

## 합리적 선택이론에 있어서 합리성의 개념\*

정 준 표  
영남대학교

### • 요약 •

본 논문은 철학적·규범적 의미의 합리성의 개념을 합리적 선택이론 내에서의 합리성의 개념과 비교하여 합리성의 개념에 관한 혼란을 일소하는데 도움이 되고자 한다. 합리적 선택이론에서의 합리성은 도구적 합리성이며, 다분히 정의의 문제이며, 현실의 제한되고 불완전한 합리성을 단순화한 합리성이며, 많은 경우 현상의 설명에 그 자체만으로는 불충분한 개념이다. 합리적 선택이론 내에서 합리성은 기대효용 극대화, 베이지의 규칙, 게임의 균형 개념에 반영되어 있다. 합리적 선택이론이 현실에 대한 설명력을 가지기 위해서는 무엇보다도 행위자가 주관적으로 인식하는 상황을 모델 속에 적절히 반영해야 하며, 이는 모델을 구성하고 해를 구하는 능력도 필요하지만 무엇보다도 문제의 상황에 대한 깊은 지식을 필요로 한다는 것을 의미한다.

주제어 : 합리적 선택이론, 게임이론, 합리성, 제한된 합리성, 기대효용, 합리적 선택 제도주의

## I. 서론

이 글은 철학적·규범적 수준에 있어서 합리성의 개념을 구체적인 실제 연구물에서의 합리성 개념과 대비함으로써 합리적 선택이론에 있어서의 합리성의 개념에 대한 혼란을 일소하는데 어느 정도 기여하고자 한다.

합리적 선택이론의 구체적 적용은 다양한 분야에서 이루어졌는데, 그러한 연구들은 어느 정도 순수 논리적·연역적이냐에 따라 형식이론(formal theory)에 가까운 것부터 그렇지 않은 것으로 분류될 수 있다.<sup>1)</sup> 형식이론에 가까울수록 행위자의 목적은 그러한 가정

\* 이 논문은 학술진흥재단의 공모과제 연구비에 의하여 수행되었음(과제번호 1996-002-C0683). 본 논문의 일부는 한국정치학회 1997년도 연말학술회의, 1999년 1월의 합리적선택연구회 모임, 그리고 한국정치학회 2001년도 하계학술회의에서 발표되었음. 당시 유익한 논평을 해 주신 토론자들에게 감사 드리며, 본 논문이 합리성의 개념에 대해 이전 판(version)과는 좀 다른 주장을 하고 있음을 아울러 밝힌다.

의 현실성의 증명 없이 특정화되는 경향이 있고, 형식이론에서 멀어질수록 다양한 목적 - 즉 협의의 합리성을 넘어서는 - 을 가정하는 경향이 있어 합리성과 적절한 이유(adequate reason)의 구분이 모호해질 가능성이 많아지며, 엄격한 논리적인 전개가 아니기 때문에 다양한 여러 가능성 중 특정 가능성을 합리적 선택의 결과라고 주장하는 경우가 많다. 이 글에서는 합리적 선택이론의 구체적 적용을 다룸에 있어 형식이론에 가까운 연구결과를 주로 논의의 대상으로 한다.<sup>2)</sup> 따라서 합리적 선택이론이 얼마나 경험적 적실성을 가지고 있는가에 대한 논의 자체는 이 글의 직접적 목적은 아니다. 다만 형식이론을 적용한 연구성과물의 구조를 살펴봄으로써, 합리적 선택이론의 적용이 경험적 성공 혹은 실패를 거두었다면 그 원인이 무엇인지를 간접적으로 파악할 수 있을 것이다. 합리적 선택이론의 구체적 적용을 살펴봄으로서 얻을 수 있는 또 하나의 성과는 과연 합리적 선택이론에 대한 다양한 비판이 과연 타당한가를 검토해 볼 수 있다는 것이다.<sup>3)</sup>

이 글의 구성은 다음과 같다. 제2장에서는 철학적·규범적 차원에서의 합리성의 개념을 합리적 선택이론과 관련하여 도구적 합리성을 중심으로 살펴보고 이를 경험적 연구에서의 합리성의 개념과 대비해 본다. 제3장에서는 합리적 선택이론의 구조를 살펴보고, 합리적 선택이론에서의 합리성의 개념을 이론 혹은 모델 내에서의 합리성으로 규정하고 합리적 선택이론의 합리성 개념의 한계를 게임이론의 한계와 연관시켜 설명한다. 제4장은 이러한 합리성의 한계와 연관시켜 최근의 신제도주의의 등장을 설명하고, 합리적 선택이론의 유용성과 한계에 관한 간단한 결론을 제시한다.

