

정신의 자율성과 인과성*†

윤 보 석‡

정신의 자율성과 유물론적 세계관의 갈등은 오래된 철학적 주제이다. 흔히 정신의 자율성을 보호하려는 욕구는 유물론의 비판을 통해 표출되었다. 그러나 유물론을 성공적으로 배제하는 것만이 모든 문제를 해결하는 것은 아니다. 이원론의 쇠퇴를 가져온 심신인과의 문제가 현대의 비환원적 유물론에도 제기된다는 지적이 있어왔다. 정신의 자율성을 보존하되 정신의 인과성을 신비스럽게 만들지 않는 길은 없는가? 이 문제에 대한 해결을 모색하는 것이 이 논문의 목적이다. 첫째, 김광수의 최근 논문을 통해 문제를 명료하게 기술하고, 둘째, 계산주의적 모델에 기초한 정신의 자율성 설명에 대하여 제기된 세 가지 비판을 비교 서술하며, 셋째, 그 중 세 번째 비판을 발전시키고자 한다.

주 제 심리철학
주요어 정신의 자율성, 유물론, 심신인과의 문제, 계산주의, 지향성, 일상 심리설명, 김광수

최근의 한 논문 “심리철학과 정신의 자율성”에서 김광수는 환원주의자, 제거주의자, 부수현상론자들 뿐만 아니라 칸트, 데이빗슨, 김재권 같은 철학자들이 공히 “철학적 스캔들”을 벌이고 있다라고 비판하였다. 스캔들의 주된 내용은 그들이 인간현상에 대한 깊은 이해에 도달하고자 의

* 접수완료: 2003. 10. 13. / 심사 및 수정완료: 2003. 11. 6.

† 이 논문은 2002년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2002-075-A00044). 이 논문의 초고에 대한 심사 의견서에서 익명의 두 심사위원들이 특히 3장의 논변을 보완하고 더욱 명료화할 것을 제안하였다. 이 제안은 논문을 향상시키는 - 필자의 생각에 - 계기가 되었다.

‡ 서울대 철학사상연구소 선임연구원.

도하면서 바로 인간 현상의 핵심을 이루는 정신의 자율성을 아예 무시하거나 혹은 말살해버리는 잘못을 범하고 있다는 것이다. 따라서, 정신의 자율성이 우리가 거부할 수 없는 근본적인 사실임을 다시 확인하고 심리철학을 올바른 방향으로 재정립하자고 주장한다. 필자 또한 현대 물리주의적 심리철학이 깊이 있는 반성을 필요로 한다는데 동감한다. 그러나 만약 위에서 언급된 다양한 이론들과 철학자들에 공통된 잘못이 있다면, 즉, 한 두 사람의 특이한 실수에 국한된 것이 아니라 상당히 편재된 문제가 있다면, 그 원인을 밝히는 것이 단순한 작업은 아닐 것이라고 기대된다. 스캔들에 빠지게된 이유를 제대로 밝히기 위해선 섬세한 논증이 요구된다고 보는 것이 적절하다. 김광수의 논문에서 출발하여, 특히 현대 분석철학내의 토론에 초점을 두고, 정신의 자율성에 대한 최근의 논의의 근본적인 문제점을 드러내고 올바른 시각을 모색해보는 것이 이 논문의 목적이다.

1

김광수는 정신의 자율성을 다음과 같이 정의한다.¹⁾

정신의 자율성: 인간은, 자연의 인과 법칙에 따라 기계적으로 작동하는 타율적인 존재가 아니라, 자신의 방식으로 문제 상황을 인식하고, 자신의 신념과 소망으로부터 가장 합리적인 판단을 도출하여 행위 하는 자율적 존재이다.

“자연의 인과 법칙”, “기계적”, 그리고 “합리성”의 개념이 정신의 자율성을 정의하는데 도입이 되고 있는데, 이 정의에 따르면 “기계적”이란 개념과 “합리성”이란 개념사이에 어떤 배타적인 관계가 있음을 암시하고 있다. 우리의 몸은 물론 자연의 인과 법칙에 따라 기계적으로 작동하는

1) 김광수 (2003), p.4.

타율적인 존재이라고 본다면 위의 정의는 전통적인 이원론에 부합한다고 볼 수 있다. 즉, 인간은 기계적인 몸과 합리적인 마음으로 구성되어 있고 그 중에서 후자가 인간의 본질이라고 말하고 있다는 것 같다. 그는 이렇게 정의된 정신의 자율성은 우리가 가지는 세계에 대한 어떤 믿음보다 근본적임을 강조하며 이를 유물론 그리고 과학적 세계관에 대한 한 심각한 도전으로 연결시킨다.

유물론자는 이러한 일(합리적 행위와 그에 기초한 인간만의 문화와 문명)이 어떻게 물리적으로 실현될 수 있는지 설명할 수 있어야 한다. 정신의 자율성이 물리적으로 실현될 수 없다고 해서 이를 외면하거나 부정하는 것은 심리철학의 목적을 망각하는 일이다. 현재의 물리학이 이 일을 감당할 수 없다면, 유물론을 견지하려 하는 한, 물리학 자체가 변해야 할 것이다.²⁾

우선 다음과 같은 질문으로 시작해 보자 - 과연 정신의 자율성과 유물론은 서로 상충하는가? 만약 서로 상충한다면 다음과 같은 선택이 가능하다. 즉, 정신의 자율성을 받아들이고 유물론을 포기하거나, 혹은, 유물론을 받아들이고 따라서 정신의 자율성을 환상으로 간주한다. 이 두 가지 중 어느 것을 선택해야 할까? 논문에서 김광수는 주로 이 질문에 관심을 가지는 것 같다. 전자를 택해야 하는 이유라고 간주될 수 있는 문장들이 있기 때문이다. 그의 입장을 정리하기 위한 한 방법으로 다음과 같은 논변을 고려해보자.

(I) 정신의 자율성과 유물론은 둘 다 참일 수 없다.

정신의 자율성은 실재한다.

유물론은 참이 아니다.

2) 김광수 (2003), p.4. 인용문의 괄호 안은 필자의 삽입임.

물론 (I)에 대해 유물론자들은 두 번째 전제의 근거를 물을 수 있다. 김광수는 두 번째 전제가 “거부할 수 없는 원초적 사실”이라고 하고 그것을 의심하면 심리철학을 할 수 없다라고 한다. 이는 우물에 독타기의 오류에 근접한다는 비판을 불러일으킬 수 있다. 그러나 그의 논문에서 비록 정신의 자율성의 존재를 증명할 수는 없으나, 왜 증명할 수 없는지에 대한 이유가 제시되어 있다. 정신의 자율성은 우리의 “경험 속에 드러난 현상”이고 “다른 사실로부터 추론”하여 알 수 있는 것이 아니라 직접적으로 알려지는 것이라고 주장한다. 만일 정신의 자율성은 그 정신을 가진 주체에게 직접적으로 알려지는 특성을 본질로 가진다면 당연히 정신의 자율성을 의심하는 사람이 있다면 바로 그 사실 때문에 우리는 그 사람이 진짜 정신의 자율성을 가지고 있는지 의심해보아야 할 것이다. 이와 같은 지적은 분명히 두 번째 전제를 의심하는 철학자들에게 향하고 있다.

그러나 정신에 관한 일인칭시점에서의 파악의 성격은 그 자체 논란의 여지가 많은 주제이다. 비록, 일인칭적 지식의 특이성을 인정하는 철학자라도 반드시 오류불가능성이나 강한 투명성을 받아들이지는 않을 수 있기 때문이다. 자신의 행동의 동기에 대한 진술이 그 사람에 대한 증거와 일치하지 않는 경우가 많이 조사되어 있다. 즉, 어떤 사람이 자신의 행동을 가장 잘 설명해주는 심리상태를 부인하거나 혹은 아예 의식하지 않는 경우들이 있다. 무의식적 심리상태나 오류가능성을 받아들이는 것이 반드시 일인칭 시점의 특이성을 부정하는 것은 아니다. 단지, 우리가 우리 자신의 마음에 대해 가지고 있는 믿음이 수정할 여지가 전혀 없고 어떠한 비판으로부터도 보호되어있는지는 의심할 수 있다. 사실, 오류가능한지 아닌지에 주목하는 것 자체가 일인칭 시점의 특이성에 관한 논의를 잘못된 방향으로 인도한다는 주장도 가능하다. 일인칭적 시점은 보다 근본적인 의미에서 삼인칭적 시점과 대비된다. 철수가 왜 사다리를 타고 지붕으로 올라가느냐는 질문을 받았을 때, “바람이 불어 모자가 지붕 위로 올라갔고 그 모자를 찾으려고 한다”라고 보고했다고 하자. 왜 우리가

철수의 설명을 받아들여야 하나? 철수가 자신의 동기에 대해 틀릴 가능성이 없기 때문인가? 어떤 사람의 보고가 다른 사람들에 의해 반박이 될 수 있다. 그러나 오류가능성 때문에 일인칭 시점의 권위가 손상된다고 우리는 생각하지 않는다. 일인칭적 시점의 손상은 철수가 자신이 제시한 믿음과 욕구가 자신의 행동설명에 대한 최선의 가설일 뿐이라고 주장할 때 드러난다. 왜냐하면 행위자가 자신의 심리상태를 단지 법칙에 의해 포섭되는 설명적인 구조로 간주할 수 없고 자신이 그렇게 행동해야만 하는 당위성을 제공하는 것으로 이해해야 하기 때문이다. 자신의 신념과 욕구가 마치 자신이 어떻게 행동할 것인지에 대한 이론적인 확신을 가져 다준다고 생각하는 것은 “신념과 욕구에 따라” 행동하는 행위 주체자의 태도라고 볼 수 없다. 만일 그렇게 생각한다면 우리는 철수가 과연 애당초 합리적인 행위설명의 대상이거조차 한지 의심이 들것이다.

