의 목표와 하위 기술은 다음의 일반적인 9단계의 절차를 통해 도덕과 수업에 적용될 수 있다.

[표 5-1] KMDD의 일반적 절차

단계	절차	내용	수업흐름
1단계	도덕적 문제사태 제시	도덕 판단을 요구하는 갈등 상황이 포함된 도덕 딜레마 제시	도입
2단계	개인별 숙고	딜레마 속 주인공의 상황에 대한 공감 주인공의 선택에 대한 자신의 입장 생각	전개
3단계	딜레마 이야기 명료화	딜레마 속 단서들에 대한 정보 확인	
4단계	첫 번째 투표	주인공 선택의 도덕적 가치에 대한 동의 여부 찬반 투표	
5단계	첫 번째 소그룹	같은 입장을 지닌 사람들끼리 모여 주장을	

	활동	정당화하기 위한 근거 함께 모으기	
6단계	학급 전체 토론	찬반 양 진영으로 나뉘어 토론 진행	
7단계	두 번째 소그룹 활동	토론에서 나온 반대 의견 중 가장 설득력 있는 상대의 근거를 선정하고 평가하기	
8단계	최종 투표	주인공 선택의 도덕적 가치에 대한 동의 여부 재투표	
9단계	정리 및 성찰	변화된 자신의 입장과 정당화 근거를 정리하기	정리

KMDD의 수업 절차에서 협력적 논증 교환 활동은 주로 4-7단계에서 이뤄진다. KMDD 절차에서는 명시적으로 자신의 도덕적 논증을 생성하는 절차가 마련되어 있지 않지만, 학생들은 개별적 숙고하고 찬반 투표를 통해 자신의 입장을 공개적으로 표현하게 된다. 또한 자신과 입장이 같은 사람들과 함께 주장을 정당화하기 위한 근거를 함께 모으는 절차가 있다. 이 절차에서는 논증을 생성해내는 과정에서 도덕적 신념이 유사한 사람들과 함께 입장을 옹호하기 위한 근거를 모으기 때문에 동기화된 추론이 활성화될 가능성이 크다. 그러나 소그룹으로 모여 함께 논증을 생성하는 작업은 개별적으로 논증을 생성하는 과정에서 발생할 수 있는 동기화된 추론으로 인한 확증 편향의 가능성을 최소화해준다. 이 과정을 통해 협력적 논증 교환 활동 참여자들은 자신과 입장이 다른 사람들을 설득하기 위한 도덕적 논증을 만들어내며, 논증 교환의 준비를 마친다.

이후 전체 토론 활동은 본격적으로 논증의 교환이 이뤄지는 단계이다. 이 절차에서는 두 가지 규칙이 적용된다. 하나는 수업에서 제시된 모든 의견은 그 자체로 존중되어야 하며 상대방의 입장(의견)에 대하여 인격적 비난을 포함한 일체의 긍·부정적 평가를 금지한다는 것이다. 다른 하나는 찬성 측과 반대 측 사이의 의견 교환을 번갈아 가면서 진행해야 한다는 것이다. 이 규칙은 동기화된 추론을 개선하기 위한 협력적 논증 교환 활동에도 주요한 시사점을 제공한다. 기본적으로 협력적 논증 교환의 양상은 '대립적 의견'의 교환이기 때문에 각자 동기화된 추론으로 인해 서로의 입장에 대해 적대적인 입장을 전제하게 된다. 따라서 상대방의 논증을 '객관적'으로 판단하고 그 과정에서 동기화된 추론을 효과적으로

개선하기 위해서는 일단 상대방이 무엇을 주장하고, 어떠한 근거로 주장을 뒷받침하는지 정확히 파악할 필요가 있다. KMDD의 전체 토론에서 제시된 첫 번째 규칙은 이처럼 정확성 목표를 강화해주는 역할을 수행한다. 상대의 입장에 대한 아무런 가치 평가 없이, 일단 있는 그대로의 논증을 확인해야 정확한 평가가 가능해지기 때문이다. 또한 찬성 측과 반대 측 의견을 교대로 교환함으로써 자신의 도덕적 지향이라는 강한 방향성에 의한 동기화된 추론의 개입을 최소화될 수 있다. 협력적 논증 교환의 참가자들은 자신의 입장만을 일방적으로 제시하고, 타인의 입장을 무조건 부정적으로 평가하는 것이 아니라 자신과 유사한 입장과 대립되는 입장을 동등하게 수집하는 과정을 거치면서 적절한 방향성과 높은 정확성에 의해 동기부여될 것으로 기대할 수 있다.

또한 KMDD의 가장 두드러지는 특징이 나타나는 7단계는 자신의 의견과 반대되는 입장 중 가장 매력적인 논증을 평가해 보는 단계이다. 높은 도덕적 역량을 지닌 사람은 자신의 도덕적 지향을 맹목적으로 고수하는 것이 아니라, 다양한 입장을 두루 고려하여 합리적이고 일관된 판단과 의사결정을 내리는 사람이다. 이를 위해서는 다른 사람과의 논증 교환을 통해 자신의 기존 신념과 판단을 미세하게 조정할 수 있어야 한다. 이를 통해 도덕 판단에서 동기화된 추론으로 인해 발생하는 문제들이 개선될 수 있다. 자신과 입장은 다르지만, 설득력 있고 매력적인 논증을 확인함으로써 참가자들은 지적인 겸손을 함양할 수 있고, 합당한 도덕적 태도의 불일치를 탐색할 수 있게 된다. 또한 높은 도덕적 역량의 함양은 자신의 도덕적 지향과 그를 바탕으로 하는 일관된 판단 및 행동을 전제하기 때문에 비도덕적 행위의 정당화 문제 또한 개선하는 데 발판을 마련해준다. 즉, KMDD의 일곱 번째 단계는 협력적 논증 교환을 통한 동기화된 추론 개선의 구체적인 방법이라고 할 수 있다.

다만, KMDD를 통해 동기화된 추론을 개선하기 위한 목적으로 협력적 논증 교환 활동을 수행할 경우, 주의해야 할 사항이 있다. 구성원 간 신뢰의 문제이다. 입장이 다른 사람들과 찬성, 반대로 양분하여 집단 대 집단으로 토론을 수행하게 되면 동기화된 추론으로 인한 문제가 심화될 여

지가 있기 때문이다. 집단 토론을 통해 집단적 사고(Janis, 1982)나 양극화의 문제가 심화(Sunstein, 2002)된다는 연구 결과들이 이를 지지한다. 자신과 입장이 다른 사람과의 토론은 오히려 자신의 사회적 정체성을 보호하기 위한 방향으로 동기화된 추론이 작동하기 때문이다. 집단적 사고나 집단 양극화는 일반적으로 모든 집단 구성원이 처음부터 동의할 때 발생한다. 그리고 나서 그들은 자신의 의견을 지지하는 주장, 즉 모든 사람들이 그들의 결론에 동의하기 때문에 비판적으로 검토되지 않는 주장을 찾기 위해 추론을 사용한다(Mercier, 2011b: 185). 따라서 이를 개선하기 위해서는 집단 내 구성원 모두가 서로를 신뢰하고 그들의 입장에 대해 진지하게 평가할 수 있는 기반이 마련되어야 한다. 만약 여건이 어렵다면, 대안은 소그룹으로 논증 교환을 하는 것이다. 찬성과 반대 양극단의 집단이 대립적으로 토론하는 것에 비해, 소그룹 내에서 소수의 구성원들 간의 인지적 작업의 대립적 분업은 서로 다른 입장을 더욱 진지하게 고려할 가능성이 높기 때문이다.

KMDD를 고안한 린트는 수업 진행에 있어 교사의 전문성을 강조한다. 교사는 학생들의 도덕 발달에 대한 전문적인 지식을 가져야 할 뿐만 아니라, 윤리적 상황을 분석할 수 있는 능력, 딜레마 토론을 운영할 수 있는 효과적인 기술뿐 아니라, 자유로운 토론이 가능하도록 협력적이고 지지적인 교실 분위기를 조성할 수 있어야 한다. 교사는 KMDD를 운영하는 과정에서 반드시 지켜야 할 원칙이 있다(Lind, 2016: 105-107). 이는 토론을 진행하는 교사의 최소한의 역할로 이해된다.

첫째, 교사는 적절한 딜레마 이야기를 선택해야 한다. 집중력을 극대화하고 학생들의 도덕적 감정을 불러일으킴과 동시에 자발적인 참여를 유도하기 위해 교사는 학생들의 특성을 고려하여 그에 따른 딜레마 이야기를 제작하거나 선택해야 한다. 특히 교사는 학생들의 진정성 있는 도덕판단을 도출해내기 위해서는 딜레마 이야기가 적정 수준의 현실성을 갖추어야 한다고 보았다. 이는 학생들에게 적절한 도덕적 감정을 이끌어내고, 실제 삶에서 적용 가능한 도덕적 역량을 기를 수 있기 때문이다.

둘째, 교사는 토론의 과정에서 학생들에게 지지(support) 및 도전

(challenge)의 경험을 제공할 수 있어야 한다. 교사는 문제를 명확히 하고, 집단 토론의 과정에서 학생들의 논증을 명료화 해주며 칭찬을 제공할 수 있다. 도전 단계에서는 의사 결정, 공개 주장, 반대 의견 청취 및 투표가 포함된다. 이러한 지원과 도전의 교대는 학생들이 집중력을 유지하고 도덕적 감정을 적절히 유지하도록 한다. 셋째, 학생들이 토론의 규칙을 잘 지키도록 도와야 한다. 자기 절제는 민주적 토론의 참여자들에게 요구되는 필수사항이다. 특히, 토론의 과정은 학생들이 자신의 입장을 정하는 동시에 편견 없이 반대 의견을 듣고 평가하는 민주적인 과정이어야 한다. 이를 위해 교사의 개입을 최소화하고, 두 가지 큰 규칙을 바탕으로 토론이 진행된다. 전체 토론에서 적용되는 규칙으로, 타인의 입장을 있는 그대로 받아들이기 위해 판단을 금지하는 것과 찬성-반대 측이 번갈아 발언한다는 규칙이다. 이를 통해 학생들은 권위자의 개입 없이도 자율적인 분위기 속에서 합리적이고 비폭력적으로 문제를 해결하는 경험을 할 수 있다.

마지막으로 토론은 사람이 아닌 문제에 초점을 맞춰야 한다. 토론에서 다뤄지는 문제에 대한 의견 공유는 개인화되어서는 안되며 학습자가 갈등을 통해 도덕의 핵심을 생각하도록 유도해야 한다. 특히, 학습자는 자신의 발화에 제한을 두지 않고 자신의 도덕 원칙에 따라 서로의 의견을 경청하고 평가한다. 참가자들은 도덕적 능력 향상을 위한 토론에서 편견, 속임수, 금기를 피하고 서로를 동등하게 대해야 한다(Lind, 2016: 106).

이와 같은 KMDD 수업의 운영 원칙은 협력적 논증 교환 활동이 원활히 이루어지기 위한 교사의 역할에 주요한 시사점을 제공한다. 첫째, 수업 목표에 대한 교사의 인식이다. KMDD의 수업은 기본적으로 도덕적 갈등 상황에서의 '평화적인 문제해결'의 능력으로서 도덕적 역량 증진을 목적으로 한다. 린트가 제시한 도덕적 역량이란 구체적으로 콜버그가 제시한 도덕 판단 능력과 밀접한 관련을 지닌다. 즉, 높은 수준의 정의 추론의 구조를 갖추고, 이를 바탕으로 도덕적 원칙을 일관되게 적용하는 능력을 갖추는 것이 KMDD 수업이 지향하는 궁극적인 목적이다. 반면, 본고에서 제안하는 협력적 논증 교환 활동의 일차적 목적은 동기화된 추

론의 개선에 있다. 이를 위해 교사는 학생들이 높은 정확성 목표에 의해 동기부여 되고, '협력'을 지향하며 적정 수준의 방향성에 의해 동기화된 추론을 발휘하도록 이끌어야 한다. 협력적 논증 교환의 활동이 KMDD 수업에서 나타나는 의견의 교환과 반대 의견에 대한 평가와 유사한 면이 있지만 활동의 목적이 상이하기 때문에 기대되는 효과 또한 다르게 나타날 수 있다. 따라서 교사는 동기화된 추론의 개선을 목적으로 협력적 논증 교환 활동을 위해 KMDD 수업의 모형을 적용하는 경우, 수업의 목표(target)를 명료화할 필요가 있다.

둘째, 교사는 학생들의 감정적 동기를 유발하고 적극적으로 논증 교환에 참여할 수 있도록 적절한 윤리적 이슈를 제시할 수 있어야 한다. 협력적 논증 교환의 전제는, 학생들의 진정성 있는 도덕적 입장에 대한 표출에 있고, 이는 자신이 가지고 있는 기존의 도덕적 신념에 기반한 논증의 생성에 있다. 그 과정을 통해 학생들은 타인을 설득하기 위한 맥락이라는 추론의 사회적 기능이 발휘되는 상황에 참여하게 되고, 동기화된 추론을 경험할 수 있다. 이를 위해 교사는 첨예하게 대립하는 주장이 존재하는 윤리적 이슈를 선정해야 한다. 나아가 교사는 학생들의 도덕적 감정을 불러일으키기 위해 적절한 발문을 제공해야 한다. 이와 관련하여 KMDD를 고안한 린트는 좋은 도덕적 딜레마 이야기의 여러 조건을 제시하였는데, 동기화된 추론의 관점에서 주목할 지점은 KMDD의 교육적 효과를 위해 '반-사실적 딜레마(semi-real)'를 사용해야 한다는 원칙이다. 완전히 가상적인 딜레마 스토리는 학생들이 토론 과정을 현실과 분리하여 받아들이게 하므로 수업의 효과가 실제적인 역량으로 전이되는 것을 방해할 수 있다, 지나치게 현실적인 스토리는 학습자가 토론의 결과에 직접적으로 영향을 받는다는 것을 의미할 수 있기 때문에 학습자들의 방어적이고, 문제를 회피하려는 태도를 강화시킬 수 있다. 따라서 딜레마 스토리는 사실에 기반하면서도 가공된 방식으로 구성되어야 하며, 교사는 이것을 제시할 때 충분히 현실에서 있을 수 있다는 분위기를 조성해야한다. 이를 통해 반-사실적 딜레마는 개인의 도덕적 원칙 사이에 갈등을 일으킬 수 있고, 참가자들 사이에 논증을 초래하여 진정한 도덕적 감

정으로 이어질 수 있도록 동기유발한다. 이는 동기화된 추론을 유발할 수 있을 만큼 충분한 도덕적 감정을 발생시키지만, 그것으로 인해 협력적 논증 교환 활동 자체를 방해할 만큼 매우 강렬하지는 않은 것이다. 이러한 딜레마는 실제 바이러스를 물리치는 신체의 능력을 향상시키기 위해 약화된 바이러스가 도입되는 백신과 유사한 목적을 제공한다. 따라서 교사가 도덕적 이슈를 제시할 때에는 학생들이 학습을 방해하기에는 너무 현실적이지도 않아야 하지만, 동기화된 추론을 유발하여 협력적 논증 교환에 참여시킬 수 있을 만큼 충분히 현실적이어야 한다.

셋째, 교사의 개입이 최소화되어야 한다. 토론 기반 수업에서 교사 개입은 학생들이 주제를 보다 정확하게 이해할 수 있도록 안내하는 기능을 수행하기도 하지만, 이러한 개입이 때때로 학생들의 기존 편견과 신념을 강화함으로써 동기화된 추론을 고착화할 가능성도 있기 때문이다. 학생들이 동기화된 추론을 바탕으로 문제에 접근할 때, 그들은 자신의 신념을 뒷받침하는 정보를 찾지만, 신념과 모순되는 정보는 무시할 가능성이 더 높다. 이는 학생들이 의미 있는 대화에 참여할 수 없게 하고, 그 대신 자신의 견해를 옹호하는 데 초점을 맞추게 되므로 양극화되고 비생산적인 토론을 초래할 수 있다. 토론 기반 수업에서 교사의 개입은 때때로 이 문제를 악화시킬 수 있다. 예를 들어, 교사가 자신의 도덕적 신념에 근거하여 토론에서 특정 입장의 편을 들거나, 특정 관점을 뒷받침하는 정보를 편파적으로 더 많이 제공한다면, 학생들은 자신의 견해가 객관적으로 검증되고 있다고 느낄 가능성이 있고, 그들의 입장에서 더 확고하게 자리 잡게 될 것이기 때문이다. 이것은 학생들이 더 이상 비판적인 사고에 참여할 수 없고 대신 그들 자신의 신념을 지키는 데 집중하는 상황으로 이어질 수 있다. 특히, 교실 수업의 상황에서 도덕 교사는 도덕적 권위자이다. 즉, 교사의 도덕적 견해는 논증 교환의 상황에 직, 간접적인 영향을 줄 수 있으므로 동기화된 추론의 개선을 위해 교사의 개입은 최소화 되어야 할 것이다.