1) 오스트롬은 합리적 선택 접근(approach), 전통(tradition) 혹은 틀/framework)을 합리적 선택 이론(theories) 및 모델(model)과 구분한다. 그녀에 따르면, 합리적 선택 접근(전통 혹은 틀)은 방법론적 개인주의(methodological individualism)에 근거하고, 개인은 행동의 전략을 선택하기 전에 그 행동의 기대 이득과 손실을 비교한다는 가정을 하는 모든 연구를 의미하고, 합리적 선택 이론은 개인의 선택에 관련된 정보, 가치 평가, 계산의 유형에 대한 연구자의 보다 구체적인 가정들을 의미하고, 합리적 선택 이론의 모델은 한 이론에 관한 특정한 형식적 재현(formal representation)을 의미한다(Ostrom, 1991). 즉, 이론들이 모여 전체로서 합리적 선택 접근을 구성한다는 것이다. 물론 합리적 선택 접근의 정의에 관해서 필자는 오스트롬과는 좀 다른 견해를 제시하겠지만, 접근·이론·모델에 관한 그녀의 분류에 따르면 이 글에서 '합리적 선택이론'이란 합리적 선택이론들(rational choice theories)이 모여 구성하는 합리적 선택 접근을 의미하고, '구체적 적용'이 그녀의 이론 및 모델을 의미하고, '형식이론'이 그녀의 모델에 해당된다. 아울러 필자는 이론과 모델을 엄격하게 구분하지 않음을 여기서 밝힌다.

2) 그린과 샤피로는 합리적 선택이론의 전형적인 특징을 형식적, 수학적 논리전개에서 찾는다(Green and Shapiro, 1994, p.10).

3) 여기서 구체적 적용이란 합리성에 대한 철학적·규범적 논의와 상반되는 의미로 사용되었다. 그렇다고 하여 구체적 연구성과물을 하나 하나 개별적으로 검토하는 것이 아니라 전형적 예를 통하여 일반적 특징을 찾고자 한다.

## II. 합리성의 개념: 철학적 · 규범적 차원과 경험적 연구의 차원

합리적 선택이론에서의 합리성(rationality)의 개념은 상당한 논쟁을 불러일으키는 개념이다. 우선 인간이 (그 합리성의 개념을 무의미할 정도로 약화시키지 않는 한) 합리적이라는 가정은 많은 실제적 상황에 들어맞지 않는다는 것은 (필자가 생각하기로는) 합리적 선택이론가들도 부인하지 않는 사실이다. 단지 인간이 합리적이라는 가정의 현실성이 인정될 수 있는 여지가 많고, 또 이러한 가정이 방법론적으로 다른 대안적 가정보다 유익한 결과를 준다고 판단되기 때문에 합리적 선택이론에서는 인간의 여러 측면 중 합리성의 측면을 특히 강조하는 것이다.<sup>4)</sup> 그러면 이러한 합리적 선택이론에서 합리성의 개념은 구체적으로 어떻게 파악될 수 있는가? 이하에서는 먼저 철학적 · 규범적 논의에서의 합리성의 개념에 대해 논의하고, 합리적 선택이론에 있어서는 이러한 합리성의 개념이 어떻게 굴절되어 나타나는가를 살펴보고자 한다.

철학적 · 규범적 차원에서의 합리성의 논의는 인간이 어떻게 행동하는 것이 가장 바람직한가에 관한 것이라고 할 수 있다. 물론 규범적 차원의 논의라고 해도 현실의 인간이 전혀 따르지 못할 권고를 할 수는 없으므로, 경험적 차원의 논의와 밀접한 관련을 가지지 않을 수 없다. 이런 측면에서 어떻게 행동하는 것이 합리적이냐에 대한 논의는 우리가 무엇을 어떻게 해야 하느냐에 관한 것뿐만 아니라 우리 자신이 어떤 존재이고 어떤 존재가 되어야 하는가에 대한 분명한 대답이 나올 수 없는 끊임없는 논의라고 볼 수 있다. 여기서 이러한 철학적 · 규범적 차원의 합리성의 개념 중 합리적 선택이론의 핵심이라 할 ‘도구적 합리성’(instrumental rationality)만을 엘스터(Elster)의 논의를 중심으로 살펴보고자 한다.<sup>5)</sup> 여러 가지 합리성의 개념이 존재한다는 것은 익히 알려져 있다.<sup>6)</sup> 하지만 합