마음에 대한 일인칭 시점의 특이성을 강조하는 사람들은 흔히 감각질을 가진 심리상태에 주목한다. 김광수의 논문에서도 감각질이 언급되어 있는데 마음은 주체에게 자명하게 드러나고 다른 사람에게는 닫혀져 있다는 직관이 사실 치통과 같이 감각질을 가진 상태가 가장 강하다. 그러나 정신의 자율성을 정의하는데 있어 언급된 소망이나 신념은 통상 감각질을 본질로 가지는 상태로 간주되지 않는다. 오히려 여기서 논의되는 합리적 행위의 정의를 위해서는 그 것들의 지향적인 성질이 특히 유관하다고 봐야 할 것 같다. 만일 감각질만이 순수한 의미의 심적인 것으로 가정했다면 이는 논란의 여지가 많다. 그 가정을 받아들일 경우 심리상태의 지향성 즉 세계와의 연관성은 마음 외적인 것이 되고 따라서 지향성이 심적 과정에 어떤 영향을 주는지 즉각 문제가 된다. 이 문제는 계속 논의가 될 것이다. 여기서는 단지 (I)을 옹호하고자 하는 사람은 두 번째 전제에 대해 제기될 수 있는 여러 비판에 적절히 대응하여야 한다는 점을 지적한다. (I)은 다음과 같은 논변과 대조된다.

(II) 정신의 자율성과 유물론은 둘 다 참일 수 없다.

유물론은 참이다.

정신의 자율성은 존재하지 않는다.

유물론은 이론 혹은 가설이고 정신의 자유성은 사실이라는 김광수의 반복되는 주장은 (II)가 주객을 전도한 심각한 오류를 범하고 있다는 지적으로 받아들일 수 있을 것이다. 그러나 논변(II)를 옹호하는 유물론자들은 나름대로 두 번째 전제에 대한 적절한 근거가 있다고 생각한다 (첫 번째 전제는 (I)과 공유하는 부분이기 때문에 따로 증명할 필요가 없다). 예를 들어, 이원론에 비하여 유물론은 경제적이며 두 가설이 동일한 설명력을 가질 경우 경제적인 가설을 택하는 원리에 따라 유물론이 선호되어야 한다고 주장할 수 있다. 물론, 두 가설이 동일한 설명력을 가지지 않는다면, 즉, 유물론에 의해 설명이 되지 않는 현상이 있다면, 위의 주장은 설득력을 상실할 것이다. (II)를 반대하는 사람들은 정신의 자율성이 바로 그러한 현상이라고 주장할 것이다. 그러나 (II)의 논변은 정신을 자율성을 인정하고 그것을 유물론이 설명할 수 있다는 것이 아니고 아예 정신의 자율성이 존재하지 않는다는 주장을 결론으로 가진다. 따라서 그 결론이 틀렸다 라고 주장하는 것은 그 논변에 대한 타당한 비판이 아니다. 물론 김광수는 (II)의 논변에 대해 강한 반발을 할 것이다. 정신의 자율성을 부정하면 윤리, 문화와 문명이 어떻게 가능한가?

이 시점에서 지적되어야 할 사항은 위와 같은 논쟁이 모든 가능한 입장을 다 포섭하지 않는다는 사실이다. 왜냐하면, 모든 유물론자들이 (II)를 옹호하지는 않기 때문이다. 아마 대부분의 유물론자는 정신의 자율성을 부정하지 않고 그것이 유물론적 세계관과 양립 가능하다는 견해를 지니고 있다. 이런 견해에 대해 김광수는 양립불가능성을 계속 옹호하거나, 혹은, 양립가능성을 인정하고 (I)에서 후퇴할 수 있다. 필자가 이해하기엔 김광수는 반드시 양립가능성을 부정하기 보다, 단지 만일 유물론자들이 양립가능성을 믿는다면 어떻게 가능한지를 보여야 할 부담이 바로 그들에게 있다고 주장한다. 즉, 유물론자들에게 “입증책임”이 있다는 것이다. 그러나 한 특정한 유물론에 대해서는 명시적으로 양립가능성을 부정

한다.

정신의 자율성을 심리철학의 피설명항으로 받아들이는 철학자들은 새삼스럽게 '환원불가능성'과 같은 점을 들어 정신의 자율성을 옹호할 필요가 없다. 환원적 유물론자들이 정신의 자율성을 물리적 현상으로 환원시켜 설명하고자 할 때, 이 시도가 성공적이지 못함을 지적하면 되는 것이다. 환원불가능성은 정신의 자율성에 대한 근거가 아니라, 정신의 자율성이 함의하는 것이다.³⁾

정신의 자율성이 환원불가능성을 함의한다면 어떤 환원주의적인 시도도 필연적으로 실패할 수밖에 없다. 따라서 구체적으로 제시된 개별 환원이론에 대해 그것이 성공적이지 않음을 보일 필요조차 없게 된다. 환원주의자들은 아마 다음과 같이 반발할 것이다. 자신들은 유물론자들이고 따라서 위에서 지적된 입증책임의 귀속을 받아들여 정신이 자신들의 형이상학 틀 내에서 허용된 어떤 성질에 - 예를 들어, 어떤 물리적 혹은 생물학적 성질 - 환원되는지 밝히려는 시도를 하고자 하는데 지금에 와서는 기다려 보지도 않고 아예 그런 시도 자체가 모순이라고 주장하는 것은 도대체 비판의 초점이 어디에 있는지 알 수 없다. 즉, 정신의 자율성을 피설명항으로 놓고 설명해야만 된다고 말했다가 설명을 제공하려고 시도하려는 순간 아예 설명이 되지 않는다고 주장하는 것이나 다를 바 없다. 물에 대한 여러 사실 (예: 100도에서 끓음)이 분자들의 물리적인 성질에 의해서 설명이 되고 따라서 물은 특정한 분자로 동일시되었다. 물론, 그런 설명이 있기 전에 환원이란 상상할 수 없었으나 그 상황이 변화하지 않았는가? 이러한 변화가 왜 정신현상에 대해서는 불가능하다는 것인가? 정신현상은 본질적으로 다른 시각에서 파악될 수 없기 때문인가? 우리의 현재 개념의 틀 내에서는 도저히 물리적 시각에서 동일한 현상이 어떻게 파악될 지 알 수 없더라도 현재 개념의 틀이란 가능한 여

3) 김광수 (2003), p.17.

러 틀 중에 하나일 뿐 아닌가? 아니면, 현재 개념의 틀을 벗어나면 그 틀에서 파악된 현상까지 사라져 버린다는 말인가? 이런 질문들에 대한 답이 김광수의 현재 논문에서 찾기 힘들며, 논의를 진행시키기 위해선 정신의 자율성이 유물론과 양립 불가능하다는 주장과 양립 가능할지도 모르지만 어떻게 가능하지 알 수 없다는 주장, 그리고 아마 우리의 지능으로는 그 문제는 해결할 수 없다라는 주장 등 여러 다른 주장들이 명확히 구분될 필요가 있다.