2. 협력적 논증 교환 수업의 사례2: 건설적 논쟁 수업 모형

논쟁 수업 모형은 인지 갈등을 일으키는 논쟁적 주제를 통해 비판적 사고 능력과 갈등 상황에서의 문제 해결 능력을 기르기 위한 목적으로 고안된 교수-학습 모형이다. 논쟁 수업은 미국에서 자유주의 교육 철학의 기조 속에서 정립된 것으로, 60년대 존슨과 존슨 형제에 의해 개발되었으며, 우리나라에는 90년대 후반 사회과와 도덕과 중심으로 수용되었다. 특히, 도덕과 학문 공동체는 주로 논쟁 문제 수업에 있어 교사의 역할에 대해 주목하였다(김윤경, 박균열, 2019; 추병완, 2020d; 정창우, 김하연, 2021). 나아가 도덕과에서는 '건설적 논쟁'의 개념을 도입하여 보다 내실 있는 윤리적 시민 교육을 위한 연구 관심이 이어지고 있다(김하연, 2020). 최근에는 건설적 논쟁 개념을 적용한 논쟁 수업 모형을 도덕과의 특성에 맞추어 수업 모형으로 개발하기도 하였다(정창우 외, 2020: 141-148).

건설적 논쟁(constructive controversy)이란 두 개인이 상반된 생각, 정보, 결론, 이론 또는 의견을 가지고 공통된 합의를 찾기 위해 노력하는 상황을 말한다(Johnson et al., 2000: 29; Johnson & Johnson, 2009: 38). 논쟁 문제는 사회가 합의에 도달하지 못한 주제이며, 각각의 제안된 해결책은 강력한 지지와 반대 의견으로 극명하게 나뉜다. 이는 소중한 신념, 경제적 이익 또는 원칙이 위협받는 것으로 인식됨에 따라 발생하는 인지적, 도덕적 갈등(conflict)을 포함한다. 이와 같은 갈등은 개인의 마음에 양립할 수 없는 생각이 존재하거나 새로운 정보가 기존 지식과 모순될 때 발생한다(Johnson & Johnson, 2009: 39).

갈등은 학생들이 새로운 정보를 찾고 더 열심히 그리고 더 오래 학습하도록 격려하는 동기를 제공한다. 수업에서 건설적인 논쟁을 구조화함으로써, 교사는 학생들의 관심을 끌 수 있고 그들이 수업에 등록했을 때 의도했던 수준 이상으로 배우도록 그들을 격려할 수 있다(Johnson et al., 2000: 32). 또한 인지적, 도덕적 갈등은 잠재적인 건설적 결과 때문에 중요한 교육 도구가 될 수 있지만, (잠재적인 파괴적 결과 때문에) 교육

상황에서는 거의 구조화되지 않는다(Johnson & Johnson, 2009: 37). 따라서 교사들이 교실 수업에서 갈등을 활용하여 이것의 긍정적인 효과를 구현하기 위해서는 연구에서 검증된 사회과학 이론에 기초해야 한다. 건설적 논쟁은 학생들 사이에 이와 같은 갈등을 유발하기 위해 고안된 교육 절차이며 이러한 기준을 충족시킨다. 이들에 따르면 지적 갈등은 다음과 같은 교육적 효과를 지닌다(Johnson & Johnson, 2009: 36-37).

1. 학습 내용과 수업 과제에 학생을 주의 집중시킨다.
2. 학생들이 과제를 완수할 수 있도록 격려하고, 새로운 정보를 찾기 위해 더 많은 시간과 노력을 투입하여 학습에 임하도록 한다.
3. 과정이 끝난 이후에도 학생들이 해당 주제에 대해 계속 관심을 가지고 배우도록 동기부여한다.
4. 관련 과제를 완료한 후에는 더 높은 수준의 인지적 추론 능력을 기대할 수 있다.
5. 역할채택을 더 정확히, 더 자주 사용할 수 있게 된다.
6. 다양한 자료를 학습하는 능력과 더 높은 성취를 보일 수 있다.
7. 더 높은 수준의 창의적인 생각을 만들어낼 수 있다.
8. 학생들 사이에서 더 긍정적인 관계가 형성된다.
9. 학생들의 자존감을 높여준다.

그러나 지적 갈등이 항상 교육의 건설적이고 바람직한 측면으로 간주되는 것은 아니다. 인지적, 도덕적 갈등의 잠재적인 이점에도 불구하고, 이것은 종종 교육적 상황에서 구조화되지 않는다. 존슨과 존슨은 교사들이 수업에서 갈등을 회피하는 대표적인 이유를 교사의 태도에서 찾는다(Johnson & Johnson, 2009). 많은 교육자들은 갈등이 긍정적인 결과보다 부정적인 결과를 낳을 가능성이 더 높다고 생각하는데, 이는 갈등을 피하는 것이 실제로 대인관계와 집단의 성과를 저해할 수 있으며, 갈등을 처리하는 과정에서 심리적으로 고통스러운 과정이 필연적으로 포함되기 때문이다. 그 밖에도 교사의 수업에 대한 통제력을 잃을 수 있다는 부담

감, 학생들 사이의 사회경제적 지위 격차로 인한 양극화의 심화 등의 이유 때문에 열린 마음으로 갈등을 처리하는 데 어려움을 겪는다(Johnson & Johnson, 2009: 38). 이러한 다양한 요인들로 인해 교사는 교실 수업에서 갈등 구조화를 꺼리게 되는 것이다. 또한 몇몇 교육 연구자들은 지적 갈등이 분노, 거부, 분열, 손상된 관계, 불신, 학습에 대한 헌신 감소 등과 같은 부정적인 결과를 초래할 수 있기 때문에 종종 문제가 있다고 인식한다. 갈등은 스트레스, 불안, 심지어 신체적 피해를 유발할 수 있으며 창의성, 결석, 규율 문제, 학업 성취도 저하로 이어질 수 있다. 또한 갈등은 수동적-공격적 행동, 학대적 행동, 신뢰성 상실, 의사소통 오류, 협력 감소, 파벌의 형성으로 이어질 수 있다(Johnson & Johnson, 2009: 37-38). 이처럼 일상생활에서 갈등과 논쟁의 과정은 자연스럽게 발생하지만, 학습 상황에서 교사들이 그것을 구성하는 방식에 의해 교육적 효과를 거둘 수도, 오히려 갈등을 심화시킬 수도 있다.

교육 환경에서 논쟁적 주제(controversial issue)와 관련하여 건설적인 대화의 모습은 '건설적 논쟁 수업'을 통해 구체화 될 수 있다(Johnson & Johnson, 2009). 이들은 이러한 유형의 갈등이 인지적 균형을 무너뜨리는 데 도움을 주고, 더 높은 수준의 인지적 또는 도덕적 발달로 이어질 수 있다는 생각을 기반으로 건설적 논쟁 수업을 개념화하고 구체화한다. 건설적 논쟁 수업은 학생들 사이의 생산적인 의견 불일치와 협력을 장려하기 위해 고안된 교수법을 말한다. 건설적 논쟁의 이면에 있는 생각은 학생들이 적극적으로 학습 과정에 참여하고 비판적이고 창의적으로 생각하도록 장려될 때 가장 잘 배운다는 것이다.

특히 이들은 아이들의 도덕 발달을 촉진하기 위한 방법으로 건설적 논쟁을 활용한 협력적 논증 교환 활동을 제안한다(Tichy et al., 2010). 건설적 논쟁이 학생들의 도덕 발달과 윤리적 기술, 사회적 상호작용에 대한 태도에 있어 다른 학습 방식에 비해 효과적임을 보이기 때문이다(Tichy et al., 2010: 779-781). 건설적 논쟁에서, 학생들은 논란의 여지가 있는 문제를 제시받고 그 문제에 대해 반대 입장을 취하도록 요청받는다. 그리고 그들은 자신의 입장을 제시하고 옹호하는 동시에 동료 학생

들의 관점을 고려하는 구조적인 토론에 참여한다. 이 활동의 목표는 합의에 도달하는 것이 아니라 학생들이 비판적으로 생각하고, 다른 관점을 고려하고, 그들 자신의 주장과 해결책을 개발하도록 격려하는 것이다. 존슨은 건설적 논쟁이 학생들이 그들의 초기 신념을 비판적으로 검토하고 이를 수정할 기회를 제공하는 효과적인 교육 방법이라고 주장한다(Johnson & Johnson, 2009). 이처럼 건설적 논쟁 수업은 학생들이 잘 구성된 논증 교환 활동에 참여함으로써, 학생들이 도덕적인 추론 능력을 개발하고, 도덕적인 문제에 더 민감해지고, 그들의 도덕적인 믿음에 따라 행동할 수 있다고 전제한다(Tichy et al., 2010: 782-783).

건설적 논쟁 수업의 일반적 절차를 제시한 존슨 연구진은 이 과정이 의견 합일(concurrence seeking), 단순 토론(debate), 개인적 노력(individualistic effort)과 구분되는 점을 강조한다. 먼저, 의견 합일은 서로 다른 의견을 가진 사람들 사이에서 의견의 일치를 추구하는 과정으로, 협력과 갈등 회피가 결합한 것이다. 즉, 양측은 각자의 입장을 준비하고, 가능한 최선의 사례를 제시하며, 일단 의견 차이가 있음을 깨닫고, 의견 차이에 대해 불안해하며, 의견 차이를 즉시 끝내는 타협적 입장을 찾아 갈등을 피하고 억제하려고 한다. 의견 합일의 추구는 표면적으로는 평화적인 갈등의 해결방안처럼 보일 수 있지만, 대안적 아이디어와 행동 방침에 대한 현실적 고려를 피하기 위해 합의를 우선시하고 의견 불일치 또는 논쟁을 억제하는 집단을 포함한다(Johnson & Johnson, 2009: 39). 이러한 유형의 행동은 의사결정 집단의 구성원들이 합의에 부합하기 위해 자신의 의심과 유보를 무시하는 집단사고의 개념과 유사하다(Janis, 1982). 이는 집단의 화합을 유지하고 외부 스트레스 요인에 대처하고 자존감을 보존하려는 강한 욕구에 의해 추진된다.

승자와 패자를 구분하는 고전적 의미의 토론은 양립할 수 없는 입장을 주장하는 개인과 논쟁의 가장 좋은 표현을 바탕으로 승자를 결정하는 최종 판결을 포함한다(Johnson & Johnson, 2009: 39). 고전적 의미의 토론의 과정은 협력적이라기보다는 경쟁적이다. 양측은 각자의 입장을 준비하고, 가능한 한 최선의 사례를 제시하고, 반대 입장에 반박을 시도하고,

상대방이 제시한 반박을 재반박한다. 특히 학생들은 승자가 되고자 하는 목적으로 참여할 것이고, 교사가 자신들을 승자 팀으로 선언해주길 기대할 것이다. 따라서 토론의 과정은 표면적으로는 협력적 논증 교환의 과정과 동일하게 시작되지만, 도전을 받아 만들어진 불확실성은 다른 관점과 부조화를 이루는 정보에 대해 폐쇄적이고 방어적인 거부를 초래한다.

끝으로 개인적 노력은 그들의 목표가 서로 관련되거나 의존적이지 않은 개인이 서로 상호작용하지 않고 독립적으로 일하는 것을 의미한다(Johnson & Johnson, 2009: 39). 개인은 문제의 양면을 연구하지만 구두 진술을 하지 않는다. 그들의 초기 결론은 이의가 없으며, 그들은 자신이 처음에 생각했던 것을 자유롭게 확인할 수 있다. 따라서 협력적 논증 교환의 과정을 포함하는 건설적 논쟁 수업에 비해 더 나은 해결책을 찾거나, 자신의 기존 신념을 수정하는 실질적인 기회를 제공 받는 것에 제한이 있다.

존슨과 존슨이 구분한 수업의 다양한 양상은 추론의 사용에 대한 메르시에의 구분과도 일맥상통한다. 메르시에에 따르면 대화 상대방을 전제하지 않는 개인의 독단적 추론은 갈등의 유무에 따라 그 방향성과 정확성 동기에 다소간 차이는 있지만 확증 편향의 경향이 짙다(Mercier et al., 2016: 693). 이는 추론의 결함이 아니라 자신의 신념을 강화하는 과정에서 사용된 추론의 정상적 결과이다. 의견 합일은 앞서 추론의 양상에 대한 메르시에의 구분에 있어 다른 의견이 없는 쌍방향 논증 교환에 해당하며, 이로 인한 동기화된 추론의 경향과 일치한다(Mercier et al., 2016: 695). 추론의 사용에 있어 논증 교환의 상대방이 전제되지만, 이들은 도덕적 사안에 대해 유사한 견해를 공유한다. 따라서 추론은 각자의 도덕 기반과 초기 직관을 옹호하는 방향으로 동기화되는 양상을 띤다. 끝으로, 승패를 가르는 단순 토론은 추론의 사회적 기능이 충분히 고려되기 위한 조건들이 간과되기 때문에 협력적 논증 교환과 구분된다. 단순 토론과 달리 협력적 논증 교환은 논증을 생성하고 평가하는 기능을 통해 의사소통 참가자 모두에게 이익이 되고, 비록 의견이 다르지만 공동의 목표를 가지고 서로에 대한 신뢰를 전제로 협력적으로 문제를 해결

하고자 하는 동기가 고려되어야 할 것이다. 이러한 조건이 결여된 상태에서의 단순 토론은 기존 신념을 강화하는 방향으로 결론에 도달할 가능성이 높기 때문이다.

건설적 논쟁 수업은 '협동 학습'과 '구조화된 지적 갈등'(학생들이 문제 해결과 합리적인 판단을 자극하기 위해 문제에 대해 찬성과 반대 의견을 주장하는 것)을 구성요소로 한다(Johnson et al., 2000: 29). 먼저 '협동 학습'의 측면에서 수업의 목표는 각 학생이 문제에 대해 깊이 이해하고, 각자 주어진 역할을 수행하고 이를 통합함으로써 '최선의 이성적 판단'에 도달하는 것에 있다. 이는 동기화된 추론의 개선의 관점에서 정확성 동기를 높이고자 하는 목표와 잘 부합한다. 이를 위해 교사는 교실 수업에서 학생들을 4인 1모둠으로 구성하고 각 모둠을 다시 2인 1팀으로 총 2팀으로 나눠 주어진 문제에 대한 찬/반, 긍정/부정 등 대립되는 입장을 배정해준다. 학생들은 제시된 문제에 대한 자신이 맡은 입장을 진술한 다음 문제에 대한 보고서 작성 등의 모둠 과제를 배정받는다. 두 명으로 구성된 한 팀은 자신이 맡은 입장과 관련된 가능한 다양한 정보를 살펴보고, 이를 바탕으로 최선의 논증을 생성하고, 옹호하는 임무를 부여받으며, 다른 한 팀 또한 반대 입장에 대해 같은 방식으로 활동을 수행한다. 독단적인 추론의 사용은 확증 편향의 위험에 노출되지만, 이 과정은 두 명이 함께하며 정확성 동기를 추구하게 된다. 물론, 논증을 생성하는 과정에서 같은 입장에 놓인 사람들 간의 대화를 포함하기 때문에 편향된 결론에 도달할 가능성은 여전히 존재한다. 이후 교사는 학생들이 각 입장에 대해 설득력 있는 주장을 구성하고, 반대 입장을 듣고 학생들이 스스로 공격에 반박하도록 돕는다. 논증 교환의 과정을 통해 자신의 입장을 위협하는 논증을 평가해보고, 자신의 입장을 점검하면서 정확성 동기를 높일 수 있다. 그 후, 두 팀은 서로 반대편의 관점을 취해본다. 그리고 나서 그들은 두 팀의 입장을 모두 종합하고 학생들의 가장 합리적인 판단을 나타내는 합의를 추구한다.

'구조화된 지적 갈등'의 측면에서 수업의 목표는 갈등 요소가 포함되어 있는 논쟁적 이슈를 이해하고 문제를 해결하거나 이에 대한 자신의 신념

을 점검하는 것에 있다. 논쟁적 이슈를 적절히 대처하기 위해 학생들은 양측의 입장과 관련된 모든 정보를 객관적으로 평가하고 최종적으로 가장 합리적인 결론을 도출해내야 한다. 건설적 논쟁 수업에서 지식은 역동적이고, 사회적으로 구성되어 있으며, 지식을 적용하여 생성한 논증과 반대되는 논증을 모두 종합하고 수정하는 과정에서 가장 잘 학습된다고 가정한다(Johnson et al., 2000: 30). 교사는 각 모둠의 활동을 모니터링하며 논증 교환이 잘 일어날 수 있도록 촉진하고, 학생들이 서로 다른 인지적, 도덕적 갈등 사이에서 각 모둠이 제시한 주장과 반박, 입장채택에 귀를 기울이도록 돕는다. 이러한 구조화된 지적 갈등의 탐구를 통해 학생들은 어떤 주제든 적어도 두 개 이상의 측면이 확인될 수 있는 한 논쟁 문제에 직면했을 때에 합리적으로 판단할 수 있는 힘을 기르게 된다.