4) 프리드만(Friedman, 1953)의 도구주의(instrumentalism) 견해와 가정 자체에 어느 정도의 현실성이 있어야 한다는 “현실주의자(realist)”의 견해 중 필자는 현실주의에 속한다고 할 것이다.(사실 프리드만이 도구주의자라는 주장 또한 논란의 대상이 된다. 프리드만이 “이론은 그 가정의 현실성이 아니라 그 이론이 설명하고자 하는 현상들에 대한 예측력에 의해 평가되어야 한다”(p.8)고 주장하긴 했지만 한편 동일한 논문에서 “이론 속의 가정들에 대한 적절한 질문은 그 가정들이 서술적으로 ‘현실적’(realistic)이냐 아니라, 왜냐하면 가정들은 결코 그럴 수가 없으므로, 당장의 목적에 충분한 근사치(approximation)이냐 이다(p.15)”라고도 말하고 있기 때문이다.) 이론가들이 만들어 내는 인간에 대한 이상형(ideal type)은 분별 있는 현실의 인간이 적어도 때때로는 그 이상형에서 가정되는 방식과 유사하게 행동할 것이라고 생각할 수 있어야 하지 누구도 흉내낼 수 없는 초인적 능력을 가정해서는 안될 것이다(이에 대해서는 Knudsen (1983) 참고). “합리적 선택이론의 기본적인 공리는 인간은 합리적이라는 지극히 단순한 가정이며 공리의 진부에 대해서는 논쟁을 허락하지 않는다”는 견해가 있으나(조기숙, 1996, pp.46-47), 인간이 합리적이라는 가정은 현실 속에 존재하는 인간이 아니라(물론 현실 속의 인간도 상당한 정도로 합리적으로 행동하는 것으로 생각되기에 그렇지만) 개개의 모델 속에서의 행위자에 대한 공리로 생각되어야 할 것이며, 이러한 합리성의 내용은 추상적 · 철학적 수준에서의 합리성과는 상당히 거리가 있음을 본문에서 보일 것이다.

리적 선택이론에서의 합리성의 개념은 도구적 합리성을 중심으로 하고 있다는 것은 거의 異論의 여지가 없다. 합리적 선택이론에서의 합리성은 주어진 목적을 달성하기 위한 최선의 수단을 찾는 것과 관련되는 것이다.

엘스터에 따르면, 행위자의 행위를 합리적 선택으로 충분히 설명하기 위해서는 최적성(optimality), 일관성(consistency), 인과성(causality)의 세 조건을 충족시켜야 한다(Elster, 1986, 1988). 합리성 개념의 구성요소인 행위자의 인지상태 혹은 믿음(belief: B), 일련의 목적 혹은 욕구(desire: D), 믿음의 바탕이 되는 증거 혹은 정보(information: I), 그리고 행위(behavior or action: A)를 중심으로 이를 설명하면 다음과 같다.

최적성은 (1) 행위자의 믿음(B)이 주어졌을 경우, 그의 행위(A)는 욕구(D)를 충족시키는 최적의 수단이다; (2) 행위자의 믿음(B)은 주어진 정보(I)에서 가질 수 있는 최선의 것이다; (3) 수집된 정보(I)의 양은 그의 욕구(D)를 고려할 때 최적의 양이다 라는 3가지 조건으로 다시 나눌 수 있다. 일관성은 믿음과 욕구에 내적 모순(internal contradiction)이 존재하지 않고, 또한 주어진 믿음(B) 하에서, 적절하게 (행위자 자신에 의해) 가중치가 부여된 모든 욕구(D)들을 전체적으로 만족시키는 최선의 행위(A)를 택할 때 충족된다. 전혀 성취할 수 없는 욕구라든지 예를 들어 “내가 비가 오지 않을 것이라고 믿으면 꼭 비가 온다”라고 믿고 우산을 가지고 가는 경우가 내적 모순이 있는 경우라면 의지박약 현상(weakness of the will, akrasia)은 욕구들을 전체적으로 만족시키는 행위를 선택하지 못하는 경우에 해당한다.<sup>7)</sup> 인과성은 행위(A)가 믿음(B)과 욕구(D)에 의해 “올바른 방법”으로 야기(cause)되고, 믿음 역시 정보에 의해 “올바른 방법”으로 야기되어야 할 것을 요구한다. 데빗슨(Davidson)에 따르면 넓은 의미의 행동(behavior)은 의도적 측면을 갖는 의도적 행동(intentional behavior) 즉 행위(action)와 하품 등의 무의식적으로 이루어지는 단순행동(mere behavior)으로 대별될 수 있다. 의도적 행동 즉 행위는 다시