더군다나, 환원주의를 배제할 수 있다고 해서 모든 문제가 다 해결되는 것은 아니다. 김광수의 주장을 따라 정신의 자율성이 환원불가능성을 함의한다고 해보자. 그렇다면, 양립가능성을 믿는 유물론자가 택할 수 있는 입장은 비환원적 유물론이 될 것이다. 여기서 전통적으로 이원론의 치명적인 문제점으로 인식되는 심신인과의 문제가 등장한다. 김재권의 널리 알려진 논변에 따르면 비환원적 유물론도 결국 동일한 문제에 봉착한다. 물론, 김재권의 논변이 모든 비환원적 유물론에 다 적용이 되는지는 논란의 여지가 있고, 필자도 다음 장에서 이와 연관된 한가지 지적을 할 것이다. 우선 김재권의 논변을 받아들여 비환원적 유물론을 배제할 수 있다 해보자. 이 또한 모든 문제가 해결되는 것은 아니라는 주장이 가능하다. 합리적인 행위자를 "자연의 인과법칙에 따라 기계적으로 움직이는 타율적인 존재"로 환원하는 것은 결국 합리적인 행위자를 소멸하는 것과 다름이 없다는 것이다. 종합하면 다음과 같은 진퇴양난이 생긴다. 인간의 정신을 자연의 인과법칙 내에 위치시키면 정신의 자율성이 말살되고 한편으로 정신을 자연의 인과법칙밖에 놓여진 것으로 보면 정신이 어떻게 합리적인 과정을 거쳐 몸을 움직이고 환경을 바꾸어 문화와 문명을 만들어 낼 수 있는지 문제가 된다. 몸이 움직이려면 어떤 물리적인 인과기재가 있어야 할 것이다. 성대와 혀, 팔과 다리 등 몸의 부위의 활동을 조절하고 통제하는 것은 인과적으로 선행하는 두뇌의 복잡한 그러나 궁극적으로 자연의 인과법칙 하에서 전개되는 과정일 것이다. 자연의 인과법칙밖에 놓여진 자율적인 정신이 이러한 과정에 어떻게 개입할 것

인가? 우리의 신념과 소망이 “세포벽을 침투해 들어가서 세포핵을 공격하는가?”⁴⁾ 아니면 어떤 염동작용(telekinesis)과 같은 기재가 있는가? 환원주의자들의 자라는 이런 당혹스런 상황을 피할 수 있다는 것이다. 즉, 만일 신념과 소망이 실제로 두뇌의 물리적인 어떤 상태라고 하면 심신인과의 문제는 아무런 신비스러운 개입을 상정하지 않아도 된다는 것이다. 물론 김광수는 환원주의자의 성공을 실패나 다름없는 성공으로 간주할 것이나 최소한 환원주의가 어떤 면에서 매력적으로 보일 수도 있음은 인정해야 할 것이다. 요약하자면, 비환원적 유물론/이원론을 택하면 심신인과가 문제가 되고, 환원주의를 택하면 정신의 자율성이 손상되는 것 같다. 이 진퇴양난에서 빠져날 율 방법은 무엇인가?

2

심신인과의 문제는 데카르트적 이원론의 몰락을 가져왔다. <마음은 내부, 세계는 외부>라는 이분법은 물리적인 세계 내에 정신의 자율성이 어떻게 가능한지에 대한 탐구를 어렵게 만든다 - 마치도 순전히 외적이고 객관적인 세계에 내적이고 주관적인 어떤 실체를 삽입해 넣어야만 하는 것처럼. 그러나 우리가 이해하는 합리적인 행위 주체는 애당초 세계와 단절되어 내적 상태에만 주목하고 반응하는 개체가 아니라 세계가 어떠한지에 대한 믿음에 따라 행위를 조절하는 세계 내의 존재이다. 이성이 물리세계에 어떻게 인과적인 차이를 가져올 수 있는가? 어떤 유물론자들은 이 오래된 형이상학적 질문에 대해 컴퓨터가 마침내 과학적인 한 대답을 제공한다고 주장한다. 인지과학의 계산주의 모델을 도입하여 정신의 자율성을 설명하려는 시도를 구체적으로 살펴보자.

“밖에 비가 온다”라는 문장이 구체적으로 칠판에 쓰이거나, 혹은, 발화되었을 때, 그 문장 사례들 혹은 개별자들(tokens)은 의미적인 성질과 더

4) Searle (1984), p.17.

불어 물리적인 사건으로서의 인과적인 성질을 가진다. 예를 들어, 어떤 파장의 빛을 반사하거나 어떤 주파수를 가진다. 밖에 비가 온다는 믿음을 내적 상태, 보다 더 구체적으로 두뇌에 어떤 문장의 사례가 쓰여지는 것으로 보자. 두뇌에 쓰여진 문장들이 있다면 그것들도 의미적인 성질과 인과적 성질을 가질 것이다. 인과적으로 유효한 물리적인 성질이 어떻게 명제태도간의 논리적, 의미론적 관계를 존중할 수 있는가? 철수가 “만일 밖에 비가 온다면 우산을 가지고 가야한다” 그리고 “밖에 비가 온다”라는 두 믿음으로부터, “우산을 가지고 가야한다”라는 믿음을 형성한다고 가정해 보자. 이러한 믿음 형성이 어떻게 가능한가는 이제 믿음의 통사론적 성질에 정의된 계산적 과정에 의해 설명된다. 형식 논리학의 증명 과정이 타당성과 같은 성질을 보존하듯이 인지 과정을 명시적인 재현에 정의된 통사론적인 과정으로 보는 것이다. 포도르에 따르면, 적절히 프로그램된 내적 계산과정이 바로 정신의 자율성을 물리적으로 실현하며, 따라서 계산주의적 접근은 명제태도가 어떻게 인과적인 힘을 가질 수 있는지, 혹은, 명제태도가 어떻게 하여 논리적인 관계를 따라 믿음과 행위를 유발할 수 있는지 설명해 준다고 한다.⁵⁾ 김광수가 유물론자에게 부여한 과제 - 정신의 자율성이 어떻게 물리적으로 실현될 수 있는가? - 에 대한 최초의 과학적으로 존중받을 수 있는 설명을 컴퓨터가 제공한다는 것이다. 요약하면, 심적 과정은 통사론적, 형식적인 과정이다(mental processes are syntactic).

계산주의 모델에 기초한 정신의 자율성 설명이 과연 성공적인가? 김광수의 논문에서 이 질문이 본격적으로 다루어지지 않는으나 그의 답은 부정적이라고 짐작된다. 필자도 이 논문에서 계산주의 모델에 기초한 정신의 자율성 설명이 잘못되었다고 가정할 것이다. 필자의 주된 관심은 계산주의에 기초한 유물론적 설명이 왜 잘못되었는가 이다. 사실 이 질문에 답하는 것도 단순하지 않다. 몇 가지 상당히 다른 비판들이 존재하기

5) Fodor (1987), p.26.

때문이다. 신념과 소망에 의거한 합리적 행동설명을 과학적으로 정당화하려는 시도에 대한 첫 번째 비판(비판-1)은 과학 내에서 찾을 수 있다. 계산주의 모델은 두뇌의 인과적인 구조에 대해 강한 경험적인 제약을 설정한다. 따라서 두뇌의 역학에 관한 다른 모델들과 경쟁하는 관계에 있게 되고, 만일 두뇌 속의 인지적인 과정이 고전적인 계산과정과 전혀 다른 새로운 원리에 기초한 역학에 의해 보다 더 잘 설명될 수 있다면, 이는 일상적 심리상태의 과학적 정당화를 손상할 수 있다. 연결주의 모델을 옹호하는 철학자들은 신념과 소망과 같은 명제태도의 구조와 내용에 해당하는 내적 상태가 존재하지 않기 때문에 명제태도는 제거되어야 할 것이라고 주장한다. 일상적인 심리학은 하나의 이론을 형성한다는 가정 하에서 그 이론은 과학의 다른 이론과 조화되기 어렵고, 지극히 제한된 설명 영역을 가지며, 오래 동안 전혀 변화하지 않은 정제된 이론이라는 것이다. 따라서 일상 심리이론은 과학적 설명에 부드럽게 흡수되기보다는 인간행동에 대한 보다 나은 이론에 의해 교체되어야 하며, 그 이론에서 상정하는 개체들은 따라서 존재하지 않는 것으로 간주한다. 첫 번째 비판은 제거주의와 연결됨으로 따라서 김광수와 같이 제거주의를 옹호하지 않는 철학자들은 다른 비판을 찾아야 할 것이다. 두 번째 비판(비판-2)으로 넘어가자.

“심적 과정이란 통사론적 과정이다”라는 주장은 의미론적 성질과 인과력을 결정하는 통사론적 성질의 분리를 전제한다. 즉, 내적 표상이 의미론적 성질과 통사론적 성질을 가지는데 통사론적 성질에 대하여 정의된 계산과정이 표상들의 의미론적 연관을 반영하는 것이다. 그러나 여기서 내적 표상의 의미론적 성질이 가지는 인과적인 역할에 대해 의문이 제기될 수 있다. 만일 의미론적 성질의 인과적인 역할이 없다면, 과연 순전히 통사론적인 과정을 진정으로 “사고 과정”이라고 간주할 수 있는가? 드레츠키는 심적 과정에 대한 통사론적 접근에 대한 불만을 다음과 같이 지적하고 있다.