건설적인 논쟁 수업은 '협력학습'의 요소와 '구조화된 지적 갈등'의 요소를 바탕으로 교사들이 활용할 수 있는 실질적인 절차로 운영되고 있다. 건설적인 논쟁은 일반적으로 다섯 단계를 통해 수업에서 구체화 된다(Johnson et al., 2000: 31-32). 그리고 이러한 일반적인 절차를 바탕으로, 우리나라 도덕과 교육에서는 논쟁 수업 모형이라는 이름으로 총 6단계의 절차가 적용되고 있다(정창우 외, 2020: 144-145). 건설적 논쟁 수업의 일반적인 절차는 다음과 같다.

① 1단계: 자신의 입장에 대한 최선의 사례 준비하기

1단계에서 한 쌍의 구성원은 다음의 세 가지를 포함하여 자신이 부여받은 입장에 대한 가능한 최선의 사례를 찾는다. ①자신의 입장에 대해 조사하고, 관련된 모든 정보를 학습한다. ②논증의 생성 즉, 설득력 있는 주장과 논지를 뒷받침할 수 있는 근거, 논지와 일관된 논리적 결론을 도출해내는 정보 등을 구성한다. ③공정하고 완전한 논증의 평가를 받을 수 있도록 자신에게 부여된 역할을 효과적으로 옹호할 계획을 세운다. 학생들은 자신이 세운 초기 입장에 대해 높은 수준의 자신감을 갖는 경향이 있다. 이 과정은 보통 제한적이긴 하지만 학생들은 이 과정에서 가

능한 많은 정보와 경험, 관점을 고려하여 정리한 것에 기초하여 자신의 입장을 정리하기 때문이다.

② 2단계 : 자신의 입장에 대한 최선의 사례를 설득력 있게 제시하기

2단계에서 학생들은 자신이 맡은 입장에 대한 최상의 사례를 다른 두 그룹 구성원에게 제시한다. 학생들은 자신의 최초 주장과 그 근거를 다른 사람들에게 공개적으로 제시할 때, 더 많은 인지적 부담을 느끼고 더 높은 수준의 추론 전략을 사용하는 경향이 있다. 그들은 자신의 주장에 대해 강력하고, 설득력 있는 옹호자가 될 필요가 있기 때문이다. 이 과정에서 학생들은 상대방의 입장을 주의 깊게 듣고 이해하려고 노력해야 한다. 이해하지 못하는 부분은 메모를 하거나 질문을 통해 반드시 명확히 해야 한다.

③ 3단계 : 공개 토론에 참여하기

3단계에서 학생들은 자신의 입장을 위해 강력하고 설득력 있게 논쟁하면서 그 문제에 대한 개방적이고 자유로운 토론에 참여한다. 그들은 반대 입장(증거와 추론)을 비판적으로 분석하고, 주장을 뒷받침할 자료를 요청하며, 정보와 추론의 부족함을 지적함으로써 반대 입장을 반박한다. 그러면서 학생들은 반대 입장을 확실히 이해하여, '불의 재판(trial by fire)'을 치른다. 즉, 자신의 주장에 대한 진실성, 헌신, 용기를 평가하는 것이다. 그리고 학생들은 자신의 입장에 대한 공격에 대한 재반박을 시도해 본다. 그러나 이 과정에서는 학생들이 제시했던 초기 결론의 정확성이 불확실해지는 경향을 보이기 쉽다. 자신이 가지고 있던 기존 신념이 도전을 받게 되면서 개념적 갈등이나 인지 불균형 상태가 유발되기 때문이다.

④ 4단계 : 관점 교환하기

4단계에서 학생들은 관점을 바꿔보고 반대 입장에 가능한 최선의 사례를 제시해 본다. 학생들은 상대방의 입장을 진실하고 강력하게 제시하면

서 새로운 사실을 추가할 수 있다. 모든 학생들은 이 문제의 두 가지 관점을 동시에 이해하려고 노력한다. 이전 단계에서 유발된 불확실성, 개념적 갈등과 인지적 불균형은 인식론적 호기심을 유발하는 경향이 있다. 그 결과, 학생들은 더 많은 정보와 새로운 경험을 더 적극적으로 찾고, 불확실성을 해소하기 위해 보다 적절한 인지적 관점과 추론의 과정을 사용한다.

⑤ 5단계 : 의견을 종합하고 입장을 통합해보기

5단계에서 학생들은 자신이 맡았던 입장에 대한 옹호를 멈추고, 양측의 최선의 증거와 추론을 요약하고 새롭고 독특한 공동 입장으로 통합함으로써 그들 모두가 동의할 수 있는 종합적인 것을 찾기 위해 노력한다. 학생들은 타인의 관점과 추론을 이해하고 비판적으로 평가하는 과정에서 수용 여부를 결정하고, 그들의 인지적 관점과 추론을 수정함으로써 새롭고 재개념화되고 재구성된 결론을 도출하게 된다. 이를 통해 도출된 최종적인 결론은 초기 결론에 비해 질적으로 더 낫고 균형 잡힌 해결책의 양상을 보인다. 네 명의 학생들은 각자의 견해를 종합하여 모둠 보고서를 작성하는데, 여기에는 협력적 논증 교환의 과정에서 제시된 새로운 정보, 증거 및 합리적 이유가 포함된다. 그리고 나서 그들은 두 가지 입장에 대해 개별적으로 검증을 하고, 집단 토론이 자신이 최종 입장을 정하는 데 얼마나 영향이 있었는지 확인한다. 집단 구성원들은 열심히 참여해준 구성원들 서로에게 감사와 축하의 인사를 나눈다.

이를 바탕으로 한 도덕과 수업에서 사용되는 논쟁 수업 모형의 절차는 다음과 같다.

[표 5-2] 논쟁 수업 모형의 구조 (정창우 외, 2020: 144-145)

단계		내용
1	수업의 준비	논쟁 주제와 그 주장에 대해 상반되는 두 입장을 정하고, 각 입장에 대한 정보와 자료 준비 한 소집단을 4명으로 구성
2	소집단 내의 미니 소집단 구성	각 미니 소집단은 상반되는 입장을 선택하고 주어진 정보와 자료, 제한된 경험과 자신의 관점에 의해 입장을 정리한 후, 이를 뒷받침할 자료와 이론을 조사
3	미니 소집단의 주장 및 근거 발표	소집단 내에서 미니 소집단이 각각 자신들의 주장과 근거를 발표
4	미니 소집단별 토론	소집단 내에서 미니 소집단별로 토론 상대방의 입장을 분석, 비판하고 자신의 주장에 대한 상대방의 비판에 반박하는 과정에서 상대방의 주장을 경청, 인지적 분석, 개념 갈등과 불확실성의 체험
5	미니 소집단의 입장 교환	미니 소집단은 입장을 바꾸어 상대방이 주장하지 못했던 자료나 논리를 제시함
6	소집단 의견의 합의점 도출	두 미니 소집단은 이제까지 나타난 여러 관점과 주장을 하나의 주장으로 종합하여 합의점 도출

이와 같은 절차로 구성된 건설적 논쟁을 위한 수업 모형은 동기화된 추론을 개선하기 위한 협력적 논증 교환의 활동이 효과적으로 적용될 수 있는 토대가 된다. 먼저, 건설적 논쟁 수업은 협력적 논증 교환을 위한 적정한 집단의 규모를 안내해준다. 추론의 논증 이론에 따르면 추론이 우리의 기능으로 적응하게 된 맥락은 소규모 사회적 상호작용의 환경이다. 협력적 논증 교환이 효과적으로 수행되기 위해서는 불특정 다수와의 논증 교환 보다는 소집단 내에서의 실질적인 정보와 논증의 교환이 필요하다. 건설적 논쟁 수업 또한 토론 활동을 위한 집단의 구성을 최대 4인으로 제한하고, 4인 내에서도 2명이 한 팀이 되어 함께 논증을 구성하도록 한다. 두 사람이 지지해야 할 하나의 입장을 공유한 상태에서 집단적으로 추론을 발휘하게 되는 것이다. 물론 두 사람이 같은 입장을 지지하기 때문에 동기화된 추론이 작동할 수 있지만, 개인적, 독립적인 추론의 사용으로 인한 동기화된 추론의 경향과 확증 편향에 빠지는 위험을 사전

적으로 예방해주어 결과적으로 합리적인 논증을 구성하는데 기여할 수 있다. 이렇게 구성된 두 명의 공통된 입장은, 소집단 내에서 그들과 입장이 다른 또 하나의 팀과 협력적으로 논증을 교환하게 되는 것이다. 이로 인해 두 집단이 제시한 논증은 더욱 세심하게 평가될 수 있으며, 반박에 대한 수용 여부 또한 보다 면밀히 검토될 수 있다.

또한 건설적 논쟁을 위한 수업 모형은 협력적 논증 교환의 목적이 효과적으로 달성되기 위한 절차로서, '의견의 종합'을 포함해야 한다는 점을 시사한다. 논쟁 수업 모형은 마지막 절차에서 토론 과정에서 제시된 모든 근거와 합리적 이유를 종합적으로 고려하여 자신의 최종 입장을 결정하도록 유도한다. 이 과정에서 학생들은 자신이 가지고 있던 기존의 신념을 비판적으로 재점검해 보게 되며 기꺼이 수정할 기회를 얻는다. 동기화된 추론으로 인해 발생하는 여러 문제의 원인이 학생들이 가지고 있던 사회적 정체성을 유지하기 위함이었다는 점을 고려해 본다면, 이와 같은 절차가 명시적으로 제공되지 않는다면 토론의 과정과 관계없이 자신의 입장을 고수할 가능성이 크기 때문이다.

그러나 건설적 논쟁 수업에서는 교사가 학생들에게 임의의 입장을 '할당'해 준다는 점에서 동기화된 추론 개선을 위한 협력적 논증 교환 수업과는 다소 차이가 있다. 학생들은 자신의 실제 관점과 관계없이 주어진 입장을 지지하기 위해 정보를 탐색하고, 논증을 구성하며, 토론의 과정에서 그 입장을 옹호하는 역할을 맡는 것이다. 교사에 의한 입장 할당은 도덕적 직관을 강하게 불러일으킬 수 있는 윤리적 이슈에 대해 학생들로 하여금 정서적 거리두기의 기회를 제공하고 이를 통해 보다 합리적인 판단과 의사결정을 내릴 수 있도록 도울 것이다. 논쟁 수업의 후반부에서 자신과 반대되는 입장을 채택하는 절차가 있기 때문에 다양한 입장을 모두 체험해볼 수 있지만, 자신의 진정한 입장을 대변하지 않는다는 점에서 동기화된 추론이 작동될 가능성이 차단될 수 있다. 우리 삶의 실제 도덕 판단과 추론은 감정적 반응과 긴밀히 결합되어 있다. 그 감정은 기존의 도덕적 신념에 의해 촉발되며, 그 신념을 보호하기 위한 방향으로 우리의 추론은 동기화되기 쉽다. 앞서 살펴본 것처럼, 방향성을 제거하는

것은 협력적 논증 교환 수업의 목표에 부합하지 않는다. 학생들은 자신의 도덕적 신념과 감정으로 동기 부여될 때 진정성 있는 논증을 생성할 수 있기 때문이다. 따라서 동기화된 추론 개선을 목표로 하는 협력적 논증 교환 수업을 위해 건설적 논쟁 수업 모형을 참고한다면 교사는 이 점을 유의하여 수업을 설계해야 할 것이다.

3. 협력적 논증 교환 수업의 사례3: 하이트식 도덕적 통찰 모형

마지막으로 동기화된 추론의 개선을 위한 협력적 논증 교환 수업의 사례로 검토해 볼 도덕과 수업 모형은 하이트식 도덕적 통찰 모형이다(정창우, 석자춘, 2021). 도덕적 통찰 모형에 대한 검토는 앞선 두 가지 수업 모형(KMDD와 건설적 논쟁 수업 모형)과는 그 방향이 다르다. 이 두 모형은 그 자체로 협력적 논증 교환의 양상을 부분적으로 포함하고 있다. 또한 이 모형들은 우리나라 도덕과 수업에 적합한 방식으로 그 절차와 운영 원리가 타당화되어 학교 현장에서 사용되고 있다. 따라서 협력적 논증 교환 수업을 현장에서 적용하기 위한 방안으로 이 두 수업 모형을 고려해볼 수 있다. 다만, 각 수업 모형이 전제하는 이론적 기반과 기법의 차이로 인해 동기화된 추론 개선을 목적으로 하는 협력적 논증 교환 수업에 직접적으로 적용되는 데는 일정한 한계를 지닌다. 이와 달리 도덕적 통찰 모형에 대한 논의는 아직 시론적인 성격이 강하다. 그러나 이 모형이 근거하고 있는 도덕 심리학 이론이 도덕 판단에서 동기화된 추론의 역할에 주목한 하이트의 사회적 직관주의를 토대로 하고 있으며 수업의 목표 중 하나가 편향된 도덕 판단에 개선에 있다는 점에서, 이 수업 모형 또한 협력적 논증 교환 수업을 현장에 적용하기 위해 참고할 만하다. 다만, 도덕적 통찰 모형은 주로 인지행동치료나 마음챙김 명상을 근거 삼아 정립되었다는 점에서 동기화된 추론의 개선에 관한 기존의 개인적 차원에서 성찰 범주에 가깝다는 점에서 협력적 논증 교환의 양상이 전면에 드러나진 않는다. 따라서 본 검토를 통해 해당 수업 모형에서 추론의 사회적 기능이 고려될 수 있는 적절한 지점을 찾고자 한다. 이 작

업은 해당 수업 모형이 상대적으로 덜 고려하고 있는 '추론된 설득'의 양상이 협력적 논증 교환 활동을 통해 보완될 수 있음을 시사한다. 또한 성공적인 협력적 논증 교환 수업의 결론은 학생들이 자신의 기존 신념을 비판적으로 되돌아보고, 이를 기꺼이 수정해보는 것에 있다. 도덕적 통찰 모형은 학생들이 자신의 기존 신념을 효과적으로 성찰하고, 자신의 도덕적 정체성을 바탕으로 변화된 신념을 삶에서 실천하기 위한 교실 수업에서 사용할 수 있는 구체적인 교육 방법을 풍부하게 제시해준다. 따라서 협력적 논증 교환 수업이 목표로 하는 동기화된 추론의 개선과 이를 기반으로 도덕적 신념의 재정립은 도덕적 통찰 모형이 제시하는 여러 절차와 기법의 도움을 받을 수 있다. 이러한 관점에서 도덕적 통찰 수업 모형의 특징과 절차를 중심으로 협력적 논증 수업과의 관계를 검토해보고자 한다.

도덕적 수업 모형은 하이트의 도덕 심리학 연구를 바탕으로, 우리나라 도덕과 교육의 여건을 고려하여 개발된 수업 모형이다. 하이트의 연구는 앞선 논의에서 중요한 이론으로 다뤄졌던 것처럼, 도덕 판단에 있어 동기화된 추론에 주목하고 그 영향을 확인해주었다. 특히, 그는 도덕적 추론이 도덕 판단을 인과적으로 이끌어낸다는 이전의 합리주의적 모델을 비판하고, 사회적 직관주의 모델을 제시하며 우리 일상의 도덕 판단에 대한 일반적인 양상을 설명하였다. 특정한 도덕적 문제나 상황에 맞닥뜨렸을 때 도덕 판단은 대개 강력한 정서적 반응을 기반으로 하는 도덕적 직관에 의해 자동적이고 순간적으로 도출되며 의식적이고 언어화된 도덕적 추론은 사후적으로 나타나며 정당화의 기능을 수행한다는 것이다. 이때 사후 정당화로서 나타난 도덕적 추론이 바로 동기화된 추론의 양상 중 하나라는 것을 3장에서 분석하였다.

그 내용을 간략히 되짚어 보면 다음과 같다. 하이트(2001)에 따르면, 도덕적 추론은 종종 동기화된 추론의 경향을 보이는데, 이는 도덕 판단의 주체가 도덕 기반으로 대표되는 기존의 도덕적 가치와 신념에 부합하는 신념과 해석을 형성하는 경향에 의해 주도됨을 의미한다. 도덕 판단에서 이러한 동기화된 추론은 사람들이 진정으로 객관적인 결론에 도달

하기보다는 그들의 도덕적 직관을 정당화하기 위해 추론과 논증을 사용하는 도덕적 합리화의 형태를 초래할 수 있다는 것이다. 특히 하이트는 도덕 판단이 주로 도덕적 자극에 노출됨으로써 촉발되는 직관, 즉 신속한 자동 평가에 의해 주도된다고 주장한다. 이러한 도덕적 직관은 도덕적 규범과 가치에 대한 노출, 개인적 경험, 정서적 및 인지적 연관성의 발달을 포함한 문화적 및 진화적 요인에 의해 형성된다. 추론과 의식적인 숙고는 도덕적 판단을 형성하는 역할을 할 수 있지만, 하이트는 그것들이 종종 직관의 영향에 부차적이며 주로 동기화된 추론의 양상을 보인다고 설명한다.