- 
- 5) 합리성의 의미 및 행위자가 실제 합리적인지 아니지가 얼마나 중요한지에 대해서는 합리적 선택이론가 사이에서도 상당한 논란이 있다. 어쨌든 엘스터는 합리성의 의미에 대한 가장 중요한 고려 사항을 제공하고 있다(Levi, 1997).
  - 6) 예를 들어 하그리브즈 힉은 합리성을 도구적(instrumental), 절차적(procedural), 표현적(expressive) 합리성으로 분류하고 도구적 합리성의 한계를 지적하고 있다(Hargreaves Heap, 1989). 하그리브즈 힉의 절차적 합리성 개념은 주로 인간이 완벽한 계산 능력을 갖추지 못했다는 인식에서 온 사이몬(Simon, 1985)의 절차적 합리성과는 달리 인간이 도구적 합리성을 넘어서는 사회적·역사적 존재라는 인식을 바탕으로 한 개념이다. 이 글의 주장과는 다소 상충되는 점도 있지만, 여러 다양한 합리성 개념에 관해서는 최종원(1995)도 참고할 수 있다.
  - 7) 금연(X)하겠다고 생각하고는 실제 흡연(Y)하는 것이 이에 해당한다. 모든 것을 고려해 X가 최선의 선택이라고 생각하고는 막상 선택의 순간이 오면 그러한 더 나은 판단에 따르지 않는 것이다. 그런데, 여기서 행위의 합리성을 평가하는 시간의 범위를 어떻게 잡을까 라는 문제가 생긴다. 합리성은 이 경우 선택의 순간을 기준으로 평가된 것이 아니라 통시적(inter-temporal)으로 평가된 것이다.

의도적 행위(intentional action)와 비의도적 행위(unintended action)로 구분할 수 있다(Davidson, 1979). 여기서 의도적 행위와 비의도적 행위는 행위의 기술(description) 방식에 따라 구분되는 것으로서, 예를 들어 갑을 죽이려고 했는데 사람을 잘못 오인하여 을을 죽였다면, '살인'이라는 기술 하에서는 이 행위는 의도적 행위가 되나, '을을 해침'이라는 기술 하에서는 비의도적 행위가 되는 것이다. 그러나, 의도적 행동 즉 행위는 어떤 기술 하에서 반드시 의도적 측면을 갖는다는 점에서 단순한 행동과 구별된다. 비의도적 행위라고 하여 항상 합리적 선택으로 설명할 수 없는 것은 아니다라는 점은 위의 예에서 을을 갑으로 오인한 믿음이 충분한 이유가 있을 수 있다는 점에서 알 수 있다. 믿음과 욕구가 "올바른 방법"으로 행위를 야기하지 않는 예로서 위기에 처한 암벽 등반가가 밧줄을 놓음으로써 무게를 줄이고 자신의 안전도 도모할 수 있다는 생각 때문에 온몸이 마비되어 밧줄을 놓아 밧줄에 달려 있는 등반가를 죽게 하는 경우를 들 수 있다. 믿음과 욕구가 있었지만 이 암벽 등반가는 밧줄을 놓는 선택을 하지도 않았고 의도적으로 그렇게 하지도 않았던 것이다.

위에서 합리성의 개념을 도구적 합리성을 중심으로 살펴보았는데, 여기서는 욕구(혹은 목적)가 일관성을 갖추어야 한다는 것 외에 그 욕구의 실질적 내용에 대해서는 언급하지 않았다. 행위자가 만족시키고자 하는 일정한 욕구들을 가지고 있고 여기에 일정한 일관성이 존재하는 경우, 선택의 가능한 결과(outcomes)에 대한 행위자의 선호(preference)가 전이성(transitivity), 완전성(completeness) 등의 일정한 특성을 가진다는 말로 대신할 수 있다.<sup>8)</sup> 욕구가 일정한 논리적 일관성을 갖추어야 한다는 형식적(formal) 요구 외에 욕구의 내용이 실질적(substantive)으로 어떠한가에 따라 합리성을 판단할 수도 있을 것이다.<sup>9)</sup> 욕구가 단순히 일정한 일관성을 가질 것을 요구하는 경우 이를 "성긴 합리성"(thin rationality)의 요구라면, "조밀한 합리성"(thick rationality)은 어떤 특정 내용의 욕