만일 의미론적 성질이 적절한 통사론적 성질과 병행한다면 의미론적 성질을 언급하여 유용한 일반화(예측을 하기에 유용한)를 구성할 수도 있다. 인과적으로 무력한 의미를 사용하여 인과적으로 유효한 내적 상태의 형식적 성질을 지칭하고 분류하는 것이 방법론적으로 필요 불가결 할 수도 있다. 그러나, 이것이 의미를 설명적인 개념으로 만들어 주지는 않는다.⁶⁾

만일 내적 상태의 인과적인 통사론적 성질이 의미론적 성질과 서로 병행하고 있을 때, 우리는 의미론적 성질을 지칭하여 행동을 예측할 수 있을 것이다. 그러나, 이것이 “우리가 P를 생각하기 때문에 A를 한다”라는 것을 참으로 만들어 주지 않는다. 무엇을 생각하느냐가 독자적인 설명력을 가지지 않는다면, 왜 우리는 다른 사람의 행동을 설명하기 위해 그들이 무엇을 생각하는지에 관심을 가져야 하는가? 만일 생각의 내용 자체가 설명력을 결여 할 경우, 의미의 제거가 설명력의 손실을 가져오지 않게 되고, 의미를 본질적인 성질로 가지는 심리상태도 존재하는 이유를 상실하게 된다. 있으나 없으나 아무런 인과적인 차이를 가져오지 않는다면 없다고 한들 무슨 상관이 있는가?

심적 과정의 통사론적 접근에 대한 드레츠키의 불만은 명제태도에 대한 실재론을 옹호하는 다른 여러 철학자들도 공유한다. 써얼의 중국어방 논증이 그 한 예이다.⁷⁾ 그 논증에 따르면, 통사론적 기호 조작은 아무리 복잡하다고 해도 진정한 심리 과정을 형성하지 못한다. 즉, 통사론적 과정이 존재하더라도 심적 과정이 부재할 수 있다는 것이다. 따라서, 통사론적 과정은 충분하지 못하다. 김재권도 그의 비환원주의적 유물론에 대한 비판에서 의미론적인 성질 자체가 그것을 가진 내적 상태에 인과적인 힘을 부여하여야 한다고 주장한다. 만일, 내적 상태로부터 의미론적인 성질을 제거하더라도 그 내적 상태의 모든 인과적 사실에 전혀 변화가 없다면, 이는 부수현상론에 귀착하고 만다는 것이다. 의미론적 성질이 한

6) Dretske (1988), p.81.

7) Scarle (1984).

중요한 심적 성질이고, 심적 성질에 대한 실재론을 옹호하고자 한다면 의미론적 성질에 독자적인 인과력을 부여해야만 한다는 비판이다.⁸⁾ 계산주의에 대한 두 번째 비판을 제기하는 철학자들은 따라서 심성의 독자적인 인과력을 부여하는 문제를 심각히 받아들인다. 심리철학 내에서 이 문제는 - 유형 부수현상론을 어떻게 반박할 것인가? - 많은 주목을 받았다.⁹⁾

비판-1과 비판-2의 차이점은 다음과 같다. 비판-1은 인지과정이 문장적 역학을 따르고 있다고 볼 증거가 희박하다는 것이고 따라서 고전적 계산주의에 의거한 명제태도 옹호가 경험적 근거가 박약하다는 주장이며, 반면에 비판-2는 비록 인지과정이 문장적인 역학을 가지고 있다고 해도 이 자체가 정신의 자율성이 물리세계 내에서 어떻게 가능한지 보여주지 않는다는 것이다. 통사론적 과정으로는 부족하고 의미론적 성질 자체가 인과력을 독자적으로 행사할 수 있어야 한다는 비판은 김광수의 정신의 자율성을 존중해야 한다는 요구에 부응하는 한 시도라고 볼 수 있다. 따라서 비판-2를 제기하는 위의 철학자들은 심리철학자들이 정신의 자율성을 외면하거나 경시한다는 지적이 자신들의 경우에는 해당하지 않는다고 항변할 것이다. 자신들이 시작한 상황에서 의미론적 성질에 인과력을 실제로 부여할 수 있는지는 별개 문제이다 - 이에 대한 필자의 견해는 사실 부정적이다. 그러나 그들의 주장은 자율성을 가지는 것처럼 보이는 것과 진정으로 자율성을 가지는 것을 분명히 구분해야 한다는 것이다. 이것이 부수현상론의 도전을 심각히 받아들이는 동기가 된다. 즉, 부수현상론의 도전을 해결하는 것은 자율성에 관한 우리의 믿음을 정당화하기 위해 필요하다.

그러나 과연 정신의 자율성을 심각히 받아들이는 것이 행동의 원인에 대한 과학적 탐구와 마찰을 일으킨다고 생각해야만 하는가? 정신의 자율

8) Kim (1990).

9) 포도르도 엄격하지 않은 법칙(non-strict, ceteris paribus laws)의 인과성에 대한 분석을 통해 이 문제를 해결하려 하였다. 그의 논증은 Fodor (1990), 5장 "Making Mind Matter More"에 제시되어 있다.

성에 대한 물리적 실현을 찾는다는 것은 정확히 무엇인지 고려해 보자. 우리는 두뇌가 의미론적 기계가 아니라고 가정하고 - 달리 생각할 이유가 있는가? - 그럼에도 불구하고 두뇌가 어떻게 우리를 의미론적 기계가 되게 해주는 가라고 묻는다. 즉, 계산주의 모델은 의미론적 기계를 가정하지 않고 의미론적 기계를 “설명”하려는 모델이다. 여기서 우리가 원하는 설명이 무엇인가? 우리가 친숙하게 인지하는 합리적 행위자의 두뇌 내부에 그의 행위를 조절하는 어떤 작은 합리적 행위자가 들어있다고 상정하는 것은 우리가 원하는 설명이 아니다. 의미론적 기계를 상정하여 의미론적 기계를 설명하려는 것은 아편을 먹으면 잠이 오는 이유가 아편이 잠이 오게 하는 성향을 지니고 있기 때문이라고 설명하는 것과 다름없다. 우리는 사람들이 자신이 무엇을 믿고 무엇을 원하느냐에 따라 행동한다고 이미 잘 이해하고 있다. 따라서 서로를 이유에 의한 설명을 적용할 수 있는 대상으로 간주하는 차원(인격의 차원, *personal level*)과 이러한 이성적 능력을 가능케 하는 두뇌의 인과 과정의 성격을 밝히는 작업의 차원(두뇌의 차원, *sub-personal level*)을 구분한다. 두뇌의 차원에서의 탐구가 인격의 차원에 대한 인과적인 설명을 제공하지 어떤 구성적인 설명(*constitutive explanation*)을 제공한다고 생각하지 않는다. 통사론적 이론이 두뇌에 관한 이론이라면 그것의 평가는 진정한 심리과정을 복구해 주어야 한다는 기준에 좌우되지 않아야 한다. 그렇다면 통사론적 과정이 진정한 심리과정이 아님으로 따라서 통사론적 이론이 틀렸다 라는 비판은 잘못되었다. 비판-2는 마치도 우리가 적절한 물리적/생물적 기제를 찾게 되면 바로 거기서 의미론적 기계를 찾을 수 있어야 한다는 주장인 것처럼 보인다. 이러한 요구는 혼동에 기초해 있는 것 같다. 왜 카멜레온의 피부색이 배경에 따라 변하는가? 하나의 설명은 “적으로부터 자신을 보호하기 위함”이고, 또 다른 설명은 “카멜레온 피부 내의 화학적 작용”이다. 이 두 설명이 경쟁관계에 있는가? 정신의 자율성에 대한 물리적인 기반을 밝히는 작업의 목적이 마치도 우리의 시각에서 없어져 버린 어떤 것을 다시 복구하는데 있고, 따라서 그것을 복구시켜주지 못할 경우 그

작업은 실패한 것인가? 이러한 생각은 우리의 설명적인 작업에 아무런 도움을 주지 못하며 오히려 과학적 탐구의 성격을 오도한다. 우리는 의미론적인 기계에 대한 믿음이 앞으로 밝혀질 물리적인 기계에 의존한다고 생각하지 않으며, 오히려 다른 사람의 생각과 욕구를 해석하고 그 입장에서 행동을 이해하려는 우리에게 이미 친숙한 설명 방식을 가정하고, 그 다음 우리는 어떤 물리적인 기제가 있는지 탐구한다. 사실 이러한 방식이 정확히 어떤 물리적 기제가 우리의 질문에 관련성이 있는지 알 수 있는 유일한 방식이다.