이처럼 하이트의 이론은 동기화된 추론에 기인한 도덕 판단의 편향을 분석하는 좋은 도덕 심리학 모델이다. 또한 그의 이론은 윤리적 다원주의의 맥락에서, 나와 다른 도덕 기반을 가진 타인에 대한 열린 사고를 권장한다는 점에서 동기화된 추론을 개선하기 위한 하나의 아이디어를 제공한다. 특히, 하이트는 인간 발달에 대한 진화와 문화의 영향을 탐구함으로써 개인이 특정 상황에서 느끼는 혐오, 분노, 수치심과 같은 감정에 대한 설득력 있는 설명을 제공한다. 즉, 우리는 일상 생활에서 특정한 도덕적인 문제에 직면했을 때, 정서에 기반한 직관적인 반응을 통해 도덕적으로 정당한지 판단한다. 생물학, 진화심리학, 인지과학, 문화심리학 등의 분야를 도덕과 결합함으로써 하이트의 이론은 자동적인 도덕 판단에 대한 독특한 관점을 제공하고, 타인에 대한 도덕적 배타주의와 부당한 비판에 대해 경고한다. 그러나 하이트의 이론은 도덕교육을 위한 구체적인 방법을 제시하지 못하고 윤리적 성찰을 깊이 있게 탐구하지 못했다는 한계가 있다(정창우, 석자춘, 2021). 이에 정창우와 석자춘은 이러한 한계를 보완하기 위한 시론적인 성격으로 하이트가 주목한 '정서에 기반한 도덕적 직관'을 바탕으로 도덕적 통찰[53] 수업 모형을 제안했다. 특히

53) 정창우와 석자춘(2021)은 통찰에 대한 다양한 의미 분석을 바탕으로 '사물 혹은 존재의 참모습과 문제의 본질을 꿰뚫어 봄으로써 문제를 해결할 수 있는 능력'으로서 통찰의 개념이 도덕적 영역에 적용된다는 윌리엄 어빈(W. Irvine)의 관점을 수용하여, 도덕적 통찰을 '자신의 도덕적 신념이나 믿음이 가진 모순과 비일관성 등을 즉각적으로 혹은 오랜 내적 논쟁 끝에 꿰뚫어 보고 삶의 방식을 새롭게 바꿀 수 있는 도덕적 능력'의 의미로 사용하고 있다(정창우, 석자춘, 2021: 10-11).

이 모형은 앞서 동기화된 추론의 개선을 위한 기존의 방법으로 제시되었던 인지행동치료나 명상의 원리와 방법을 수업의 차원에서 반영했다는 점에 주목할만 하다. 수업 모형의 구조는 다음과 같다.

[표 5-3] 하이트식 도덕적 통찰 모형의 구조(정창우, 석자춘, 2021: 13)

1	생각 나누기 및 자료 제시	→	· 학습 목표 달성을 위해 필수적이면서도 인지 편향의 가능성이 높은 도덕적 이슈(issues)나 토픽(topics)에 대한 간략한 생각 나누기 · 선택한 도덕적 이슈/토픽에 관련되어 있고 정서적으로 강한 충격/자극을 줄 수 있는 자료(예: 영상, 이미지, 텍스트 자료 등) 제시
↓			
2	자신의 반응 관찰하고 표현하기	→	· 자신의 반응(예: 몸의 반응, 느낌, 직관적 판단) 관찰하기 (마음챙김 명상 기법이나 SEL의 자기인식 기술 활용) · 내면의 풍경에 대한 ‘관찰’ 내용을 표현하기
↓			
3	원인 분석 및 탐구 활동	→	· 자신의 반응에 대한 원인을 분석하고 공유하기 · 자료에 제시된 사태 관련 사실 탐색 및 유사 사례와 비교·분석하기
↓			
4	도덕적 신념의 한계 및 비일관성 검토	→	· 대상에 따른 도덕적 신념의 (비)일관성 검토 · 인지 편향 검토 질문지(〈표1〉 참조)나 심리요법 기반 질문지(〈표2〉 참조)을 활용한 자신의 도덕적 신념 검토
↓			
5	도덕적 통찰	→	· 자신이 가진 도덕적 신념의 한계 및 비일관성 자각 · 자신의 삶에 대한 윤리적 성찰 및 새로운 삶의 방식 채택
↓			
6	‘되고 싶은 자아’ 및 ‘도덕적 실천	→	· ‘되고 싶은 자신의 이미지’를 구체적으로 설정하기 · 앞으로 지켜나갈 ‘도덕적 실천 규정’을 1-2개 정하기

또한 연구자들은 이 수업 모형이 현재로서는 시론적 성격을 지니며, 이후 후속 연구를 통한 타당화 작업이 필요함을 밝히고 있다. 본고에서는 이 수업모형이 하이트의 도덕 심리학 이론을 도덕과 수업에 적용할 수 있도록 모형화 했다는 점에 주목하며, 이 수업에서 동기화된 추론의 문제를 개선하기 위한 협력적 논증 교환 활동이 활용될 수 있는 방안과 그 시사점을 중심으로 논하고자 한다.

	규정’ 정하기		
	↓		
7	확언 및 공유하기	→	· 앞 단계에서 설정한 것을 다른 사람들 앞에서 확언하고 공유하기 · 동료들에게 도움을 주고 싶다면, ‘추론된 설득’ 적용하기
	↓		
8	반복적 실천 및 모니터링하기	→	· 자신이 설정한 것을 일상생활에 반복적으로 연습 및 적용하기 · 실천 과정에서 자신이 제대로 하고 있는지 모니터링하기

먼저, 첫 번째 단계인 ‘생각 나누기 및 자료 제시’는 수업의 도입에 해당한다. 이 단계에서 교사는 학습목표의 달성을 위해 인지 편향에 빠지기 쉬운 도덕적 이슈(issue)나 토픽(topics)을 제시하고 학생들은 이슈에 대해 가볍게 생각을 나눈다. 해당 주제에 대해 간략히 이야기를 나눈 다음 교사는 해당 이슈나 토픽과 관련하여 학생들에게 정서적으로 강한 충격이나 자극을 불러 일으킬 수 있는 자료를 제시한다. 이를 통해 학생들은 순간적인 도덕적 정서를 느끼게 되고, 이는 해당 이슈에 대한 그들의 도덕 판단과 동기화된 추론의 기초가 될 것이다. 이때 나타나는 감정은 학생들의 관심을 편향시키고, 상황에 대한 우리의 인식을 바꾸며, 그들의 추론에 영향을 줌으로써 도덕 판단에 영향을 미칠 수 있다(Uhlmann et al., 2018). 예를 들어, 특정 행동에 대한 분노나 혐오감의 강한 감정은 우리가 그것을 더 도덕적으로 잘못되었다고 인식하게 하는 반면, 한 사람에 대한 연민이나 공감의 감정은 우리가 그들의 행동을 덜 도덕적으로 비난받을 가치가 있다고 인식하게 할 수 있다.

다음은 ‘원인 분석 및 탐구활동’ 단계로, 해당 도덕적 이슈나 토픽에 대한 자신의 최초 반응을 관찰하고 이를 언어적으로 표현하는 단계이다. 이 단계에서 학생들은 내면을 들여다보기 위해 마음챙김 명상 기법을 활용할 수 있다. 하이트는 여러 명상 기법이 있지만, 공통적으로 ‘분석적이지 않은 방법으로 주의를 집중하려 의식적으로 노력한다’는 점에 주목한다(Haidt, 2006/2010: 74). 일반적으로 명상을 위해서는 조용하고 편안한

상태에서 앉아 눈을 감고 심호흡을 하며 몸과 마음을 진정시킨 다음, 신체적 감각의 변화에 주의를 기울인다. 몸의 각 부분을 탐색하면서 떠오르는 감정이나 생각을 알아차리면서 이러한 감정이나 생각을 판단하거나 저항하지 말고 있는 그대로 바라본다. 교실 상황에서 학생들이 마음챙김 명상 시간을 갖기 어려울 수 있다. 이때는, 자신이 느낀 정서에 이름을 붙여보거나, 자신이 느끼고 있는 정서의 이유를 생각해보거나 그러한 정서가 느껴지게 된 상황적 맥락을 살펴보는 등의 방법을 포함한 사회정서학습에서 제시하는 자기인식의 기술 또한 적용해 볼 수도 있다. 이러한 방법으로 자신의 내면에서 느껴지는 정서들을 확인하고, 관찰한 결과를 표현함으로써 해당 도덕적 이슈나 토픽으로부터 촉발된 자신의 도덕적 직관을 파악할 수 있게 된다. 이는 동기화된 추론을 개선하는 데 있어서도 매우 중요한 지점이다. 도덕 판단에서 문제가 되는 동기화된 추론은 강한 '방향성'에 의해 동기화된 것이며, 이때의 방향성은 정서에 기반하기 때문이다. 학생들은 이 과정에서 자신의 추론이 편향된 원인을 파악할 수 있다. 물론, 이 부분은 개인적으로 수행되기 때문에 그 과정 자체가 자칫 왜곡될 여지가 있다. 그럼에도 교사는 학생들이 객관적으로 자신의 정서 상태를 관찰하고 표현할 수 있도록 명상이나 자기인식의 과정을 잘 안내해야 한다. 이를 통해 학생들은 자신이 순간적으로 느낀 정서를 왜곡하지 않고, 그로부터 도출된 도덕적 직관을 마주할 수 있게 될 것이다.

자신의 도덕적 직관에 관한 분석은 3단계인 '도덕적 신념의 한계 및 비일관성 검토' 단계에서 구체화된다. 이 단계에서는 해당 도덕적 이슈와 토픽에 대한 자신의 초기 반응의 원인을 분석하고, 다른 학생들과 공유한다. 교사는 학생들이 명상 등을 통해 마주하고 표현한 자신의 정서 상태가 어디에서 비롯되었는지 다층적인 차원에서 생각해보도록 안내한다. 도덕적 이슈나 토픽에 대한 우리의 정서 상태는 여러 가지 방법으로 도덕 판단에 있어 동기화된 추론에 영향을 미칠 수 있다. 우리가 도덕 판단을 내릴 때, 정서는 우리가 관심을 갖는 정보, 그 정보를 어떻게 해석하는지, 그리고 우리가 가장 설득력 있다고 생각하는 주장에 영향을 미

칠 수 있다. 이것은 우리가 기존의 신념과 정서를 뒷받침하는 방식으로 정보에 선택적으로 관심을 기울이고 해석하는 동기화된 추론에 참여하도록 이끌 수 있다(Ditto et al., 2009). 예를 들어, 만약 우리가 특정한 이슈에 대해 강한 분노의 감정을 가지고 있다면, 우리는 그 행동이 도덕적으로 잘못되었다는 우리의 견해를 뒷받침하는 정보에 주의를 기울이는 반면, 다른 것을 암시하는 정보는 무시할 가능성이 더 높을 수 있다. 마찬가지로, 만약 우리가 누군가에게 강한 연민의 감정을 가지고 있다면, 우리는 그들의 행동을 더 유리하게 해석하고 그들에 대한 우리의 도덕적 판단에 더 관대해질 가능성이 있다. 따라서 이 단계를 통해 감정의 이유를 분석한다면, 이는 동기화된 추론을 개선하기 위한 성찰의 기준으로 활용될 것이다. 특히, 추론의 사회적 기능을 고려한다면 이 과정에서 해당 도덕적 이슈에 대해 다른 관점과 감정을 가진 다른 사람들과 대화하는 것도 효과적이다. 이것은 여러분의 도덕적 판단에 대한 무의식적인 편견이나 감정적인 영향을 밝히는데 도움을 줄 수 있기 때문이다. 이는 교사가 해당 단계에서 경청과 상호 존중의 분위기 속에서 활동이 진행될 수 있도록 지도해야 하는 이유이다.

도덕적 통찰의 단계인 4단계에서는 이전에 분석한 원인을 바탕으로 자신의 도덕적 신념의 한계를 검토한다. 특히, 자신의 도덕적 신념이 상황에 따라 다르게 적용되는 것은 아닌지 일관성에 대해 주목할 필요가 있다. 도덕적 직관에 의해 도출된 도덕 판단은 동기화된 추론을 통해 사후적으로 정당화되는데, 이때 우리의 도덕적 신념은 비일관적으로 적용되는 경향이 크기 때문이다(Cusimano & Lombrozo, 2021). 특히, 도덕적 신념은 짧은 기간에 임시방편으로 형성된 것이 아니라, 삶의 오랜 기간 속에서 누적된 경험의 산물이자 그 이전에 자신이 속한 사회, 문화적 환경의 맥락에서 진화적으로 자리 잡은 도덕 기반으로 구성된 일종의 도덕 매트릭스이다. 그렇기 때문에 자신의 도덕적 신념의 한계를 스스로 확인하기란 매우 어려운 일이다. 도덕적 통찰 모형은 이를 위해 자기 보고식인지 편향 검토 질문지나 심리요법 기반 질문지를 개발하여 제공하고 있다(정창우, 석자춘, 2021: 13-14). 특히, 심리요법 기반 질문지는 하이트

가 마음챙김 명상과 함께 직관을 교정하기 위한 효과적인 방법으로 제시한 인지행동치료에서 유래한다. 인지행동치료는 앞서 동기화된 추론의 기존 개선방안에서도 분석한 것 바와 같이, 왜곡된 자동적 사고를 검토하여 스스로 인식하도록 하는 기법이다. 추론의 사회적 기능을 고려한다면, 이와 같은 질문지가 효과적으로 기능하기 위해서는 학생들 각자 개별적으로 수행하기 보다는, 짝이나 소그룹을 이뤄 함께 질문하고 답해가는 과정이 포함되어야 한다. 자기 보고식 질문지는 자기기만과 가장(假裝)의 문제가 있기 때문에 해당 질문을 서로에게 던지고, 그에 대한 답을 교환하면서 자신의 기존 신념이 가지고 있는 한계와 비일관성을 보다 객관적으로 검증할 수 있을 것이다.

이와 같은 검토는, 학생들의 도덕적 통찰을 가능케 한다. 도덕적 통찰 수업 모형의 절차에 따르면 5단계 즉, '되고 싶은 자아' 및 '도덕적 실천 규정' 정하기 단계에 이르러 학생들은 자신이 가지고 있던 기존의 도덕적 신념의 한계를 확인하고, 비일관성을 스스로 깨닫게 된다. 동기화된 추론의 관점에서 본다면, 자신의 도덕 판단이 잘못된 방향성에 의해 동기화된 추론에 기반했다는 사실을 마주하는 순간이다. 이를 통해 학생들은 자신의 삶에 대한 윤리적 성찰 및 기존의 도덕적 신념의 수정을 통해 새로운 삶의 방식을 채택할 기회를 갖게 된다. 이 모형에서 참고하고 있는 도덕적 통찰의 개념을 제시한 어빈에 따르면, 도덕적 통찰에 도달하기 위한 두 가지 방법으로 간디형 통찰과 클라크슨형 통찰이 있다. 마하트마 간디는 남아프리카에서 인도인들이 직면한 차별과 억압을 목격한 후 개인적인 도덕적 통찰을 경험했다. 이 깨달음은 그가 비폭력 저항의 철학을 채택하고 인도 독립 운동의 지도자가 되도록 이끌었다(Irvine, 2015: 79-80). 한편, 토마스 클라크슨은 끊임없는 연구와 내적인 고민을 바탕으로 이성에 의해 촉발된 도덕적 통찰을 바탕으로 영국에서 흑인 인종차별 폐지 운동의 선구자가 되었다(Irvine, 2015: 80-83). 즉 간디형 통찰은 감정에 의해 촉발되어 즉각적으로 획득되는 방식이고, 클라크슨형 통찰은 이성에 의해 촉발되어 비교적 오랜 내적 논쟁 끝에 획득되는 방식이다(정창우, 석자춘, 2021: 27). 도덕적 통찰 수업 모형은 간디형 통찰

의 방식에 초점을 맞춘 방식이다. 그러나 클라크슨형 통찰의 획득 방식의 가능성을 고려한다면, 이 과정은 추론의 사회적 기능을 고려한 협력적 논증 교환 활동을 통해 보완될 여지가 있다.