8) 전이성은 임의의 세 결과에 대해  $x$ 를  $y$ 보다 선호(무차별한 경우 포함)하고  $y$ 를  $z$ 보다 선호할 경우,  $x$ 는  $z$ 보다 선호되어야 한다는 것이다. 완전성은 모든 임의의 두 결과에 대해 둘 중 어느 것을 선호하는지 혹은 둘 사이에 무차별한 것인지를 행위자가 밝힐 수 있어야 한다는 것을 의미한다. 전이성을 위반하는 것은 비합리적이라고 대체로 간주할 수 있지만 ( $x$ 를  $y$ 보다 선호하는 경우,  $x$ 를 위해 약간의 돈을 지불할 수 있을 것이다. 전이성을 위반하는 경우,  $y$ 에서  $x$ 로 가기 위해, 또  $x$ 에서  $z$ 로 가기 위해,  $z$ 에서  $y$ 로 가기 위해 약간의 계속 지불한다면 들고 돌아 결국은 엄청난 돈을 지불해야 할 것이다.), 완전성은, 이의 위반이 비합리적이기보다는, 전이성과 더불어 최선이 무엇인지를 항상 찾을 수 있게 하는 조건이 된다.

9) 여기서 말하는 실질적·실체적 합리성의 개념은 사이몬(Simon, 1985)의 실질적 혹은 객관적 합리성(substantive or objective rationality) 개념과 구분되어야 한다. 사이몬은 외부적 상황 및 행위자의 제약된 능력을 함께 고려할 때의 합리성을 절차적 혹은 제한된 합리성(procedural or bounded rationality)이라 하고, 이에 비해 상황요소만을 고려한 합리성을 실질적 혹은 객관적 합리성이라 정의한다. 이 글의 목적 중의 하나는 이러한 사이몬의 구분에 따르면 현실의 합리적 선택이론은 대부분 객관적 합리성에 기초한 것이 아니라 오히려 주관적, 제한된 합리성에 기초하고 있다는 것을 보이는 것이다.

구만을 합리적으로 간주하는 것에 해당한다. 합리적 선택이론에서 말하는 소위 “협의 합리성”(narrow rationality)은 조밀한 합리성의 하나로서 물질적이고 이기적인 자기이익(material selfish self-interest)의 추구만을 합리적이라고 가정하는 것이다.

위에서 다른 내용 외에도 합리성의 개념에 대한 많은 논란이 존재한다. 예를 들어 믿음의 합리성에 대해 살펴보자.<sup>10)</sup> 예를 들어 오늘날의 과학적 지식에 비추어 사실이 아닌 믿음도 합리적인가? 일상적 의미에서 우리는 비가 내리도록 하기 위해 기우제를 지내는 행위를 비합리적이라고 생각한다.<sup>11)</sup> 그러나 실제 어떤 사회에서 기우제가 비를 내리게 한다고 믿고 있고, 그 믿음에 근거하여 기우제를 지냈다고 하자. 이러한 행동을 합리적이라고 볼 수 있는가? 믿음-합리성에 대한 해석은 상대주의(relativism) 논쟁과 밀접하게 연관되어 있다.<sup>12)</sup> 우리와 다른 사회를 이해하기 위해서는 이러한 행동도 합리적이라고 불려야 하지 않을까? 주관적 믿음-합리성에 따르면 실제 개인이 가지고 있는 믿음에 따라 행동하는 경우 합리적이라 부를 수 있을 것이다.