비판-2가 제기하는 문제의식 자체에 대한 비판을 제기하였다. 여기서 일단 계산주의 모델에 의한 정신의 자율성 설명에 대한 마지막 비판(비판-3)을 보자. 비판-1을 얘기할 때 두뇌의 인과과정에 대한 두 가지 모델에 기초한 명제태도의 위상에 대한 상반된 견해를 보았는데, 이 견해들은 다음과 같은 공통점을 가진다. 일상심리 설명이론이 참이고 그 것에서 도입되는 명제태도가 실재한다면 이는 내적 인지과정에 대한 과학이론에 의해 정당화되어야만 한다는 가정이 그 것이다. 한 견해는 합리적인 믿음과 행동의 형성과정이 실제로 내적 인지과정에 - 사고언어의 문장에 정의된 통사론적 과정 - 반영되어 있기 때문에 정당화된다고 하고, 다른 견해는 내적 인지과정이 그렇지 않기 때문에 따라서 정당화되지 않는다는 것이다. 그러나, 애당초 명제태도와 그에 정의된 합리적 행위와 과연 내적 인지과정에 대한 어떤 과학 이론에 의해 정당화(vindicate)될 필요가 있는가? 명제태도의 상호작용에 의한 행동설명이 두뇌의 구조에 관한 과학적 이론에 의해 정당화되지 않으면 그런 설명이 폐기되어야 하고 따라서 명제태도는 실재하지 않고 기껏 실용적인 가치만 지니게 되는가? 연결주의가 참이더라도 여전히 신념과 소망에 의한 설명이 보존될 수도 있지 않을까? 다시 말해, 합리적인 행위가 가능하기 위해서는 반드시 특정한 어떤 종류의 내적 기제가 필요하다고 보아야할 이유를 찾기가 쉽지 않다. 이런 회의적인 태도를 취하는 철학자들은 우리가 서로에게 부여하는 신념과 소망이 과연 애당초 내적인 인과과정 - 즉, 물리적으로

기술 가능한 자극과 반응의 사이에 놓인 인과적 매체 - 에 관한 경험적인 규제를 요구하는지에 대해 회의적이다. 심적 상태가 내적인 물리 상태와 동일시되어야 만 심적 상태가 행위의 원인이 될 수 있음을 부정하는 것이다. 만일 심적 상태가 내적 물리 상태와 동일시(token-identity)되지 않으면 그 심적 상태가 바로 심적 성질 때문에 가지는 인과력이 있는지를 물을 수조차 없게 된다. 따라서 비판-3은 비판-2에 비해 더 극단적인 비판을 제기한다고 볼 수 있다.

비판-3을 제기하는 데넷과 같은 철학자들은 신념과 소망을 바이러스와 같이 (데넷이 사용하는 비유) 우리의 몸 속에서 발견될 수 있는 개체들로 보는 것을 반대한다. 우리가 개체 내부에 대해 전혀 무지하더라도 지향적 관점에서 행동을 예측하거나 설명할 수 있고, 그런 작업이 성공적이라면 그 개체는 신념과 소망을 가진다는 것이다. 이성적 심리설명은 상당히 이질적인 구성을 가진 개체들의 구체적인 특성으로부터 추상화하여 그들이 가지는 보다 높은 단계의 규칙성에 착안하고 있으며, 따라서 하위단계에서 명제태도에 해당하는 구조를 요구하지 않는다. 명제태도가 바이러스와 같이 우리 두뇌 속에서 발견될 수 있다는 것을 부정하지만 그렇다고 명제태도가 존재하지 않는다는 것은 아니며 이를 무게중심과의 유비를 통해 주장한다. 무게중심에 해당하는 점을 물체의 내부에서 찾을 수 없으나 무게중심에 중력이 작용한다고 보는 것이 그 물체의 운동을 예측하거나 설명하는데 과학적으로 유용함과 마찬가지로 합리성을 가정하는 해석과 예측 방식은 명제태도에 해당하는 개체내의 어떤 구조들을 요구하지 않으나 개체의 전반적인 행위를 예측/설명하는데 유용하기 때문에 진화해 왔고, 우리가 진화했기 때문에 잘 작동하는, 이상적이고, 추상화된, 도구적 방법으로 간주한다.¹⁰⁾ “점심무료제공”이라는 간판이 서울의 한 음식점 앞과 뉴욕의 한 음식점 앞에 놓여 있다고 할 때 우리는 그 표시가 가질 인과적인 차이에 대해 상당한 확신을 가진다. 사람들의 행

10) Dennett (1987), p.49.

동에 대한 성공적인 예측은 어떤 규칙성에 기초해 있고 그 규칙성은 지향적인 어휘(intentional terms)에 의해서만 포착될 수 있다. 만일 지향적인 어휘를 배제하고 물리적인 단계에서 예측을 하려고 하면 이는 원리적으로 가능할 지도 모르나 계산적으로 제어불능의(computationally intractable) 작업이 되어버릴 것이다. 지향적인 예측이 진화해온 것은 비록 그런 예측 방식이 여러 요인에 의해 잘못될 수 있고 따라서 그것에 의거하여 행동하는 것이 위험성을 가지나 한편으로 그런 예측을 사용하지 않을 때 오는 엄청난 비용을 절감해 주었기 때문이다. 진화과정에서 선택된 훌륭한 도구라는 사실이외에 더 이상의 정당화가 필요한가?

3

위에서 정신의 자율성이 어떻게 물리적으로 실현되는가에 대하여 계산주의적 모델에 근거한 설명을 언급하고 이에 대한 세 가지 비판을 나열한 다음 그들의 차이점에 대해 간략히 언급하였다. 비판-2에 따르면 명제 태도에 해당하는 통사론적 구조가 두뇌에 있다는 것이 정신의 자율성을 보이기 위해 충분조건이 아니라는 것이고 비판-3에 따르면 필요조건도 아니라는 것이다. 현대 물리주의적 심성이론들이 이원론부터의 진정한 해방을 제공하지 못하고 오히려 이원론의 근본적인 문제점을 계승하고 있다는 판단을 비판-3에서 찾을 수 있다고 본다. 즉, 비판-3은 이원론으로부터의 진정한 탈피를 위한 모색을 담고 있다. 그렇다고 해도 완성된 그림이 되기까지는 해결되어야 할 문제점들이 많이 있다. 이 장에서는 비판-3에 대한 반발을 알아보고 이에 대한 가능한 대처를 모색함으로써 논의를 더 진전시켜 보겠다.

명제태도를 물리적으로 기술될 수 있는 어떤 내적인 상태로 동일화시키는 것에 대한 데넛의 반발은 분명히 그를 비환원주의자로 간주하게 해주며 이는 김광수의 입장과 부합한다. 그러나 비환원주의에도 여러 유형이 있을 수 있고, 우선 비판-2를 제기하는 철학자들은 데넛의 비환원주의

가 반실재론에 귀착된다고 비판할 것이다. 비록 지향적 전략을 택하는 것이 유용하다고 하더라도 이 자체가 우리가 부여하는 지향적 태도의 성격이 무엇이며 그것이 실재하는 지에 대한 답은 제공하지 않으며, 위에서 언급하였듯이 사실 데넷은 명제태도들이 문자 그대로 두뇌 속에 실재하는 것은 아니라고 주장한다. 데넷의 견해에 동정적인 철학자들도 신념과 소망을 가진 지향적인 개체(true believer)란 바로 그의 행위가 지향적인 시점에서 충분히 잘 예측되는 개체일 뿐이다 (행위의 내적 인과구조가 어떠한 지간에) 라는 데넷의 주장이 반실재론 혹은 도구주의에 귀착한다는 비판을 해왔다.

우리가 물이 0도에서 언다는 사실을 성공적으로 예측할 수 있다해도 물의 그러한 성질은 우리가 혹은 누군가 물에 대한 사실을 예측할 수 있는 능력과 독립되어 존재한다. 데넷의 이론에 따르면 물이 0도에서 언다는 믿음을 누가 가지고 있느냐는 다른 사람이 그의 행동을 예측하기 위한 어떤 전략을 가지고 있느냐에 따라 결정된다.¹¹⁾

만일 베이커의 이런 해석이 옳다면 데넷의 이론은 받아들이기 어렵다. 철수가 영희의 행동을 예측하기 위해 영희가 물이 0도에 언다는 믿음을 부여하더라도 실제로 영희가 그런 믿음을 가지느냐 아니냐는 철수의 부여와 독립된 사실이다. 그러나 데넷은 지향적 전략(Intentional Stance)에서 부여된 명제태도와 그들의 관계가 객관적으로 존재한다고 주장한다. 어떤 패턴이 특정 관점에서만 보인다고 해서 그것이 주관적임을 필하지는 않고, 만일 화성에서 온 외계인이 지향적 전략을 택하지 않는다면 그들은 우리에게 대한 전적으로 객관적인 사실을 놓치고 만다는 것이다. 그렇다면 위의 해석은 - 데넷의 견해에 따르면 믿음이 전략 의존적(stance-dependent) 이다 - 다소 지나친 감이 있다. 왜냐하면, 그 해석에 따르면 데넷은 외계인이 지향적 전략을 택하지 않는다면 우리가 가진 믿음이 없

11) Baker (1994), p.334.