도덕적 통찰은 자신이 추구하는 가치나 도덕적 신념이 비일관적이라는 사실을 깨닫는 것에서 시작한다. 그러나 그러한 통찰에 도달하는 과정은 동기화된 추론으로 인해 기존의 신념과 편견에 근거하여 자신의 의견을 합리화하는 경향에 의해 방해될 수 있다. 따라서 동기화된 추론을 극복하기 위해 자신의 신념에 대한 개방적이고 정직한 검토와 다른 관점을 가진 다른 사람들과의 협력적 논증 교환이 도움이 될 수 있다. 협력적 논증 교환은 개인이 자신의 기존 전제에 도전하고, 다른 사람들의 관점에 대한 이해를 확장하며, 궁극적으로 자신의 도덕적 신념에 대한 더 명확한 이해에 도달할 수 있도록 하기 때문이다. 이를 위해 가장 우선적으로 준비되어야 하는 것은 자신의 신념과 의견을 표현할 수 있는 안전하고 존중되는 환경이다. 자신이 가지고 있던 기존의 도덕적 신념에 대해 공개적으로 주장하고, 그것을 객관적으로 검토하고 평가받기 위해서이다. 그리고 그러한 논증을 교환하는 집단의 구성원들은 각자 가진 의견이 달라야 더욱 효과적이다. 환경이 확립되면 참가자들이 당면한 주제에 대한 신념과 의견을 공유하고, 서로의 관점에 귀를 기울이고, 다른 사람들의 주장에 대해 특정 동기나 의도를 가정하지 않도록 주의해야 한다. 토론이 진행됨에 따라 참가자들이 서로 다른 신념과 가치를 가지고 있다는 것이 명백해지고, 그들 자신의 관점에 대한 의견 불일치와 도전에 직면할 수 있다. 이러한 의견 차이가 자신의 도덕적 신념에 대한 귀중한 통찰력을 제공할 수 있다는 것을 인식하면서 열린 마음과 다른 사람들의 관점에 참여하려는 의지로 접근하는 것이 중요하다. 가령, 클라크슨은 "자신의 의지에 반하여 다른 사람들의 노예를 만드는 것이 옳은가?"를 주제로 에세이를 쓰기 위해 연구하는 과정에서 도덕적 혼란을 느꼈다. 연구 자료를 조사하는 과정에서 접한 책에서 노예 무역의 잔혹성을 사실적으로 묘사한 장면을 마주했기 때문이다. 만약 그 내용이 사실이라면 노예제도는 분명히 도덕적으로 혐오스러운 것이었다(Irvine, 2015: 92). 그

럼에도 불구하고 그의 주변 사람들, 심지어 교육을 받고 교회에 다니는 사람들도 노예제도를 용인했을 뿐만 아니라 그것으로부터 이익을 얻기 위한 조치를 취했다. 그들과 함께 있었다면 클라크슨은 도덕적 통찰을 얻지 못했을 것이다. 노예제도를 옹호하는 신념을 지닌 사람들로만 구성된 집단이기 때문이다. 비록 클라크슨은 혼자만의 시간을 통해 노예제도의 비도덕성에 대해 통찰을 얻었지만, 이를 수업 환경에 적용한다면 학생 개개인이 가진 특성의 차이에 따라 도덕적 통찰을 얻지 못하고 오히려 기존의 신념이 강화되는 방식으로 결론 내려질 수 있다. 따라서 지지적인 환경 속에서 자신과 입장이 다른 다양한 사람들 사이에서 이루어지는 협력적 논증 교환 활동이 포함된다면, 도덕적 통찰의 여부가 개인적인 능력만으로 결정되진 않을 것이다.

이전 단계를 통해 얻은 도덕적 통찰을 바탕으로 '확언 및 공유하기' 단계에서는 자신의 새로운 도덕적 삶을 위한 구체적인 방법을 세우게 된다. 이를 위해 '되고 싶은 자아'를 구체적으로 생각해보고, 앞으로 지켜나갈 '도덕적 실천 규정'을 정해보는 활동이 포함된다. 도덕적 통찰 수업 모형의 개발자들은 이 절차의 이론적 근거를 행동 윤리학에 두고 있음을 밝혔다(정창우, 석자춘, 2021: 27-28). 이들은 '도덕적 자아에 관여하는 것이 비윤리적 행동에 대한 유혹에 저항하고 자신이 추구하는 높은 가치나 윤리적 기준(ethical standards)에 헌신하도록 돕는다'고 주장한 행동 윤리학의 관점을 바탕으로 도덕적 자아의 이미지를 강조하는 것의 정당성을 제시한다. 또한 동기화된 추론의 문제가 발생하는 이유가 사회적 정체성의 유지하고자 하는 욕구에 의한 것이라는 점을 고려한다면, 자신이 되고 싶은 도덕적 자아의 이미지를 세움으로써 우리의 추론을 도덕적으로 바람직한 방향으로 이끌 수 있을 것이다.

끝으로 '반복적 실천 및 모니터링하기' 단계에서는 지금까지 절차를 통해 확인한 자신의 윤리적 기준과 되고 싶은 자신의 모습을 통합적으로 고려하여, 관련 상황이 발생하였을 때마다 그 기준을 바탕으로 구체적인 실천 계획을 세우고 계획대로 실천할 수 있도록 모니터링 해야한다. 이와 같은 통찰이 삶에서 전이되기 위해서는 습관화가 동반되어야 한다.

이를 위해서는 일상적인 연습과 엄청난 양의 반복이 필요하다(정창우, 석자춘, 2021: 15). 이와 같은 절차는 동기화된 추론 개선에 효과적으로 기여할 것이다.

한편, 도덕적 통찰 수업 모형을 제안한 연구자들은 '확언 및 공유하기' 단계에서 하이트가 주장한 '추론된 설득'의 기제를 고려할 수 있다고 제안한다. 그러나 이들은 '필요하다면'이라는 조건을 붙임으로써 도덕적 통찰에 있어 추론된 설득의 기제를 활용하는 제안에 있어서 소극적인 입장을 취한다. 그 이유는 두 가지로 추측 가능하다. 하나는 도덕적 통찰 모형이 전제하는 이론적 근거에서 찾을 수 있다. 연구자들은 해당 수업 모형의 절차적 순서와 관련한 이론적 근거를 '인지행동치료'와 '명상' 등에서 찾는다. 특히 3세대 인지행동치료는 마음챙김을 기반으로 하는 사적 경험으로서의 '알아차림(awareness)'를 통찰의 출발점으로 삼는다. 이는 동기화된 추론 개선을 위해 고려될 수 있는 개인적 차원의 성찰 측면과 관련이 깊다. 반면, 한 사람의 직관적 판단에 따른 추론이 다른 사람의 직관에 영향을 미침으로써 그 사람의 판단에 영향을 주는 과정인 '추론된 설득'은 추론의 사회적 사용에 가깝다. 따라서 해당 수업 모형에서는 추론된 설득 기제에 대해 상대적으로 덜 고려하게 되었을 것이다.

다른 하나는 도덕적 통찰 수업 모형이 전제하는 '통찰'의 유형에서 찾을 수 있다. 연구자들은 어빈이 제시한 간디형 통찰과 클라크슨형 통찰 중 본 수업 모형을 구상함에 있어 간디형 통찰에 초점을 맞추고 있다. 어빈(Irvine)은 간디형 통찰과 클라크슨형 통찰의 차이를 우리 내면에서 통찰이 나타나는 방식에서 찾는다.

> *간디형 통찰은 격정에 의해서 촉발된다. 우리는 뭔가가 잘못되었음을 본능적으로 느끼고, 따라서 뭔가가 잘못되었다고 결론 짓는다. 이 과정에서 이성의 역할은 없거나 미미하다. 물론 사후에 우리의 결론에 대해서 그럴싸한 이유를 제시하기 위해서 이성이 작동할 수는 있겠지만 말이다. 반면에 클라크슨형 통찰은 이성에 의해서 촉발된다. 이성은 우리의 믿음들에 비일관성이 있음을 알아채고, 그것을 해소하기 위해서 애쓴다(Irvine, 2015: 92).*

이처럼 도덕적 통찰 수업 모형은 간디형 통찰을 지향하기 때문에 타인에 의한 언어적 설득을 그 내용으로 하는 '추론된 설득' 기제에 상대적으로 소극적을 입장을 취하게 된다.

'추론된 설득'은 직관에 의해 도달된 자신의 도덕 판단을 정당화하기 위한 추론이 타인의 직관에 영향을 미치는 기제이다. 이는 자신의 판단을 언어적으로 정당화하고 타인에게 논증의 형태로 제시하는 과정으로 이를 통해 한 사람의 판단은 다른 사람이 자신의 입장을 수정하는 계기가 마련될 수 있다. 추론된 설득은 개인의 직관을 바꾸는 데 타인의 추론에 의한 논증이 개입될 여지를 보여준다. 도덕적 통찰 모형의 연구자들 또한 이 점을 고려하여, 교실 상황에서 학생들은 상호 간의 이해도가 전제되기 때문에 타인의 정서적 반응을 고려하면서 추론된 설득을 효과적으로 활용할 가능성에 대해 열어두고 있다(정창우, 석자춘, 2021: 27).

협력적 논증 교환은 도덕적 통찰 모형의 '확언 및 공유하기' 단계를 더욱 효과적으로 수행하는 데 활용될 수 있다. 협력적 논증 교환 수업은 '추론된 설득' 기제를 포함하여 추론의 사회적 사용을 바탕으로 학생들이 초기 직관을 수정하는 기회를 제공할 수 있기 때문이다. 협력적 논증 교환은 서로 다른 입장 사이에서 서로를 설득하기 위해 논증을 생성하고 평가하는 과정이 포함된다. 학생들은 자신이 재정립한 도덕적 신념을 동료 학생들에게 공언하고, 동료 학생들은 이를 객관적으로 평가해 줌으로써 더 나은 신념과 의사결정에 도움을 줄 수 있다. 협력적 논증 교환을 이 단계에 적용함으로써 학생들은 이전 단계들에서 제시된 다양한 개인적 성찰과 숙고 방법을 통해 자신의 초기 직관을 들여다보고 수정한 입장을 바탕으로 반복적인 연습과 실천으로 나아가기에 앞서 최종적으로 함께 고민하고 검토의 기회를 갖게 된다. 협력적 논증 교환 절차를 명시적으로 고려함으로써 학생들은 자신의 신념을 공유하고, 수정하는 과정에서 동료 학생들과 상호 긍정적인 역할을 경험하게 된다. 이처럼 동기화된 추론을 개선하기 위한 교육적 조치는 개인적 차원의 성찰과 숙고, 이를 통한 도덕적 통찰과 함께 추론의 사회적 기능을 고려한 협력적 논증 교환 활동이라는 두 축의 순환적 상호작용을 통해 목표에 더욱 가까

워질 수 있을 것이다.

제 6 장 결 론

안개가 자욱한 어두운 밤, 헤드 랜턴 빛에만 의존해 울창한 숲길을 나아가는 여행자를 생각해 보자. 랜턴은 여행자의 시선이 머무는 곳을 환하게 비춘다. 그러나 그의 시선이 향하는 부분을 제외한 나머지 숲 전체는 그림자로 뒤덮일 것이다. 랜턴은 여행자가 보고 싶은 것을 더 크게 보여주지만, 올바른 길을 가기 위해 그가 꼭 보아야 하는 것을 놓치게도 한다. 어둠 속 숲을 헤쳐나가는 여행자의 머리 위에서 빛을 비추는 랜턴이 오히려 그의 시야를 제한하는 것처럼, 우리의 도덕 판단은 때로 윤리 문제에 대한 우리의 이해를 편협하게 제한하기도 한다. 여행자가 향하고 싶은 곳만을 강하게 비추는 랜턴처럼, 동기화된 추론은 우리가 내리고 싶은 도덕적 결론에 도달하기 위해 세상을 왜곡하여 이해하기도 한다.

도덕 판단에서 동기화된 추론의 존재는 도덕적 추론의 공정성을 훼손한다. 그 결과, 우리는 자신의 신념에 부합하는 정보와 근거는 과대평가하고, 신념을 위협하는 정보와 근거는 무시하거나 과소평가한다. 이와 같이 도덕 판단에서 작동하는 동기화된 추론의 심리 기제는 다양한 문제를 야기한다. 대표적으로 동기화된 추론은 개인과 집단의 비도덕적 행위를 정당화하는 심리적 기제로 작용하기도 하며, 도덕적, 정치적 불일치의 양극화를 심화시켜 집단적 혐오와 사회적 분열을 가속화하는 데 기여한다.

도덕 판단에서 동기화된 추론이 발생하는 원인은 크게 두 가지로 분석되었다. 하나는 자신의 신념을 보호하고자 하는 마음의 경향이다. 도덕적 사태에 대한 우리의 초기 직관은 강한 도덕적 감정으로 인해 검토되지 않은 자신의 신념을 보호하고자 하는 방향으로 추론을 동기화한다는 것이다. 다른 하나는 자신이 범하는 편향에 대한 편향된 인식이다. 자신의 편향에는 관대하지만 타인의 오류에는 엄격한 이중적 잣대를 적용하는 편향의 사각지대 현상 때문에 도덕 판단에서 동기화된 추론의 문제가 발생한다. 그러나 우리의 도덕적 신념은 수정 가능하며, 우리는 편향의 사

각지대로부터 벗어날 수 있다.

기존의 인지 편향을 개선하고자 제안된 도덕적 메타인지 함양을 위한 방안들은 동기화된 추론의 문제를 개선하기 위해 활용될 수 있는 좋은 아이디어다. 구체적으로 인지행동치료기법이나 비판적 사고를 위한 교육을 참고할 수 있다. 이와 같이 인지 편향을 개선하고자 하는 여러 시도들은 정서로부터 거리두기를 시도함으로써 자신의 판단을 비판적으로 점검하고, 스스로를 성찰해 본다는 점에서 의미가 있지만 동기화된 추론의 개선을 개인적 차원에 국한한다는 점, 추론으로부터 동기적 요인을 제거하고자 하는 시도가 현실적으로 매우 어렵다는 점 등의 측면에서 보완될 필요가 있다. 특히, 집단 내 도덕적 평판을 유지하고 긍정적 사회적 정체성을 유지하고자 하는 인간의 기본적 동기에서부터 동기화된 추론이 비롯된다는 점을 고려할 때, 이 문제를 개선하기 위해서는 추론의 사회적 기능에 대한 고려도 병행될 필요가 있다. 나아가 추론에 관한 최근 심리학 연구는 추론의 본질적 기능이 사회적인 것에 있음을 시사한다.

진화 심리학적 관점에서 추론의 기능을 규명하고자 한 추론의 논증 이론은 추론이 사회적 상호작용의 대화적 맥락에서 논증을 생성하고 평가하는 과정에서 적응한 진화의 산물임을 보여준다. 우리는 의사소통 상황에서 자신의 입장을 타인에게 설득하고자 적절한 논증을 생성해낸다. 한편, 타인이 생성한 논증의 신뢰 여부를 결정하기 위해서는 논증을 객관적으로 평가하는 능력도 필요하다. 이 과정에서 발휘되는 것이 '이유를 대는 논리적 활동' 즉 추론이다. 추론에 대한 진화적 설명은 인간 종의 차원에서는 추론의 기원과 기능에 대해 설명해줄 수 있지만, 개체발생적 관점에서 발달의 측면은 간과할 수 있다는 한계를 지닌다. 이에 발달심리학자들은 추론의 기능이 논리적 이해와 함께 발달해가며, 높은 수준의 추론 능력은 타인과의 상호작용을 통한 공동의 문제를 해결하는 기능을 수행한다는 사실을 보여준다.

이와 같은 추론의 사회적 기능은 협력적 논증 교환이라는 구체적 활동을 통해 동기화된 추론의 개선에 기여한다. 협력적 논증 교환은 대화의 맥락에서 주제에 대한 공유된 이해를 구축하기 위해 서로 다른 관점에서

증거와 추론을 제시하는 두 명 이상의 사람 사이의 사회적 상호작용이다. 협력적 논증 교환은 윤리적 문제에 대한 자신의 진정성 있는 입장을 바탕으로 자신과 상반된 입장을 가진 대화 상대와 서로의 논증에 대해 공개적으로 평가하는 작업을 반드시 포함하는 의견 교환이라는 틀 안에서 이루어진다. 이를 통해 우리는 비도덕적 행위를 정당화하려는 마음의 경향을 사전에 예방하여 도덕적 위선의 문제를 최소화하고, 지적 겸손을 갖추게 하여 합당한 도덕적, 정치적 불일치의 가능성을 탐색할 수 있도록 해준다.

동기화된 추론을 개선하기 위해 도덕 수업은 협력적 논증 교환의 활동을 활용할 수 있어야 한다. 협력적 논증 교환 수업은 토론 수업의 모습으로 구현될 것이다. 그러나 협력적 논증 교환 수업은 수업의 목표와 양상, 교사의 역할 측면에서 일반적인 토론 수업과 구별되는 특징을 지닌다. 먼저 수업의 목표는 동기화된 추론의 개선에 있다. 일상의 도덕 판단에서 정서적 반응이 추론의 과정과 분리하여 접근하는 것이 현실적으로 불가능하기 때문에 추론에 영향을 주는 동기 요소를 적절히 설정함으로써 목표에 도달할 수 있을 것이다. 따라서 이를 위해 수업은 정확성 목표를 높이고, 학생들이 적절한 방향성을 찾도록 동기 부여해야 한다. 교실 수업에서 협력적 논증 교환은 '인지적 작업의 대립적 분업'의 양상으로 나타날 것이다. 이는 소크라테스의 논박술을 모태로 하지만, 논증을 생성하는 사람과 평가하는 사람의 역할이 수업의 과정 내내 순환한다는 특징을 갖는다. 끝으로 협력적 논증 교환 수업이 효과적으로 운영되기 위해 교사는 추론의 사회적 기능이 발휘될 수 있도록 집단을 적절히 구성할 수 있어야 한다. 특히, 진정한 반대 의견을 포함한 이질적 견해의 공존이 필수적으로 요구되는데, 교실 상황에 따라 이 조건이 달성되기 어려울 수 있다. 이때 교사는, 스스로 혹은 학생에게 역할을 부여함으로써 의도적으로 반대 의견 또는 다른 의견을 제시할 수 있는 착한 악마의 옹호자를 설정할 수 있어야 한다.