이러한 철학적·추상적 수준에서의 합리성 개념이 여러 측면을 포함하고 있고 또한 상당한 논란의 대상이 될 뿐만 아니라, 합리성이 무엇이나에 대한 논란을 일정한 방향으로 해결한다 하여도 (예를 들어 성긴 합리성과 주관적-믿음 합리성을 합리성으로 인정하면서 최적성, 일관성, 인과성을 충족하는 것을 합리적으로 간주함) 이러한 합리성의 개념은 논리적으로 모순을 가질 뿐만 아니라 특정 상황에서는 행위를 설명하기에는 불충분하다. 예를 들어, 의사결정 전에 어느 정도의 정보수집을 하는 것이 합리적이냐의 문제는 합리적 선택의 내적 “개념적 문제”(conceptual problem: Knudsen, 1993)를 드러내고 있다. 즉, 위에 언급된 합리성의 개념으로는 어느 정도의 정보를 수집하는 것이 최적(optimal)이라는 것을 결정하지 못하기 때문이다. 정보의 한계가치가 정보 수집의 한계비용과 일치해야 한다는 것은 모든 정보를 가지고 있는 행위자나 할 수 있는 것이지, 정보를 찾고 있는 행위자가 할 수 있는 것은 아닌 것이다.<sup>13)</sup> 또한 위에서 논한 합리성의 개념은 뒷장에서 자세히 살펴듯이 의사결정의 상황에서는 충분할지 몰라도 ‘상호의존의 결정상황’(interdependent decision situation)인 게임상황에서는 더 많은 제약을 요구한다.<sup>14)</sup>

10) 햄린은 믿음, 욕구, 그리고 결정(decisions, 본문의 행위에 해당)이라는 3 차원과 관련하여 합리성의 개념을 수단-합리성(means-rationality), 믿음-합리성(belief-rationality), 목적-합리성(ends-rationality)으로 나누어 논의하고 있다(Hamlin, 1986, Chapter 2).

11) 기우제의 진정한 목적이 마을주민의 단합에 목적이 있다면 합리적이라고 말할 수 있을 것이다.

12) 이에 대해서는 Hollis and Lukes, 1982 참고.

13) 이 문제를 인지하고 엘스터는 정보수집에 있어서의 최적성의 요구를 주어진 상황에 따라 결정되는 상한선과 하한선의 사이에서 정보를 수집해야 한다는 조건으로 대체하고 있다.

14) 뒤에 언급하겠지만 어떤 상황을 게임으로 볼 것인지 의사결정의 문제로 볼 것인지는 연구자의 선택에 달

이상과 같이 추상적·철학적 수준에서의 합리성 개념을 살펴보았다. 합리적 선택이론에 의한 실제의 연구에서는 그 관심이 주로 최적성의 조건, 그것도 주어진 믿음과 욕구 하에서 최선의 행동을 택했는가에 놓여지게 된다. 엘스터는 합리적 선택이론에 의한 설명이 주로 많은 사람의 행동에 의해 구성되는 사회적·집단적 현상에 관한 것이기에 인과성의 원칙을 검증하는 것은 불가능할 뿐 아니라 의미가 없다고 한다 (Elster, 1986).<sup>15)</sup> 그런데 필자가 생각하기로는 규범적이 아니라 경험적·예측적 연구에서는 행위자의 실제 믿음(belief)과 욕구를 - 그것이 비록 위에서 논의된 바와는 달리 비합리적으로 형성되었을지라도 - 중심으로 최적의 행동을 택했는가에 관심이 놓여질 수밖에 없을 것이다.<sup>16)</sup> 따라서 사회적·정치적 현상을 설명하려는 합리적 선택이론에서의 합리성은 앞에서 설명한 추상적·규범적 의미에서의 합리성과는 상당한 거리가 있을 수밖에 없다. 합리적 선택이론의 실제 연구에 있어서 욕구의 내용 또한 성긴 합리성에서부터 여러 가지 조밀한 합리성까지 개별 연구에 따라 합리성의 개념이 다양하게 정의되어 있다. 예를 들어 개인의 선호가 어떻게 사회적 선호로 집약되는가를 다루는 공리적 사회선택이론(axiomatic social choice theory), 죄수의 딜레마(prisoners' dilemma), 혹은 겁쟁이게임(chicken game) 등의 예에서는 욕구(이 경우, 욕구는 개인의 선호 혹은 효용함수로 대체되어 나타난다)가 실질적으로 어떤 내용을 반영하고 있는 지에 대한 가정이 없다고 할 수 있으므로, 성긴 합리성의 가정만으로도 유용한 결과를 도출한다고 할 것이다. 그러나 대부분의 경우, 특히 경험적 연구에서는 합리적 선택이론의 유용성 및 그 강력함은 상황에 따라 적절하게 행위자의 목적을 조밀하게 설정함으로써 이루어진다(정준표, 1993). 선거연구분야에서 정치가의 목적은 당선의 확률을 제고하는 것이다 라든지, 관료집단의 연구에서는 관료가 예산의 증가를 목적으로 한다는 등의 가정이 바로 그 예라고 할 것이다. 합리적 선택이론의 지금까지의 공헌 중 많은 부분이 선거에 관련된 연구에서 얻은 결과라고 할 수 있는 바, 대부분의 연구에서 위에 언급한 바와 같이 정치가는 당선의 확률을 높이려 한다

려있는 것이다. 대부분의 사회현상은 여러 행위자의 행동이 결합하여 어떤 결과를 만들어내기 때문에 게임이라고 간주해야 하나, 상황의 특성에 따라 의사결정의 문제로 볼 수 있는 것이다.