어져 버린다고 주장하기 때문이다. 아마 데넛은 지향적인 전략을 택하지 않으면 믿음은 없는 것이나 마찬가지로 주장하는 것 같다. 데넛은 어떤 것이 “객관적”이라고 해서 그 것이 모든 관점과 독립되어 존재한다거나 혹은 어떤 임의의 다른 관점에서 서술 가능하다는 뜻으로 주장하는 것 같지는 않다. 데넛은 반실재론 비판에 대응하여 명제태도를 무게 중심이나 적도선과 같은 추상적인 존재자로 보아야한다고 주장하기도 한다. 무게 중심이 있다고 우리가 믿는 것처럼 명제태도도 실재한다는 것이다. 적도선이 있기는 하지만 배가 적도선에 걸려서 항해에 지장을 받는지는 않는다. 우리가 적도선에 대한 실재론자라고 해서 적도선이 물리적인 인과 과정에 개입한다고 생각해야 하는가? 물론 믿음이나 욕구는 추상적인 존재자라기보다 행동과 인과 관계가 성립하는 구체적인 대상이다. 철수가 사다리를 타고 지붕에 올라가는 행동이 지붕위로 날아간 모자를 찾자 하는 그의 의도에 의해 설명될 경우 그의 의도는 행위의 구체적인 원인이다. 그 구체적 사건은 철수, 특정 시간, 그리고 “지붕 위의 모자를 찾는다”라는 내용에 의해 개별화(individuate)된다. 그러나 여기서도 적도선과 마찬가지로 믿음과 의도가 철수 내부의 신경생체학적인 인과과정에 개입해야만 실재한다고 생각할 필요가 없다.

데넛의 도구주의적 경향은 명제태도를 두뇌의 신경생체학적 상태로 환원하는 것에 대한 반발로 해석할 수 있다. 이 반발은 유형동일론 뿐만이 아니라 개별자동일론 에도 확장된다. 물론 심-신동일화가 매력적으로 보여질 수는 있다 - 누구도 두뇌의 신경생체학적 상태가 존재하고 신경생체학적 상태가 원인임을 부인하지는 않을 것임으로. 역사적으로 믿음, 의도, 소망을 두뇌의 물리 상태, 뇌신경학적 구조와 동일화하게 된 동기는 행동주의에 대한 반발이 그 하나이다. 심적 상태를 가진다는 것이 행동의 성향을 가지는 것으로 이해되지 않는 한, 즉, 우리와 행동의 성향은 동일하나 그럼에도 불구하고 심적인 차이가 가능하다고 보면, 심적 상태를 행위의 내적 원인으로 간주하는 것은 당연한 것처럼 보일 수 있다. 한 보편적인 견해에 따르면 사다리를 타고 지붕에 올라가는 행위는 심적

개념과 독립되어 우리에게 주어진 기본적인 자료이고 심적 개념의 인식적 내용은 그 자료를 넘어서 그 자료를 설명하는 내적 인과구조를 서술하기 위해 도입된 이론에서 도출된다. 물론, 이 정의는 인과적 역할을 실제로 수행하는 내적 구조가 물리적이냐 아니냐와 독립되어 주어진다. 물리적인 실현의 가능성을 열어두지만 심적 상태의 기능화는 심리상태가 물리상태임을 요구하지는 않고 보다 추상적인 단계에서 제시하기 때문에 개별자 동일론의 틀 내에서 위에서 말한 인격과 두뇌의 두 단계의 구분이 유지할 수 있는 것 같은 인상을 준다. 즉, 인격의 차원은 기능적으로 명시된 내적 상태로서 심리상태를 파악하는 것이고 두뇌의 차원은 그 기능들이 실제로 실현된 구체적인 상태로서 심리상태를 파악하는 것으로 이해될 수 있는 것처럼 보인다.

그러나 기능적인 단계와 물리적 단계의 구분이 인격의 차원과 두뇌의 차원의 구분과 동일한지는 의심스럽다. 우선 자극과 반응을 중계하는 내적 인과적 역할을 수행하는 상태는 지향적인 내용을 본질적으로 가지지 않는다. 물리적인 성질간의 관계를 기술할 때 수학적 등식을 사용한다고 해서 (예: $f = ma$) 물리적 대상과 추상적인 대상간의 어떤 관계를 상정해야 할 필요가 없는 것처럼 명제내용간의 논리적 관계를 사용하여 내적인 상태들의 관계를 기술한다고 해서 내적인 상태가 명제내용을 본질적으로 가짐을 요구할 필요는 없다. 그러나 어떤 상태가 명제태도로 간주되기 위해서는 내용을 본질로 가져야 한다. 따라서 행동주의 반발로 시작하여 심적 상태를 내적인 인과구조로 기능화하는 작업이 인격의 차원과 두뇌의 차원의 구분을 보존하였다고 보기 힘들다. 오히려 기능화 작업은 사실 상 그 구분을 파괴했으며 이 것이 바로 의미론적 성질에 독자적인 설명력을 부여하여야만 인격의 차원을 복구할 수 있다는 비판-2의 동기가 되었다고 생각할 수 있다. 물론 비판-2는 문제의 근본적인 원인을 진단하기보다는 현상적으로 드러난 증세에 주목한다. 필자의 생각은 그 증세는 표준적 견해의 깊은 문제점을, 즉, 일단 표준적 견해에서 출발하여 의미의 설명력을 복구한다는 것은 이미 늦었다는 것을, 드러낸다.

개별자 동일론을 받아들이게 되면 심-물 인과관계는 더 이상 아무런 -최소한 물-물 인과관계가 가진 문제 이상의 - 문제가 없다. 물론 개별자 동일론이 모든 문제를 다 해결한다고 주장하는 물리주의자는 없다. 비판-2를 제기하는 철학자들은 개별자 동일론을 받아들여더라도 심적 유형이나 성질이 어떤 인과적 차이를 가져오는 가는 여전히 문제로 남는다는 것을 인정한다. 그들의 주장은 단지 개별자 동일론은 심성의 인과와 설명에 대한 타당한 출발점을 제공한다는 것이며, 사실 최근 심리철학에서 관심의 초점은 심적 유형의 인과력으로 전이하였다. 그러나 개별자 동일론을 받아들인다고 해도 여전히 의미의 설명적 역할이 문제가 되는 것이 아니고 혹시 개별자 동일론을 받아들이기 때문에 그러한 것이 아닌가? 즉, 의미론적 성질과 독립되어서 심적 사건을 규정할 수 있다는 근거가 미약함에도 불구하고 심적 사건을 내적 물리적 사건으로 환원하였기 때문에 심적 사건이 의미론적 성질을 가짐으로서 어떤 설명적인 공헌을 하는지가 문제가 된 것은 아닌가? 이 추측이 옳다면 개별자 동일론에 의하여 심-물 인과관계를 해명하고 그 다음 심적 유형이나 성질의 설명적인 차이를 논하는 전략은 재고되어야 한다. 이 점을 좀 더 상세히 살펴보자.

인격의 차원에서는 의미적 성질이 심리상태의 설명력에 본질적으로 공헌한다. 즉, 세계가 어떠하다는 재현적인 성질이 그 심리상태의 설명력에 기여하지 않는다면 그 심리상태는 우리가 인격의 차원에서 파악된 상태가 아니다. 철수가 “지붕 위에 모자가 있다”라고 믿는다면 사다리를 타고 지붕으로 올라가는 그의 행위는 “적절하다”. 다음 두 쌍의 문장들을 비교해 보자.

- (1) “철수는 P를 믿는다”, “철수는 -P를 믿는다”
- (2) “철수는 P를 믿는다”, “철수는 P를 믿지 않는다”

(2)의 두 문장은 모순관계임으로 둘 다 참일 수 없다. 지향적 전략에 따르면 우리가 어떤 개체에게 명제태도를 부여할 때 믿음이 전반적으로

참이 되게 하고, 정합성 등의 합리성을 최대한 만족시키도록 해야 한다고 요구한다. 명제태도의 부여는 합리성의 가정 하에 전개된다 - 데이빗슨의 표현을 빌리자면, 명제태도에 의한 이유설명(reason-giving explanation)은 어떤 합리성의 후광(the aura of rationality)을 가지고 있다. 명제태도 부여는 당연히 전체적인 성격(holistic)을 가지게 된다. 즉, 믿음 하나를 놓고 이 믿음이 어떤 개체에 귀속되느냐 아니냐를 묻기보다 이 믿음이 다른 상태 그리고 행동과 합리적인 패턴을 형성하느냐를 물어야 하기 때문이다. 따라서 (1)의 두 믿음은 철수에 대한 사실을 묻기도 전에 선험적으로 서로를 배제한다. 합리성을 가정하고 전개되는 설명적 작업 밖에서는 믿음과 같은 명제태도는 존재하지 않는다. 명제태도는 내적인 연관을 가지고 행동의 설명에 공헌을 하는 상태들이기 때문에 물리적인 과정에 환원될 수 없다.