마지막으로 협력적 논증 교환 수업이 적절하게 활용되기 위해 기존 도덕과 수업 모형의 사례를 검토해보았다. 콜버그의 도덕 딜레마 토론 수

업에서 ‘반대 의견’에 대한 평가를 명시적으로 포함하고 있는 콘스탄츠 딜레마 토론 수업 모형은 협력적 논증 교환 수업을 직접 적용할 수 있는 좋은 무대가 된다. 교사는 적절한 딜레마 이야기를 통해 학생들을 정서적으로 동기부여하며 수업을 시작하고, 학생들은 입장이 같은 동료와 논증의 생성 작업을 함께 수행한다. 이 과정에서 나타날 수 있는 편향된 도덕 판단은 전체 토론에서 반대 의견에 부딪히면서 조정될 수 있다. 반대되는 입장 중 가장 매력적인 논증을 평가하는 단계를 통해 학생들은 신념의 다양성을 확인하고 경우에 따라 자신의 기존 의견을 수정해 보는 기회를 갖기도 할 것이다. 또한 KMDD 모형은 협력적 논증 교환이 효과적으로 수행되기 위한 교사의 역할을 시사해준다. 교사는 수업 목표에 대한 명확한 인식을 바탕으로, 학생들의 정서적 반응에 영향을 주어 도덕 판단과 논증 교환을 촉진시킬 수 있도록 적절한 윤리적 이슈를 제시할 수 있어야 한다. 나아가 협력적 논증 교환이 학생들의 주체적인 참여로 수행될 수 있도록 교사의 개입은 최소화되어야 한다.

이와 유사한 맥락에서, 윤리적 논쟁을 수업에서 적절히 다루기 위해 고안된 건설적 논쟁 수업 모형은 협력적 논증 교환이 효과적으로 수행되기 위한 절차와 세부 지침에 시사점을 제공한다. ‘협력’의 과정과 ‘인지적 대립을 유발하는 구조화된 갈등’이 적절히 고려될 때 협력적 논증 교환 수업은 성공적으로 운영될 수 있다. 또한 건설적 논쟁 수업 모형이 제시하는 집단 구성은 협력적 논증 교환이 적정 인원으로 구성된 소규모 집단 내에서 활성화될 수 있음을 시사한다.

끝으로 아직 시론적인 성격이지만, 하이트의 도덕 심리학 이론을 최초로 수업 모형의 절차로 구체화했다는 점에서 하이트식 도덕적 통찰 모형도 협력적 논증 교환 수업의 적용을 위한 검토의 대상이 되었다. 도덕적 통찰 모형이 지향하는 바는 개인적 차원의 성찰과 숙고를 통한 자신의 초기 직관을 관찰하고, 수정할 기회를 갖는 것에 있다. 이때, 도덕적 통찰 모형의 기존 절차에 협력적 논증 교환의 단계를 명시적으로 추가함으로써 추론된 설득의 기제를 더욱 활성화할 수 있다. 이는 개인적 차원의 성찰과 사회적 차원의 협력적 논증 교환이라는 두 방안의 상호보완적인

선순환 관계를 가능하게 한다. 이러한 노력을 통해 동기화된 추론 문제가 효과적으로 개선될 수 있을 것이다.

협력적 논증 교환 수업은 동기화된 추론이 발생하는 맥락과 도덕 판단이 갖는 사회적 성격, 학제적 접근이 밝혀낸 추론의 사회적 기능에 관한 논의를 바탕으로 도출했다는 점을 고려했을 때, 효과적인 도덕교육적 개선방안으로 제안될 수 있다. 다만, 본 연구에서는 개선방안이 교실 수준의 수업으로 제한되었다는 한계를 갖는다. 도덕 판단의 사회적 성격 등을 고려한다면, 사회적, 문화적, 제도적 차원 등의 고려 또한 동기화된 추론 개선을 위한 주요한 전략이 될 수 있기 때문이다. 이후의 연구에서는 사회적 관습과 분위기 등의 개선을 통해 동기화된 추론의 문제들을 해결할 수 있는 거시적 차원의 개선방안에 대해서도 밝힐 필요가 있다.

어두운 숲길을 헤쳐나가는 여행자의 이야기로 돌아와 보자. 랜턴은 시선이 향하는 곳만 제한적으로 비추기 때문에 여행자가 볼 수 있는 더 넓은 숲을 보지 못하게 한다. 랜턴을 끄는 것도 하나의 근본적인 방법이 될 수 있다. 랜턴 빛이 사라지면 시간이 흘러 우리의 눈이 어둠에 적응한다면 오히려 숲 전체를 보게 될 수도 있다. 우리의 도덕 판단에서 정서적 개입을 완전히 제거할 수 있다면, 그래서 도덕 판단에 아무런 사심이 들어가지 않고 어떠한 동기도 작동하지 않는다면, 동기화된 추론은 일어나지 않을 것이다. 따라서 이로 인한 문제 또한 원천적으로 차단된다. 그러나 이는 수준 높은 차원의 인지적 기술과 고도의 훈련을 요구한다. 어쩌면 추론에서 동기를 제거하지 못하는 대다수 사람은 그 앞에서 좌절할지도 모른다.

다행히도 어두운 숲길을 헤쳐나가는 여행자는 혼자가 아니다. 숲길에 들어선 모든 여행자는 저마다 헤드 랜턴을 끼고 앞으로 나아가고자 한다. 랜턴의 종류도, 빛의 양도, 시선이 향하는 방향도 모두 다르지만 여행자들은 어두운 숲길을 나아가고자 하는 공통의 목표를 가지고 있다. 이들이 향하는 서로 다른 불빛은 각자의 시선이 닿지 않는 곳을 비춰준다. 서로가 서로의 빛에 의존한다면 이내 자기 앞에 놓인 전체 숲의 윤곽이 드러날 수 있다. 숲 전체가 훤히 보이지 않더라도, 적어도 이전보다

는 여행자가 원하는 방향을 더 정확하게 안내해줄 것이다. 우리는 저마다 산적한 윤리적 문제들 사이에서 보다 나은 삶을 위해 앞으로 나아가고자 한다. 도덕적 삶을 위해 자신 앞에 놓인 문제를 해석하고, 해결하는 것은 개개인 각자의 몫이다. 그러나 그 과정은 비록 다른 입장일지라도, 서로를 신뢰하는 동료들과 함께할 때 더욱 명료해질 수 있다. 신뢰를 바탕으로 논증을 교환하고 비판적으로 평가하는 과정을 통해 서로 협력한다면 자신의 기존 신념에서는 편협하게 보였던 문제의 전체를 보게 되고, 복잡하고 어려운 윤리적 문제들을 적절히 대응해낼 수 있을 것이다.

우리는 종종 편향된 도덕 판단을 도출한다. 우리의 도덕적 추론은 도덕적이지 않을 때도 있다. 그러나 신뢰하는 사람과 함께하는 추론은 나의 도덕적이지 않은 결론을 객관적으로 되돌아보게 하고, 더욱 적절한 방향으로 수정할 기회를 제공해주며, 다양한 신념과 판단을 고려하는 계기가 된다. 도덕적 삶으로 나아가는 여정에 함께하는 동료가 있음을, 그리고 서로가 서로를 도와 우리의 앞길을 함께 비출 수 있음을 상기한다면, 때로는 어둡고 어려워 보이는 그 험난한 여정도 이전보다 한층 밝아질 것이고 우리는 앞으로 함께 나아갈 힘을 얻을 것이다.

참 고 문 헌

1. 국내문헌

1) 단행본

김구(2011), 『사회과학 연구조사 방법론의 이해』, 서울 : 비앤엠북스.

박전규(1995), 『아리스토텔레스의 실천적 지혜』, 경기: 서광사

정창우(2013), 『도덕과 교육의 이론과 쟁점』, 서울: 울력.

정창우(2020a), 『도덕교육의 새로운 해법(2판)』, 경기: 교육과학사.

정창우(2020b), 『도덕교육 연구 및 실천의 방향과 과제』, 경기: 교육과학사.

정창우(2022), 『변혁적 도덕 역량 증진을 위한 도덕교육론』, 경기: 교육과학사.

정창우 외(2020), 『도덕과 교재 연구 및 지도법』, 경기: 교육과학사.

2) 논문

강민겸(2019), 「역량기반 도덕과 교육과정 실현을 위한 이론 연구 - 린트(G. Lind)의 도덕교육론을 중심으로」, 서울대학교 윤리교육과 석사학위 논문.

강민겸, 박균열(2019), 「도덕적 딜레마를 활용한 중학교 청렴교육 콘텐츠」, 『인문사회 21』, 10(6), 1245-1258.

강인구(2015), 「인성교육에서 도덕적 추론과 직관의 통합」, 『한국교육문제연구』, 33(4), 73-97.

김동창(2015), 「인지과학의 두 마음 이론과 도덕교육」, 『도덕윤리과교육연구』, 48, 171-189.

김미라, 민영(2014), 「지지 후보와 추론 동기가 유권자의 선택적 노출과 교차노출에 미치는 영향 : 선거에서의 인지부조화를 중심으로」. 『한국방송학보』, 28(2), 7-49.

김윤경, 박균열(2019), 「논쟁수업에서 도덕과 교사의 역할과 발문에 관한 연구」, 『도덕윤리과교육』, 65, 163-188.

김하연(2020), 「건설적 논쟁의 도덕교육적 활용 방안 연구」, 『초등도덕교육』,

68, 1-31.
문경호(2017), 「도덕적 의사결정에서 직관과 추론의 역할」, 『윤리교육연구』, 44, 85-115.
박균열(2016). 「중등 도덕과에서 도덕적 딜레마를 활용한 통일교육 교수학습방법」. 『통일정책연구』, 25(2), 51-88.
박균열(2017), 「도덕적 역량 검사도구(MCT)의 구성원리를 활용한 통일 · 안보 의식측정도구 개발 기초 연구」. 『국방연구』, 60(3), 67-92.
박은진(2006), 「토픽맵에 기초한 철학 고전 텍스트들의 체계적 분석 연구와 디지털 철학 지식지도 구축(베이컨, 『신기관』)」, 『철학사상』 별책, 제7권, 제12호.
박장호(2012), 「신경과학의 도덕적 직관 이해와 도덕교육」. 『윤리교육연구』, 29, 137-164.
석자춘(2020), 「게오르그 린트(Georg Lind)의 도덕적 역량 및 측정 방법, 『도덕윤리과교육』, 66, 197-225.
석자춘, 정창우(2017), 「나바에츠 통합적 윤리교육모형을 활용한 도덕교육 개선 방향」, 『도덕윤리과교육』, 54, 23-52.
신세인, 이준기, 하민수(2018), 「과학적 사고의 걸림돌 동기기반추론 -과학기술 분야 학문후속세대들의 사례를 중심으로-」. 『한국과학교육학회지』, 38(5), 635-647.
신원동(2022), 「사회인지 이론의 도덕교육적 함의: 반두라의 도덕적 이탈을 중심으로」. 『초등도덕교육』, 76, 61-86.
양해성(2016), 「하이트의 사회적 직관주의 모델의 도덕 교육적 함의」, 『도덕윤리과교육연구』, 53, 145-170.
양해성(2020), 「민주시민 역량으로서 반성적 성찰의 교육 방안 연구- 2015 개정 도덕과 교육과정을 중심으로 -」, 『도덕윤리과교육』, 67, 207-226.
이기연(2023), 「인격특질에 대한 상황주의 비판의 재조명과 도덕교육적 함의」. 『도덕윤리과교육』, 78, 107-139.
이인태(2016), 「또래 압력이 인지적 왜곡에 미치는 영향에 대한 개인내적 중재

요인으로서 도덕적 정체성」, 『윤리연구』, 1(111), 305-334.
이인태(2022), 「부도덕한 행동에 대한 합리화 기제로서 인지 왜곡에 대한 연구」, 『도덕윤리과교육』, 74, 213-239.
이인태, 박균열(2019), 「도덕적 사고 능력 향상을 위한 교수학습방법 및 평가 도구로서 KMDD와 MCT의 활용 가능성에 대한 연구」, 『도덕윤리과교육연구』, 65, 1-33.
장동익(2004), 「토픽맵에 기초한 철학 고전 텍스트들의 체계적 분석 연구와 디지털 철학 지식지도 구축(흄, 『인성론』)」, 『철학사상』 별책, 제7권, 제12호.
정창우(2011), 「도덕 심리학 연구의 최근 동향과 도덕교육적 함의」, 『초등도덕교육』, 37, 95-130.
정창우(2012), 「나바에츠(D. Narvaez) 도덕발달 이론의 특성과 도덕교육적 함의」, 『도덕윤리과교육』, 37, 329-361
정창우, 김하연(2020), 「학문적 요구를 반영한 도덕과 교육의 개선 방향」, 『도덕윤리과교육』, 67, 227-260.
정창우, 김하연(2021), 「논쟁 문제 수업에 대한 이해와 도덕 교사의 역할 」, 『도덕윤리과교육』, 71, 245-271.
정창우, 석자춘(2021),「도덕교육 방법 개선을 위한 하이트 도덕심리이론의 적용 방안」, 『도덕윤리과교육』, 72, 1-27.
추병완(2020a), 「도덕 이탈의 도덕교육적 함의」. 『도덕윤리과교육』, 67, 75-98.
추병완(2020b), 「도덕 정당화의 심리 기제 분석」. 『초등도덕교육』, 67, 87-118.
추병완(2020c), 「도덕교육에서 직관과 추론의 관계 분석: 하이트와 나바이즈를 중심으로」, 『초등도덕교육』, 68, 83-117.
추병완(2020d), 「논쟁 이슈 교수·학습에서 교사의 역할」. 『倫理硏究』, 1(129), 23-50.
추병완(2021), 「도덕 교과 교수·학습에서 행동 윤리학의 적용 방안」, 『윤리연구』, 1(133), 1-26.
추병완(2022a), 「행동 윤리학에 대한 비판적 평가」. 『도덕윤리과교육』, 76, 161-195.

추병완(2022b), 「도덕 판단에서 정서의 역할」. 『윤리연구』, 1(138), 23-47.
하민수(2016), 「합리적 문제해결을 저해하는 인지 편향과 과학교육을 통한 탈인지 편향 방법 탐색」. 『한국과학교육학회지』, 36(6), 935-946.

2. 국외문헌

1) 단행본 및 번역서

Allport, G. W. (1937). *Personality: A Psychological Interpretation.* New York, NY: Holt, Rinehart & Winston.

American Psychological Association. (2010). *Publication manual of the American Psychological Association.* Washington, DC: Author.

Andriessen, J., Baker, M. & Suthers, Dan. (2003). *Arguing to Learn*, Dordrecht, Netherland: Springer.

Aronson, E., Wilson, T.D. and Akert, R.M. (2010). *Social Psychology.* 8th, 박권생 등 역(2015), 『사회 심리학』, 서울: 시그마프레스.

Bacon, F. (1620/1952). *Novum Organum from Great Books of the Western World Vol. 28*, Encyclopaedia Britannica, 진석용 역(2001), 『신기관: 자연의 해석과 인간의 자연 지배에 관한 잠언』, 서울: 한길사.

Baggini, J. (2016). *Edge of Reason,* 박현주 역(2017), 『위기의 이성』, 경기: 아르테.

Bandura, A. (2016). *Moral disengagement,* 김의철, 이상미, 박선영, 박은실 역(2018), 『도덕성의 분리와 비도덕적 행위의 정당화』, 경기: 교육과학사.

Bazerman, M. H. & Tenbrunsel, E. (2011). *Blind Spots: Why we fail to what's right and what to do about it.* Princeton University Press., 김영욱, 김희라 역(2014), 『이기적 윤리』, 서울: 커뮤니케이션북스.

Beauchamp, T. L., and Childress, J. F. (1994). *Principles of Biomedical Ethics* 4th ed., Oxford, University Press, New York.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders.* International

Universities Press.

Brooks, J. G., & Brooks, M. G. (1993). *In Search of Understanding: The Case for Constructivist Classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Callard, A. (2018), *Aspiration: The Agency of Becoming.* NY: Oxford University press.

Dennett, D. (1995), *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* NY: Simon & Schuster Paperbacks.

Dorkins, R. (1989). *The Selfish Gene, (2nd edition),* 홍영남, 이상임 역(2010), 『이기적 유전자』, 경기: 을유문화사.

Festinger l. (1957). *A Theory of Cognitive Dissonance,* 김창대 역(2016), 『인지부조화 이론』, 경기: 나남.

Festinger l. Riecken, H. & Schachter S. (1956). *When Prophecy Fails*, 김승진 역(2020), 『예언이 끝났을 때』, 경기: 이후.