- 15) 또한 일관성은 대체로 충족되는 것으로 가정된다고 한다. 한편, 의지박약 현상과 같은 보다 더 나은 판단에 따르지 않는 자기 파괴적 행동(self-destructive behavior)에 지금까지보다 더 많은 관심을 보여야 한다고 주장한다.
- 16) 하딘(Hardin, 1996)의 경우, 인종적 갈등 및 폭력을 종족집단의 구성원이 가지고 있는 잘못된 믿음을 중심으로 설명하고 있다. 이러한 '분명히' 잘못된 믿음에 따른 (효율적인) 행동을 과연 합리적이라고 할 수 있는지는 의문이지만, 합리적 선택이론에서 합리성의 개념이 얼마나 다양하게 이해될 수 있는가를 보여주는 좋은 예라고 할 것이다. 물론 합리성의 개념이 이렇게 확장될 경우, 합리적 선택이론의 병리(?) 중 하나가 '사후적 이론 발전'(post hoc theory development)이라는 그린과 샤피로의 비판을 어느 정도 수용해야만 할 것이다(Green and Shapiro, 1994, pp.34-35).

는 가정을 하고 있으나, 실상 이러한 목표는 아주 좁은 의미의 합리성 - 개인의 물질적 이익 -을 넘어서는 개념으로서 타인으로부터 인정받고자 하는 욕구 등의 측면도 포함하고 있다고 볼 수 있다. 또한 경우에 따라서는 비이기적인 규범 - 예를 들어 공동체의 기준(Ostrom, 1990), 공정성(Levi, 1998) 등 -을 따르려는 욕구 역시 포함될 수 있다. 이렇게 볼 때 합리성이라는 개념은 개인의 물질적 이익 추구 외에도 다양한 측면을 포함할 수 있는 신축적 개념으로서 어떤 행동이 합리적이냐 아니냐는 상당 부분 정의(definition)의 문제라고 볼 수 있다.<sup>17)</sup>

### Ⅲ. 합리적 선택이론에서의 합리성

#### 1. 합리적 선택이론의 내용적 구조

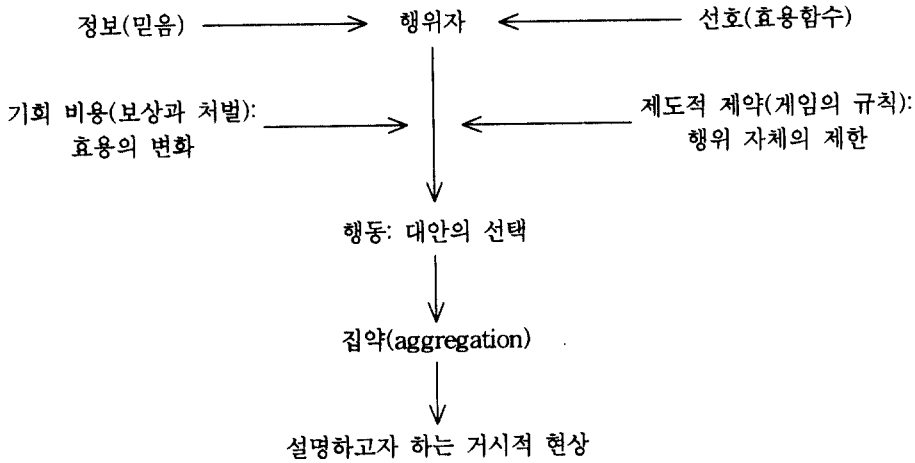
합리적 선택이론은 기본적으로는 의도적 혹은 목적지향적 행위자를 전제로 하고 있다.<sup>18)</sup> 이러한 행위자는 특정의 정보와 아울러 선호, 가치, 혹은 효용을 가지고 일정한 주어진 제약 하에서 대안을 선택하는 것으로 간주된다. 이러한 행위자의 대안 선택, 혹은 행동을 집약(aggregation)하여 記述 혹은 說明하고자 하는 (거시적) 사회현상을 도출하