합리성이 명제태도의 구성적인 원리(the constitutive ideal of rationality)임을 거부하는 한 이유는 그 원리가 믿음과 소망에 대한 실재론과 상충한다는 것이다. 이런 비판에 따르면 (1)의 두 문장은 모순이 아니고 우리가 철수에게 둘 다 부여할 수 있으며 우리의 부여가 맞는지 아닌지는 철수에게 달렸다. 철수 내부의 실제 사정이 어떠하냐에 따라 두 문장 다 참일 수 있다. 물론, 'P'와 '-P'는 모순관계이고 철수가 둘 다 믿는다면 비합리적일 것이나 철수가 실제로 둘을 다 믿느냐 아니냐는 어디까지나 경험적인 사실이고 “모순된 믿음은 존재하지 않는다”는 법칙이 성립하느냐에 의존한다. 그러나 믿음의 체계가 예외 없이 정합적이라 하더라도 여전히 합리적인 행위자는 부재하지 않는가? 우리가 법칙에 의해 항상 정합성을 유지하더라도 이것이 우리가 자율적인 사고와 행위의 주체임을 보여주지 않는다. 그러나 한편으로 합리성이 명제태도를 구성한다면 명제태도가 어떻게 진정한 설명력을 가지는가?

명제태도에 의한 행위설명 성격과 내용에 대한 모란의 한 흥미로운 관찰을 하고 있다.¹²⁾ 모란은 지향적 전략이 행위자의 시점을 무시하고 순전히 삼인칭적 시점에서 하나의 행동예측의 도구로서 명제태도를 부여

하는 것이 아니라 오히려 일인칭 시점에서의 규범적인 고려를 핵심적인 요소로 포함하고 있다고 주장한다. 행위자의 시점에서는 신념, 소망, 의도가 자신의 행위를 예측하는데 도움을 주는 심적 자료라기보다는 바로 자신의 행위를 정당화해주어야 한다. 만일 의도, 신념, 소망 등이 행위자 자신의 행동을 정당화를 해주는 것이 아니라면, 왜 애당초 우리가 합리성을 가정한 예측 전략을 그 사람에게 적용할 수 있다고 기대할 수 있는가? 우리가 서로에게 부여하는 명제태도는 물리적인 법칙과 같이 어떤 법칙에 포섭되는 상태들이 아니라, 바로 그 주체가 따라야 할 규범에 부합하는 것이다. 즉, “내가 어떤 행동을 할 것인가”라는 질문에 대한 답은 “내가 어떤 행동을 해야만 하는가?”라는 질문과 무관하지 않다. 자신이 무엇을 믿고 무엇을 원한다는 사실은 바로 후자를 답하는 것으로 인식되고 따라서 전자에 대한 답이 된다. 만일 철수가 지붕 위의 모자를 찾고자 하는 의도가 단지 자신이 지붕 위로 올라갈 확률을 높이는 증거로 채택한다면 우리는 철수가 과연 의도의 주체인지 한지 의심스러울 것이다. 지향적 상태를 부여함이 단순히 어떤 개체의 행위를 예측하고 설명하고자 하는 관심에 부합하기 위해 우리가 임의로 채택한 전략인지 아니면 우리의 전략과 독립된 실체들인지에 관한 실재론 논쟁은 따라서 문제의 핵심에서 벗어난다. 중요한 점은 지향적인 전략이 이론적인 고려와 함께 행위자 자신의 관심과 시점을 포함하고 있다는 것이고 따라서 철수가 지향적인 상태의 진정한 주체인지는 우리가 그렇다고 믿는데 의존하지 않는다. 그 것은 그야말로 철수 나름이다. 이는 철수가 스스로를 어떤 믿음을 가지고 있다고 믿는다고 해서 그 믿음을 가진다는 말이 아니다. 철수 자신에게 믿음과 소망이 무엇인가, 즉, 믿음과 소망이 부여되는 주체의 “내적”인 문제이다.

자신의 믿음이 행동을 정당화해주는 것으로 인식되려면 그 믿음의 내용이 알려져야 되고 자신이 무엇을 믿는지 안다는 것은 그 믿음의 내용

12) Moran (1994), pp.171-172.

이 참입을 받아들이는 것이다. 삼인칭 시점에서 그 사람의 믿음을 설명적인 작업에 도입할 때 우리가 그의 믿음을 참으로 받아들일 필요는 없다. 비록 철수가 오판을 하여 모자가 지붕 위에 있다고 잘못 믿고 있다 해도, 즉 그의 믿음이 거짓이라고 우리가 생각하더라도, 지붕 위로 올라가는 그의 행위는 똑같이 잘 설명된다. 반면에, 행위자의 시점에서선 믿음의 파악이나 부여가 그 내용의 진리치와 독립되어 혹은 믿음을 참으로 만드는 세계의 상황과 독립되어 진행될 수 없다. 만일 어떤 사람이 자신의 세계관과 아무 상관없이 자신의 행위와 정합적인 믿음의 집합을 자기에게 부여하는데 만족한다면 과연 그가 믿음의 주체인지 의심스럽다. 믿음이 참일 증거가 바로 자신이 그 믿음을 가지고 있다고 믿을 증거가 되어야 한다 - “나는 지붕 위에 모자가 있다고 믿는가?”라는 질문은 “지붕 위에 모자가 있나?”라는 질문에 답하는 순간 답해진다. 그렇지 않을 경우 자신의 믿음이 자신의 행위에 정당성을 부여해 줄 수 없다. 왜 지붕 위에 올라가는가? 이에 대해 철수는 자신의 내부를 지칭하는 것이 아니고 지붕 위를 가리킬 것이다.

지향적 전략의 시점에서 전제된 합리성이 행위자의 시점에서 요구된 것으로 봄으로써 명제태도가 본질적으로 주관적인지를 보이고자 하는 모란의 시도를 간략히 소개하였다. 지향적 전략과 해석이론이 지나치게 삼인칭적 시점에 경도되어 있으며 일인칭 시점에 의해 보완되어야 하고 그 이론들은 실제로 일인칭 시점을 필요 불가결하게 포함한다는 주장이다. 일인칭적 시점을 중요하게 고려하지 않는 데넷의 입장에서 모란의 설명이 어떻게 받아들여질 지 궁금하며, 모란의 설명이 받아들여진다고 해도 과연 데넷의 지향적 전략이론을 반실재론, 도구주의비판으로부터 구제할 수 있을 지도 더 논의가 되어야 할 것이다. 필자의 견해로는 전략개념에 대한 모란의 옹호와 재해석이 가지는 중요한 점은 의미론적 성질이 어떻게 행동설명에 본질적으로 개입하는 지에 대한 한 설명을 제공한다는 것이다. 일인칭 시점에서는 행동 설명이 의미적 성질에 내적으로 의존한다. 즉, 행동이 의미적 성질을 가진 심리상태로부터 법칙에 의해 도출됨으로

서 그 행동이 설명되는 것이 아니라, 그 행동이 그 것의 원인이 되는 심리상태의 재현적인 내용에 의해 정당화됨으로서 그 행동이 설명이 되는 것이다. 물론 모란의 설명이 가능한 유일한 설명이 아닐 수 있다. 그리고 모란의 설명이 비판의 여지가 없는 지는 더 논의해야 될 것이다. 그러나 만일 모란의 설명을 받아들인다면 명제태도의 내용과 설명력을 분리시킬 수 없다는 것은 더 이상 설명을 요구하는 신비한 사실이 아니다. 이는 중요한 성과이다. 왜냐하면, 명제태도를 의미론적 성질과 독립된 어떤 내적 상태로 규정하고 그 다음 그 내적 상태가 바로 의미론적 성질 때문에 어떤 설명력의 차이를 가져오는지 확보하려는 시도가 아무런 동기가 없음을 보여주기 때문이다. 그러한 시도는 마치도 명제태도의 행동조절에 관한 역할이 내용과 독립되게 규정될 수 있는 것처럼 가정하기 때문이다. 의미론적 성질과 설명력의 분리를 그대로 두는 것은 오로지 일인칭 시점이 손상된 상황에서나 가능하다. 명제태도를 일인칭 시점과 분리시켜 삼인칭적/설명적인 시점에서 상정한 내적인 상태로 동일화시킨 다음 그 내적 상태가 다시 의미론적 성질을 가짐으로서 어떤 설명적인 차이를 가져오는 지 묻는 것은 명제태도에 의한 설명이 본질적으로 일인칭 시점을 포함한다는 사실을 무시한 것이다. 물론, 어떤 믿음이 “밖에 비가 온다”라고 기술하는 내용을 가진다는 것과 그 믿음이 우리로 하여금 어떤 행동을 하게 하는 성향을 가진다는 것은 구분되어야 한다. 전자는 세계를 재현하는 역할이고 후자는 행동설명에의 역할이다. 삼인칭적 시점에서는 한 믿음이 참이건 거짓이건 동일한 설명력을 가지는 반면 일인칭 시점에서는 진리치에 대한 고려와, 혹은 그 믿음이 참일 증거와, 독립되게 생각할 수 없다는 것이다. 바로 이런 의미에서 믿음이란 그것의 의미론적 성질이 그것의 설명력에 본질적인 공헌을 한다.