Fiske, S. T., & Taylor, S. E. (2013). *Social Cognition: From Brains to Culture.* Thousand Oaks. CA: Sage.

Galef, J. (2021). *The Scout Mindset,* 이주만 역(2021), 『스카우트 마인드셋』, 서울: 와이즈베리.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge: Cambridge University Press.

Ginsburg, H.P. & Opper, S. (1988). *Piaget's Theory of Intellectual Development,* 김정민 역(2006), 『피아제의 인지발달이론』, 서울: 학지사.

Gould, S. J. (1996). *The Mismeasure of Man*, 김동광 역(2003),『인간에 대한 오해』, 서울: 사회평론.

Haidt, J. (2006). *The happiness hypothesis,* 권오열 역(2010), 『행복의 가설』, 경기: 물푸레.

Haidt, J. (2012). *The Righteous Mind,* 왕수민 역(2014), 『바른 마음』, 경기: 웅진씽크빅.

Hart, C. (1998). *Doing a literature review: Releasing the social science research*

imagination. Thousand Oaks, CA: Sage.

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life.* Free Press.

Hume, D. (1980). *A Treatise of Human Nature,* 이준호 역(1998), 『인간 본성에 관한 논고』, 서울: 서광사.

Irvine, W. B. (2015), *AHA! : The Moments of Insight that Shape Our World*, 전대호 역(2015), 『아하! 세상을 바꾸는 통찰의 순간들』, 서울: 까치글방.

Janis, I. L. (1982). *Group think*. Boston: Houghton Mifflin.

James, W. (1890). *The Principles of Psychology Vol 1.*, 정양은 역(2005), 『심리학의 원리 1』, 서울: 아카넷.

James, W. (1968). *Psychology: The Briefer Course*, 정명진 역(2014), 『윌리엄 제임스가 한 권으로 간추린 심리학의 원리』, 서울: 아카넷.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis.* Minneapolis: University of Minnesota.

Johnson-Laird, P.N. & Wason.P.C. (1977). *Thinking: Readings in Cognitive Science*. Cambridge, UK: Cambridge University Press.

Kahneman, D., Slovic, P. & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases, Cambridge University Press, 이영애 역(2001), 『불확실한 상황에서의 판단』, 서울: 아카넷.

Kahneman, D. (2011). *Thinking Fast and Slow*, 이진원 역(2012), 『생각에 관한 생각』, 경기: 김영사.

Kohlberg, L. (1981). *The Philosophy of Moral Development,* 김민남, 김봉소, 진미숙 역(2000), 『도덕발달의 철학』, 서울 : 교육과학사.

Kohlberg, L. (1984). *The Psychology of Moral Development,* 김민남, 진미숙 역(2001), 『도덕발달의 심리학』, 서울 : 교육과학사.

Kohlberg, L., Levine C. H. & Hewer, A. (1983). *Moral Stages: A Current*

Formulation and a Response to Critics, 문용린 역(2000), 『콜버그의 도덕성 발달 이론』, 서울 : 아카넷.

Kunda, Z. (1999). Social cognition: Making sense of people. The MIT Press.

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Laughlin, P. R. (2011). *Group problem solving*. Princeton University Press.

Leslie, I. (2021). *Conflicted,* 엄윤미 역(2021), 『다른 의견』, 서울: 어크로스.

Lind, G. (2016). *How to Teach Morality: Promoting Deliberation and Discussion, Reducing Violence and Deceit*. 박균열, 정창우 역(2017), 『도덕적 민주적 역량, 어떻게 기를 것인가』, 서울: 양서각.

Mead, G.H. (1934). *Mind, Self, and Society from the Standpoint of a Social Behaviorist.* University of Chicago Press: Chicago.

Mercier, H. & Sperber, Dan. (2017). *The Enigma of Reason,* 최호영 역(2018), 『이성의 진화』, 서울: 생각연구소.

Mooney, C. (2012). *The Republican Brain,* 이지연 역(2012), 『똑똑한 바보들』, 경기: 동녘사이언스.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality, and identity* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Moshman, D. (2020). *Reasoning, Argumentation and Deliberative Democracy*, NY: Routledge.

Narvaez, D. (2014). *Neurobiology and the development of human morality: evolution, culture, and wisdom.* New York: W. W. Norton & Company.

Nemeth, C. (2018). *In Defense of Troublemakers: The Power of Dissent in Life and Business,* 신솔잎 역(2020), 『반대의 놀라운 힘』, 경기: 청림출판.

Nucci, L. P. (2001). *Education in the moral domain.* Cambridge University Press.

Page, S. E. (2007). *The Difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Perelman, C. & Olbrechts-Tyteca, L. (1958). *The New Rhetoric: A Treatise on Argumentation.* University of Notre Dame Press.

Petty, R.E. & Cacioppo, J.T. (1986). *Communication and Persuasion.* NY: Springer.

Piaget, J. (1928). *Judgment and reasoning in the child.* Harcourt, Brace.

Piaget, J. (1965). *The moral judgement of the child,* 송명자, 류웅달, 김상윤 역(2000), 『아동의 도덕판단』, 울산: 울산대학교 출판부.

Pinker, S. (2021), 김한영 역,『지금, 다시 계몽』, 서울: 사이언스북스.

Popkin, R. (2003). *The History of Scepticism.* NY: Oxford Univ. Press.

Rawls, J. (2005), *Political Liberalism,* 장동진 역(2016), 『정치적 자유주의』, 경기: 동명사.

Reeve, J. (2015). *Understanding Motivation an Emotion, 6th,* 김아영 등 역(2018), 『동기와 정서의 이해』, 서울: 박학사.

Reimer, J. Paolitto, D. R. & Hersh, R. H. (1990). *Promoting Moral Growth: From Piaget to Kohlberg*, Longman.

Siegel, D. J. (2010). *The mindful therapist: A clinician's guide to mindsight and neural integration*. W W Norton & Co.

Smith, A. (1789). *The Theory of Moral Sentiments,* 박세일, 민경국 역(2009), 『도덕 감정론』, 서울: 비봉출판사

Stanovich, K. (2004). *The robot's rebellion.* IL: Chicago University Press.

Stanovich, K. (2021). *The Bias That Devide Us*, 김홍옥 역(2022), 『내 편 편향』, 서울: 바다출판사

Stebbins, R. (2001). *Exploratory Research in the Social Sciences.* Thousand Oaks, CA: SAGE.

Sunstein, C. R. & Hastie, R. (2015). *Wiser,* Harvard Business Press, 이시은 역(2015), 『와이저』, 서울: 위즈덤하우스

Toulmin, S, E. (1983). *The Uses of Argument.* Cambridge University press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*, NY: Cambridge University Press.

van Eemeren, F. H. & Grootendorst, R. (2003). *A Systematic Theory of Argumentation.* Cambridge University Press.

Vygotsky, L.S. (1978). *Mind in Society: The development of higher psychological processes*. MA: Harvard University Press.

Wilson, E.O. (1975). *Sociobology : The New Synthesis*, 이병훈, 박시룡 역(1992), 『사회 생물학』, 서울: 민음사.

2) 논문

Ancell, A.J. (2017). Public unreason- Essays on political disagreement, Dissertation for the degree of Doctor, Department of Philosophy Duke University.

Aquino, K., Freeman, D., Reed, A. II, Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology,* 97(1), 123-141.

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

Aronson, E. (1999). Dissonance, hypocrisy, and the self-concept. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 103-126), American Psychological Association.

Aronson, Z. H., & Reilly, R. R. (2006). Personality Validity: The Role of Schemas and Motivated Reasoning, *International Journal of Selection and Assessment*, 14(4), 372-380.

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

Bandura, A. (2002). Selective Moral Disengagement in the Exercise of Moral Agency, *Journal of Moral Education*, 31(2), 101-119.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure

to opposing views on social media can increase political polarization, *the National Academy of Sciences of the United States of America*, 115(37), 9216-9221.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2), 230-244.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation, *Psychological Bulletin,* 117(3), 497-529.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes, *Personality and Social Psychology Bulletin*, 20(1), 3-19.

Bayes, R., & Druckman, J. N. (2021). Motivated reasoning and climate change, *Current Opinion in Behavioral Sciences,* 42, 27-35.

Blair, J.A. & Johnson, R.H. (1987). Argumentation as dialectical, *Argumentation*, 1(1), 41-56.

Blatt, M.M., & Kohlberg, L. (1975). The Effects of Classroom Moral Discussion upon Children's Level of Moral Judgment, *Journal of Moral Education*, 4, 129-161.

Blum-Kulka, S., & Snow, C. E. (2004). Introduction: The Potential of Peer Talk, *Discourse Studies,* 6(3), 291-306.

Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion, *Political Behavior,* 36(2), 235-262.

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning, *Judgment and Decision Making*, 3(2), 121-139.

Carlo, G., McGinley, M., Hayes, R., Batenhorst, C., & Wilkinson, J. (2007). Parenting styles or practices? Parenting, sympathy, and prosocial behaviors among adolescents, *The Journal of genetic psychology*, 168(2), 147-176.

Carter, J. A., & McKenna, R. (2020). Skepticism motivated: On the skeptical import

of motivated reasoning, *Canadian Journal of Philosophy,* 50(6), 702-718.

Castañer, X. & Oliveira, N. (2020). Collaboration, Coordination, and Cooperation Among Organizations: Establishing the Distinctive Meanings of These Terms Through a Systematic Literature Review, *Journal of Management,* 46(6), 965-1001.

Cialdini, R. B.; Goldstein, N. J. (2004). Social Influence: Compliance and Conformity, *Annual Review of Psychology,* 55(1), 591-621.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and Systematic Information Processing within and beyond the Persuasion Context, In J. S. Uleman, & J. A. Bargh (Eds.), *Unintended Thought,* 212-252.

Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs, *Journal of Personality and Social Psychology*, 85(5), 808-822.

Collins, E. & Batson, B. (2007). Moral Hypocrsy, in Baumeister, R. F. & Kathleen, D., *Encyclopedia of Social Psychology, (589-590),* CA: SAGE.

Cushman, F., Young, L., & Hauser, C.(2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm, *Psychological Science,* 17(12), 1082-1089

Cusimano, C., & Lombrozo, T. (2020). Morality justifies motivated reasoning. 960-966. Paper presented at 42nd Annual Meeting of the Cognitive Science Society: *Developing a Mind: Learning in Humans, Animals, and Machines,* CogSci 2020, Virtual, Online.

Cusimano, C., & Lombrozo, T. (2021). Morality justifies motivated reasoning in the folk ethics of belief,*Cognition,* 209, 1-24.

Davis, J. R. (2013). Improving Students' Critical Thinking and Classroom Engagement by Playing the Devil's Advocate, *Theory, Research, and Action in Urban Education,* 2(1), online.

Dearden, R. (1981). Controversial issues and the curriculum, *Journal of Curriculum Studies,* 13(1), 37-44.

Dieckmann, N. F., Gregory, R., Peters, E., & Hartman, R. (2017). Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning, *Risk Analysis,* 37(3), 471-486.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions, *Journal of Personality and Social Psychology,* 63(4), 568-584.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral judgment and decision making* (pp. 307-338). Elsevier Academic Press.

Druckman, J.N. & McGrath, M.C. (2019). The evidence for motivated reasoning in climate change preference formation, *Nature Clim Change* 9, 111-119.

Drummond, C., & Fischhoff, B. (2017). Development and validation of the Scientific Reasoning Scale, *Journal of Behavioral Decision Making*, 30(1), 26-38.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments, *Journal of Personality and Social Psychology,* 71(1), 5-24.

Egan, L., Santos, L. & Bloom, P. (2007). The Origins of Cognitive Dissonance: Evidence From Children and Monkeys. *Psychological science*. 18. 978-83.

Ehrich, K. R., & Irwin, J. R. (2005). Willful Ignorance in the Request for Product Attribute Information, *Journal of Marketing Research*, 42(3), 266-277.

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017, *Personality and Social Psychology Review,* 23(4), 332-366.

Ellis, J. (2022). Motivated reasoning and the ethics of belief, *Philosophy Compass.* 17. 1-11.

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning, *Journal of Economic Perspectives,* 30(3), 133-140

Evans, J. (2002). Logic and human reasoning: An assessment of the deduction paradigm, *Psychological Bulletin,* 128(6), 978-996.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance, *Journal of Abnormal and Social Psychology,* 58, 203-211.

Fishkin, J.S., & Luskin, R.C. (1999). Bringing Deliberation to the Democratic Dialogue. M. Mccombs & A. Reynolds, *The poll with a human face (p*p. 3-36). NJ: Lawrence Erlbaum Associates.

Fiske, S. T. (2010). Interpersonal stratification: Status, power, and subordination. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), Handbook of social psychology (pp. 941-982). NJ: John Wiley & Sons, Inc..

Frederick, S. (2005). Cognitive reflection and decision making, *Journal of Economic Perspectives,* 19(4), 25-42.

Freund, T, Kruglanski, A. W, & Shpitzajzen, A. (1985). The freezing and unfreezing of impressional primacy: Effects of the need for structure and the fear of invalidity, *Personality and Social Psychology Bulletin,* 11, 479-487.

Gino, F, Michael I. N, & Roberto A. W. (2016). Motivated Bayesians: Feeling Moral While Acting Egoistically, *Journal of Economic Perspectives,* 30 (3): 189-212.

Gose, M. (2009). When Socratic Dialogue is Flagging: Questions and Strategies for Engaging Students, *College Teaching*, 57(1), 45-49.

Graça, J., Calheiros, M. & Oliveira, A. (2016). Situating moral disengagement: Motivated reasoning in meat consumption and substitution, *Personality and Individual Differences.* 90. 353-364.

Graham, J., Haidt, J, Koleva, S., Koleva, S., Motyl, M. Iyer, R., Wojcik, S. & Ditto, P., (2012). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In P. Devine, & A. Plant(eds), *Advances in Experimental Social Psychology*, Vol. 47(pp. 55-130), Burlington: Academic Press.

Hadit, J. (2001). The emotional dog and its rational tail: A social intuitionist

approach to moral judgment, *Psychological Review*, 108(4), 814-834.

Hadit, J. (2007). The new synthesis in moral psychology, *Science,* 316(5827), 998-1002.

Hadit, J. Björklund, F, & Murphy. S. (2000). Moral Dumbfounding: When Intuition Finds No Reason, Unpublished manuscript.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize, *Social Justice Research,* 20, 98-116.

Hardy, S. A., & Carlo, G. (2011). Moral identity: What is it, how does it develop, and is it linked to moral action? *Child Development Perspectives,* 5(3), 212-218.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information, *Psychological bulletin,* 135(4), 555-588.

Haselton M. G, Nettle D, Andrews P. W. (2016). The evolution of cognitive bias. In Buss D. M (ed.). The Handbook of Evolutionary Psychology, 김한영 역 (2019), 『진화심리학 핸드북 2』, 565-596.

Hindriks, F. (2015). How Does Reasoning Contribute to Moral Judgment? Dumbfounding and Disengagement, *Ethical Theory and Moral Practice* 18 (2):237-250.

Hornsey, M., Harris, E., Bain, P. & Fielding, Kelly. (2016). Meta-analyses of the determinants and outcomes of belief in climate change, *Nature Climate Change*. 6. 622-626.

Huemer, M. (2016). Why People Are Irrational About Politics, in Anomaly, J. et al. Philosophy, *Politics, and Economics: An Anthology (pp* 457-467), NY: Oxford University Press.

Iwry, J. (2021). Open to Debate: Reducing Polarization by Approaching Political Argument as Negotiation, *Negotiation Journal*, 37: 361-391.

Jacobs, G. M. (2015). Collaborative Learning or Cooperative Learning? The Name Is Not Important; Flexibility Is, *Beyond Words,* 3(1), 33-52.

Johnson, D. W., & Johnson, R. T. (2009). Energizing Learning: The Instructional Power of Conflict, *Educational Researcher*, 38(1), 37-51.

Jurchiş, R., & Opre, A. (2018). A dual process approach to current issues from cognitive behavior therapies: An integrative perspective, *International Journal of Cognitive Therapy,* 11(4), 374-409.

Kahan, D. M. (2013). Ideology, Motivated Reasoning, and Cognitive Reflection, *Judgment and Decision Making*, 8(4), 407-424.

Kahan, D. M. (2015). Ideology, The Politically Motivated Reasoning Paradigm, *Emerging Trends in Social & Behavioral Sciences,* 1-24.

Kahan, D. M.(2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition, *Cultural Cognition Project Working Paper* Series No. 164, Yale Law School, Public Law Research Paper No. 605.

Kahan, D., Braman, D., Gastil, J., Slovic, P. & Mertz, C. (2007). Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception, *Journal of Empirical Legal Studies*. 4, 465 - 505.

Kahan, D. M, Peters, E. Wittlin, M., Slovic, P., Ouellette, L, L. Braman, D. & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks, *Nature Climate Change*, 2, 732-735

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality, *American Psychology,* 58(9):697-720.

Killen, M. (1991). Social and moral development in early childhood. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development*, Vol. 1. Theory; Vol. 2. Research; Vol. 3. Application (pp. 115-138).