지금까지 개별자 동일론을 거부하는 다른 틀이 바람직함을 보이고자 했다. 개별자 동일론의 거부가 단지 심적 사건이 행위의 원인이라는 믿음을 아무런 정당화 없이 그냥 주장할 뿐이라는 비판은 적절하지 않다. 물론 데이빗슨은 심성의 무법칙성을 주장하면서도 심적 사건을 물리적

사건으로 동일화 하고자 한다. 그러나 데이빗슨을 비판하는 철학자들은 비록 무법칙적 일원론이 심적 사건이 원인이 될 수 있음을 허용하나 그 대가로 심적 유형이나 성질이 인과적으로 아무런 차이를 가져오지 않게 된다고 주장한다. 이에 대해 데이빗슨은 그의 무법칙성이론이 유형부수 현상론을 함축하지 않는다고 반발하며 수반개념에 의존하여 유형 부수현상론을 배격하고자 한다. 데이빗슨의 해결책이 성공적인지는 논란의 여지가 있다. 그러나 필자는 개별자 동일론을 논증하고자 하는 데이빗슨의 시도가 애당초 분명한 동기를 가지고 있는지에 대해 회의적이며 데이빗슨에 대한 비판은 바로 개별자 동일론이 문제가 있음을 드러낸다고 생각한다. 개별자 동일론을 거부하는 이유는 바로 심적 사건이 심적 사건으로서 가지는 설명력을 존중하기 위함이다.

데이빗슨은 그를 비판하는 철학자들과 함께 심적 사건이 행위의 원인이라고 믿으면서 개별자 동일론을 거부할 수는 없다는 생각을 공유하는 것 같다. 개별 심적 사건이 물리적인 속성을 결여하고서 물리적 사건의 원인이 될 수 없다고 생각하는 철학자들은 예를 들어 물리세계의 인과적 폐쇄성(causal closure of the physical world)에 대한 고려를 그 이유로 제시한다. 동일한 사건에 대한 별개의 두 원인들이 있다면 하나의 원인이 다른 원인을 배제하게 되면 이를 방지하자면 심적 원인이 물리적 인과과정에 개입해야만 하는데 이는 물리세계의 인과적 폐쇄성 원리를 파기하게 된다는 것이다. 따라서 유일한 해결책은 개별자 동일론 이 된다. 그러나 필자는 동일한 사건에 대한 별개의 두 원인의 존재가 우려하는 만큼 큰 문제를 가지고 있지 않다고 생각한다. 심적 사건이 물리적인 사건과 별개로 행위의 원인이 될 수 있다면, 심적 설명의 자율성을 존중하면서 동시에 심적 속성의 인과적 차이를 허용할 수 있는 길이 열린다. 즉, 개별자 동일론의 거부는 심적 속성을 환원하지 않고 심적 속성의 인과적인 차이를 가능하게 해준다. 만일 심적 유형이나 속성이 인과적인 차이를 가진다면, 바로 인과관계가 성립되는 개별적 사건들의 차이에 있게 되고, 심적 속성을 이 세계로부터 제거했을 때 사건들 사이의 인과관계는 달라

지게 되기 때문이다. 그러나 행위에 대한 별개의 두 원인이 가능함을 옹호하는 것은 하나의 독립된 다른 논문에서 따로 본격적으로 다루어야 한다.

4

이 논문은 김광수의 정신의 자율성에 대한 정의에서 출발하였고 정신의 자율성에 관한 논의를 내용의 설명력에 관한 문제의 틀 내에서 발전시키고자 하였다. 즉, 정신의 자율성이 무엇인지, 그리고 왜 문제가 되는지 명확히 하려고 시도하였으며 또한 넓게 보아 정신의 자율성에 대한 옹호를 제시하려고 하였다. 이 옹호는 비판-2와 대비되어 전제되었으며, 비판-2에서 제기된 문제의식 자체를 의심하는 것이 당연히 받아들여야 할 의무를 기피하는 불성실한 태도의 표현이 아님을 보이려고 하였다. 명제태도에 대한 무비판적 실재론은 명제태도를 자연계를 초월하여 존재하는 어떤 초자연적인 사실로 강요하는 경향이 있다. 만일 우리가 그런 강요에 굴복하게 되면 우리 앞에 놓인 선택은 환원주의, 제거주의, 아니면 부수현상론 등이 될 것이다. 애당초 초자연적인 사실이 강요되지 않음을 이해하는 것이 궁극적으로 심신문제를 해소하는 관건이라고 생각한다. 합리적 행위자가 애당초 세계와 단절된 내부에 갇혀있는 - 따라서 세계 내에 다시 삽입되어야 할 필요가 있는 - 존재라고 보지 않으면, 물리 세계 내에서 정신의 위상을 밝히는 것이 더 이상 형이상학적 난제로 다가오지 않는다.

이 논문은 정신의 자율성과 유물론적 세계관과의 갈등에 중점을 두었으나 물론 그 문제는 내용 외재론, 일인칭 시점의 권위, 현상적 의식 등 많은 철학적 현안들과 겹치는 부분이 있다. 비록 이 논문에서 연관된 주제들을 본격적으로 다룰 수는 없었으나 김광수의 지적대로 정신의 자율성에 대한 올바른 이해가 철학의 중심적인 과제임을 충분히 보여주었다고 생각한다. 그러나 필자는 김광수가 언급한 많은 철학자들이 정신의 자율성을 무시했다고 보지 않으며 오히려 그 반대로 그들은 정신의 자율

성을 (도대체 정신의 자율성의 본질이 무엇인지, 그것이 물리주의와 공존 가능한지, 등) 자신들 철학의 핵심 문제로 심각하게 받아들였음을 의심하지 않는다. 단지 불행하게도 진지함이 깊은 철학적 문제에 대한 잘못된 사고를 방지할 충분조건은 아니며 사실 이에 대해 누구도 예외라고 장담하기는 어려울 것이다.

참 고 문 헌

- 김광수 (2003), “심리철학과 정신의 자율성”, 『철학적 분석』 7호, 한국분석철학회, pp.1-27.
- Baker, L. (1994), “Instrumental Intentionality”, in *Mental Representation: A Reader*, edited by S. Stich and T. Warfield, Oxford, England: Blackwell, pp.332-344.
- Dennett, Daniel (1987), *The Intentional Stance*, Cambridge, Ma: MIT Press.
- Dretske, Fred (1988), *Explaining Behavior: Reasons in a World of Causes*, Cambridge, Mass.: MIT Press.
- Fodor, Jerry (1987), *Psychosemantics*, Cambridge, Mass.: MIT Press.
- _____ (1990), *A Theory of Content and Other Essays*, Cambridge, Mass.: MIT Press.
- Moran, Richard (1994), “Interpretation Theory and The First Person”, *The Philosophical Quarterly* 44, No.175.
- Searle, John (1984), *Minds, Brains and Science*, Cambridge, Mass.: Harvard University Press.
- Jaegwon Kim (1990), “Explanatory Exclusion and the Problem of Mental Causation”, in *Information, Semantics, and Epistemology*, edited by Enrique Villanueva, Basil Blackwell, pp.33-56.

Mental Autonomy and Mental Causation

Bo-Suk Yoon

Materialistic theories of mind have been criticized on the ground that they can't preserve the autonomy of the mental. But, to accept some form of non-materialistic or non-reductive account of the mental does not solve all the problems. It is argued that the problem of mental causation plagues various forms of non-reductive theory. The dilemma is that if we try to save mental autonomy, mental causation becomes mysterious; we can solve the problem of mental causation, but, only at the cost of losing the autonomy of the mental. I do three things in this paper. First, I discuss a recent paper of Prof. Kim Kwang Su for the purpose of introducing and clarifying the issue. Second, I describe and compare three criticisms made against the representational theory of mind. Finally, I pursue the third criticism further in order to develop a satisfactory way out of the dilemma.

[Subject] Philosophy of Mind

[Key Words] Mental Causation, Autonomy, Computation, Anti-Reductionism,
Kwang-Su Kim