Killen, M. & Dahl, A. (2021). Moral Reasoning Enables Developmental and

Societal Change, *Perspectives on Psychological Science*, 16(6), 1209-1225.

Kirschner, P. A. (1996). A vision of higher education: An educational technology approach . InJ.H. Coetzee & T.G. Smith (Eds.) *Proceedings International Conference on Education and Change*, Pretoria: University of South Africa.

Kohlberg L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, *Handbook of socialization theory and research,* Chicago : Rand McNally.

Kohlberg, L. (1971). "From is to ought: How to commit the naturalistic fallacy and get away with it inthe study of moral development" . In Mischel, T., ed. *Cognitive Development and Psychology.* New York: Academic Press. 151-235.

Kohlberg L. (1975). Moral education for a society in moral transition, *Educational Leadership,* 46-54.

Kriegeskorte, N., Simmons, W., Bellgowan, P. & Baker, C. (2009). Circular analysis in systems neuroscience: The dangers of double dipping, *Nature neuroscience,* 12, 535-540.

Krettenauer, T. (2021). Moral sciences and the role of education, J*ournal of Moral Education,* 50(10), 77-91.

Kruglanski, A. W. (1980). Lay epistemo-logic—process and contents: Another look at attribution theory, *Psychological Review,* 87(1), 70-87.

Kruglanski, A. W., & Klar, Y. (1987). A view from a bridge: Synthesizing the consistency and attribution paradigms from a lay epistemic perspective, *European Journal of Social Psychology,* 17(2), 211-241.

Kuhn, D. (2000). Metacognitive development. Current Directions, *Psychological Science,* 9(5), 178-181.

Kuhn, D. (2002). What is scientific thinking, and how does it develop? In Goswami, U. (Ed.). (2008). *Blackwell Handbook of Childhood Cognitive Development*. Hoboken, NJ: John Wiley & Sons.

Kuhn, D. (2015). Thinking Together and Alone, *Educational Researcher*, 44(1), 46-53.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories, *Journal of Personality and Social Psychology,* 53(4), 636-647.

Kunda Z. (1990). The case for motivated reasoning, *Psychological bulletin,* 108(3), 480-498.

Kunda, Z. & Nisbett, R. E. (1986). The psychometrics of everyday life, *Cognitive Psychology,* 18(2), 195-224.

Kunda, Z. & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition, *Psychological Inquiry*, 10(1), 12-22.

Landemore, H. & Page, S. E. (2015). Deliberation and disagreement: Problem solving, prediction, and positive dissensus, *Politics, Philosophy and Economics*. 14 (3):229-254.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks, *Journal of Experimental Social Psychology*, 22(3), 177-189.

Leeper, T. J., & Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation, *Political Psychology,* 35(1), 129-156.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin,* 85(5), 1030-1051.

Lind, G. (2012). Moral competence and democratic ways of life. In: W. G. Weber, M. Thoma, A. Ostendorf, & Lynne Chisholm, eds., *Democratic competences and social practices in organizations*, (pp. 62-85). Wiesbaden: Spinger VS.

Lodge, M. & Taber, C. (2000). Three steps toward a theory of motivated political reasoning. Lupia, A. McCubbins, M & Popkin, S(2000). *Elements of Reason: Cognition, Choice and Bounds of Rationality.* Cambridge University Press.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence, *Journal of Personality and Social Psychology*, 37, 2098-2109.

Lord, C. O, Lepper, M. R.,& Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment, *Journal of Personality and Social Psychology,* 47, 1231-1243.

Loxton, D. (2021). How cognitive dissonance distorts reality? *Skeptic*, vol. 28. 51-79.

Luo, Yu & Zhao, Jiaying. (2019). Motivated Attention in Climate Change Perception and Action, *Frontiers in Psychology*, 10. eCollection.

Mata, A., Ferreira, M., & Sherman, S. J. (2013). Flexibility in motivated reasoning: Strategic shifts of reasoning modes in covariation judgment. *Social Cognition,* 31(4), 465-481.

Mata, A., Vaz, A. & Mendonça, B. (2022). Deliberate ignorance in moral dilemmas: Protecting judgment from conflicting information, *Journal of Economic Psychology*, 90, 1-12.

May, J. (2018). The Limits of Emotion in Moral Judgment. *The Many Moral Rationalisms*, Karen Jones & Francois Schroeter (eds.), Oxford: Oxford University Press (2018), 286-306.

May, J. (2021). Moral Rationalism on the Brain, *Mind & Language* 38(1): 237-255.

May, J. & Kumar, V.(2018). Moral Reasoning and Emotion, *The Routledge Handbook of Moral Epistemology*, eds. Karen Jones, Mark Timmons, and Aaron Zimmerman, Routledge, 139-156.

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001-2010. *The Sociological Quarterly,* 52(2), 155-194.

Mercier, H. (2011a). What Good is Moral Reasoning?. *Mind & Society*, 10(2), 131-148.

Mercier, H. (2011b). Reasoning serves argumentation in children, *Cognitive Development,* 26(3), 177-191.

Mercier, H. (2013). The Function of Reasoning: Argumentative and Pragmatic Alternatives, *Thinking and Reasoning,* 19(3-4), 488-494.

Mercier, H. (2016). The argumentative theory: predictions and empirical evidence, *Trends in Cognitive Sciences.* 20(9), 305-324

Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation, *Political Psychology,* 33(2), 243-258.

Mercier, H., Castelain, T., Hamid, N., & Marín-Picado, B. (2017). *The power of moral arguments.* In J.-F. Bonnefon & B. Trémolière (Eds.), Moral inferences (pp. 107-122). Routledge/Taylor & Francis Group.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is There a Universal Positivity Bias in Attributions? A Meta-Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias. *Psychological Bulletin*, 130(5), 711-747.

Moody-Adams, M. (1999). The idea of moral progress, *Metaphilosophy*, 30(30), 168-185.

Molden, D. C., & Higgins, E. T. (2005). Motivated Thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 295-317). Cambridge University Press.

Moshman, D. (1995). The construction of moral rationality, *Human Development,* 38(4-5), 265-281.

Moshman, D. (2004a). False moral identity: self-serving denial in the maintenance of moral self-conceptions, In D. K. Lapsley, & D. Narvaez. (Eds.). *Moral development, self, and Identity* (83-110). Mahwah, NJ: Erlbaum.

Moshman, D. (2004b). From inference to reasoning: The construction of rationality, *Thinking & Reasoning,* 10(2), 221-239.

Moshman, D. (2007). Us and them: Identity and genocide, *Identity: An International Journal of Theory and Research*, 7(2), 115-135.

Moshman, D. (2012). The morality of social identity, *Journal of Applied Developmental Psychology,* 33(6), 320-322.

Moshman, D. (2018). Essay Review of The Enigma of Reason of Hugo Mecrier and Dan Sperber, *Human Development*, 61, 60-64.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality, *Thinking and Reasoning,* 4(3), 231-248.

Momsen, K. & Schneider, S. O. (2022). Motivated Reasoning, Information Avoidance, and Default Bias. *MPI Collective Goods Discussion Paper,* No. 2022/3.

Narvaez, D. (2005). The Neo-Kohlbergian Tradition and Beyond: Schema, Expertise, and Character, Pop-Edward, C., Carlo, G., *Moral Motivation through the Lifespan,* 51(N. E.: University of Nebraska Press), 121-156.

Narvaez, D. & Lapsley, D. K. (2009). Moral identity, moral functioning, and the development of moral character, *Psychology of Learning and Motivation,* 50, 237-274.

Nemeth, C.J., Connell, J.B., Rogers, J.D. and Brown, K.S. (2001a), Improving Decision Making by Means of Dissent. *Journal of Applied Social Psychology*, 31, 48-58.

Nemeth, Charlan & Brown, Keith & Rogers, John. (2001b). *Devil's advocate versus authentic dissent: Stimulating quantity and quality*. European Journal of Social Psychology. 31. 707-720.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.

Nowak, E. (2013). Democracy begins in the mind : Developing democratic personality. In: E. Nowak, ed., *Experimental ethics*, 141-155. Munster: LIT-Verlag.

Nucci, L. P. (1984). Evaluating teachers as social agents: Students' ratings of domain appropriate and domain inappropriate teacher responses to transgressions. *American Educational Research Journal,* 21(2), 367-378.

Nucci, L. (2016). Recovering the role of reasoning in moral education to address inequity and social justice, *Journal of Moral Education,* 45(3), 291-307.

Oyserman, D., & Lee, S. W. S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin,* 134(2), 311-342.

Paharia, N., Vohs, K. D., & Deshpandé, R. (2013). Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior and Human Decision Processes*, 121(1), 81-88.

Pasek J. (2018). It's not my consensus: Motivated reasoning and the sources of scientific illiteracy. *Public understanding of science,* 27(7), 787-806.

Paternotte, C. (2014). Constraints on joint action. In M. Gallotti & J. Michael (Eds.), *Perspectives on social ontology and social cognition* (pp. 103-123). Springer Science + Business Media.

Payne, J. W, Bettman, J. R, & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 14, 534-552.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making,* 10(6), 549-563.

Pennycook, G., Bago, B., & McPhetres, J. (2023). Science beliefs, political ideology, and cognitive sophistication. *Journal of Experimental Psychology: General,* 152(1), 80-97.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. Journal of Personality and

Social Psychology, 41(5), 847-855.

Piaget, J. (1976). The affective unconscious and the cognitive unconscious. B. Inhelder, H. H. Chipmand, eds., *Piaget and his school,* 63-71, New York : Springer.

Pillow, B. H., & Pearson, R. M. (2009). Children's and Adults' Evaluation of Their Own Inductive Inferences, Deductive Inferences, and Guesses. *Merrill-Palmer Quarterly*, 55(2), 135-156.

Porumbescu, G. A., Cucciniello, M., & Gil-Garcia, J. R. (2020). Accounting for citizens when explaining open government effectiveness. *Government Information Quarterly*, 37(2), 1-10.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology,* 43(4), 565-578.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin,* 28(3), 369-381.

Pyszczynski, T, & Greenberg, J. (1987). Toward and integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297-340). NY: Academic Press.

Resnick, L., Salmon, M., Zeitz, C., Wathen, S & Holowchak, Mark. (1993). Reasoning in Conversation. *Cognition and Instruction.* 11. 347-364.

Rest, J., Narvaez, D., Beabeau M. J. & Thoma, S. J. (1999). A Neo-Kohlbergian Approach: The DIT and Schema Theory. *Educational Psychology review,* 11(4), 291-324.

Rest, J., Narvaez, D., Thoma, S. J. & Beabeau M. J. (2000). A Neo-Kohlbergian Approach to Morality Research. *Journal of moral education,* 29(4), 381-395.

Rest, J., Narvaez, D., Thoma, S. J. & Edward , L. (1997). Postconventional

moral thinking: A neo-Kohlbergian approach. *Development Psychology,* 21, 709-714.

Robinson, P. D. (2020). *Social Theories of Reasoning*. Doctor of Philosophy, Ohio State University, Philosophy.

Ryfe, D. M. (2006). Narrative and deliberation in small group forums. *Journal of Applied Communication Research,* 34(1), 72-93.

Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science,* 61(10), 2468-2486.

Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences.* 34(2), 57-74.

Stangor, C., & Ruble, D. N. (1989). Strength of expectancies and memory for social information: What we remember depends on how much we know. *Journal of Experimental Social Psychology,* 25(1), 18-35.

Stevens, K. & Cohen, Daniel H. (2020). *Angelic Devil's Advocates and the Forms of Adversariality.* Topoi 40(5). 899-912.

Sunstein, C. R. (2002). The Law of Group Polarization, *The Journal of Political Philosophy*, 10(2), 175-195.

Taber, C. & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755-769.

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly,* 46, 285-292.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, 52(4), 700-709.

Tetlock, P. E., & Levi, A. (1982). Attribution bias: On the inconclusiveness of the cognition-motivation debate. *Journal of Experimental Social Psychology*, 18(1), 68-88.

Tichy, M., Johnson, D. W., Johnson, R. T., & Roseth, C. J. (2010). The impact of constructive controversy on moral development. *Journal of Applied Social Psychology*, 40(4), 765-787.

Thibodeau, R., & Aronson, E. (1992). Taking a closer look: Reasserting the role of the self-concept in dissonance theory. *Personality and Social Psychology Bulletin*, 18(5), 591-602.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143(5), 1958-1971.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* 5, 207-232.

Tversky, A, & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185, 1124-1131.

Uhlmann, E. L., Korniychuk, A., & Obloj, T. (2018). Initial prejudices create cross-generational intergroup mistrust. *PloS one*, 13(4), 1-14.

Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making,* 4(6), 476-491.

Valdesolo, P., & Desteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science,* 17(6), 476-477.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18(8), 689-690.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5), 1334-1338.

van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature neuroscience*, 12(11), 1469-1474.

Wainryb, C. (2011). 'And so they ordered me to kill a person': Conceptualizing the impacts of child soldiering on the development of moral agency.

Human Development, 54(5), 273-300.

Wason, P.C. (1969). Regression in Reasoning? *British Journal of Psychology*, 60, 471-480.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual Processes in Reasoning? *Cognition*, 3, 141-154.

Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: an FMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election. *Journal of cognitive neuroscience,* 18(11), 1947-1958.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). MA: Academic Press.

Abstract

A Study on the Problem and Improvement of Motivated Reasoning in Moral Judgment

Mingyeom Kang
Ethics Education
The Graduate School
Seoul National University

This study aims to analyze This study aims to analyze the factors contributing to biased moral judgment, using the concept of motivated reasoning, and explore potential methods for improving by moral education. Motivated reasoning refers to a phenomenon where the process of reasoning is influenced by motivational factors. Generally, these motives tend to be biased as they seek to justify pre-existing beliefs. In the context of moral judgment, motivated reasoning can compromise the objectivity and fairness of moral reasoning by distorting evidence to protect unquestioned beliefs and identities.

The prevailing discourse in moral psychology, exemplified by Kohlberg's work, assumes that moral reasoning is the psychological mechanism driving moral judgments. However, Haidt's research

challenges the causal role of reasoning in moral judgment. According to Haidt's social intuitionist model of moral judgment, most moral judgments are immediately informed by moral intuition, and moral reasoning serves as a post-hoc justification consistent with already-formed conclusions. From Haidt's perspective, moral reasoning can be seen as a form of motivated reasoning.

In moral judgment, motivated reasoning justifies immoral actions and reinforces moral and political polarization within society. Firstly, motivated reasoning provides a plausible rationale for individuals to view themselves as moral agents despite their immoral actions and judgments. Additionally, it reinforces dogmatic beliefs, perpetuates exclusionary tendencies and group divisions against those with differing viewpoints, and deepens the polarization of moral and political beliefs.

The problems stemming from motivated reasoning in moral judgment are driven by two primary factors: the mind's inclination to protect beliefs and the phenomenon of bias blind spots. People tend to perceive unexamined challenges to their beliefs as threats to their moral and social identity, leading them to instinctively reject such challenges. This motivated reasoning process defends one's beliefs and identity, sometimes even justifying immoral actions and judgments while providing a rationale for ostracizing others with divergent beliefs. Furthermore, individuals often lack objective awareness of their biased thinking. Those with bias blind spots perceive their moral judgments and supporting evidence as objective and impartial, while considering the judgments and evidence of others that conflict with their own opinions as biased. This becomes particularly problematic as bias blind spots are common even among cognitive elites.

To address these challenges, fostering moral meta-cognition offers an opportunity for reflection on unexamined beliefs and a chance to overcome bias blind spots. Moreover, recent interdisciplinary approaches in cognitive science and psychology suggest considering the social function of reasoning to improve motivated reasoning. Collaborative argument exchange is one such approach where individuals exchange arguments from different perspectives to build a shared understanding of ethical issues.

A collaborative argument exchange involves a circular process where participants express their moral judgment on a particular issue in the form of an argument and then exchange evaluations of these arguments. Through this interaction, participants can objectively identify, examine, and revise their initial beliefs.

Collaborative argument exchange classes can be applied as a method of instruction in moral education within schools. Teachers should design lessons that allow students to publicly justify their moral judgments and genuinely evaluate their peers' arguments. Additionally, students should have the opportunity to engage in collaborative argument with peers who hold different viewpoints within trusted small groups.

Lastly, this paper reviews three previously developed models for moral education that use collaborative argument exchange. The Konstanz Method of Dilemma Discussion(KMDD), which includes the process of evaluating opposing arguments, provides insights into the necessary conditions for collaborative argument and considerations for teachers. The constructive controversial issue class model addresses controversial topics in the classroom, offering procedures and detailed guidelines for the lesson. The Haidt-style Moral Insight Model, a direct application of Haidt's moral psychology theory, is still in its

preliminary stage but is expected to contribute to improving motivated reasoning through a complementary virtuous circle with collaborative argument exchange classes.

keywords : Moral Judgment, Motivated Reasoning, Moral Reasoning, Cognitive Bias, Moral Meta-cognition, Collaborative Argument Exchange

Student Number : 2020-39021