17) 정의의 문제라고 하여 그 문제시되고 있는 상황에서 실제 행위자가 가지고 있는 욕구 혹은 가지고 있을 것이라고 추정되는 욕구를 무시한다면 잘 된 경험적 연구가 되기는 불가능할 것이다. 한편, 합리성의 개념이 이렇게 다양할 수 있다면, 지역주의 투표가 과연 합리적이냐 아니냐에 대한 최근의 논쟁은 필자의 생각으로는 그렇게 유익하지 않은 것 같다. 지역주의 투표를 두고 자기 지역 출신의 대령을 뽑아야 재화의 분배나 인사에서 (내가) 이득을 볼 것이라는 유권자의 계산에서 비롯되었으니 합리적이라는 논의(조기숙, 1997)는 무엇을 합리적이라고 보는가에 크게 달려 있다. 문제는 유권자가 이러저러한 동기로 선택을 하니가 합리적이다가 아니라 과연 이러저러한 동기에서 선택하는가란 경험적 검증에 있는 것이다. 지역주의 투표가 과연 개인적 혜택을 가져올 수 있는지를 주로 공공재 이론을 중심으로 비판적으로 검토한 강원택(2000)의 논문 역시 경험적 검증을 강조한 것이라고 해석할 수 있다. 그런데 지역출신이 권력을 장악하여도 실제로 개인적 혜택을 받는 사람이 거의 없더라도(지역주의 투표를 설명하는데 있어 합리적 설명은 지역출신의 권력장악이 '빠짐없이' 그리고 '자동적'으로 전환되는 경우에만 성취된다는 강원택의 논의는 좀 너무 강한 강조라고 생각된다. 상당히 많은 사람에게 개인적 혜택이 되면 지역주의 투표 현상을 설명할 수 있기 때문이다.) 지역주민은 그렇게 믿을 수 있고 그에 따라 (도구적으로) 합리적인 행동을 했다고 주장할 수도 있을 것이다. 이 경우 문제는 이것이 합리적이냐 혹은 그렇지 않느냐가 아니라 정말 그렇게 믿었느냐 이다. 특히 그러한 믿음과 행동과의 연계가 너무나 직접적이고 눈에 뵈할 때, 다른 경험적으로 확인되는 현상이 도출되지 않는 경우, 그러한 믿음을 가졌다는 가설은 직접적으로 확인되어야 할 것이며, 더 나아가 다른 대안적 가설보다 더 설득력이 있다는 증거를 제시하여야 할 것이다(cf. Chong, 1996).

18) 여기서 기본적이라는 말을 쓴 이유는 모델에 따라 행위자가 비합리적 행동을 할 가능성을 포함하여 분석할 수도 있기 때문이다. 예를 들어 불완전정보하의 게임모델에서는 종종 행위자가 기계적으로 행동하는 유형일 가능성을 포함하고 있다.



는 것이 합리적 선택이론에 기반한 분석 방법이라고 할 수 있다. 아래 그림 1은 이를 도식적으로 표현한 것이다.<sup>19)</sup>



〈그림 1〉 합리적 선택이론의 도식적 구조

그림 1에서 나타난 합리적 선택의 제 요소 중 어느 부분이 개별 연구에서 핵심적 역할을 차지하는가는 개별 연구에 따라 다르게 나타난다. 예를 들어 셸링(Schelling, 1978)의 인종간의 주거분리 현상 및 죄수의 딜레마는 개인들의 선호의 배열이 가장 핵심적 요소이다.<sup>20)</sup> 제도적 제약을 강조하는 연구는 각종 선거제도의 연구, 의회 규칙의 영향 연구 등에서 잘 나타난다.<sup>21)</sup>

합리적 선택이론은 사회현상을 행위자들의 의식적 행동의 특정한 배합의 결과로 파악하는 방법론적 개인주의(methodological individualism)에 입각해 있다(황수익, 1989). 이

19) 유사한 논의로는 Friedman and Hechter(1988) 참고.

20) 죄수의 딜레마의 경우 실제 연구의 대상이 되는 상황은 오히려 제도적 제약 혹은 기회비용에 중점이 놓여질 수 있다. 그러나 이러한 것을 감안하여 이를 게임으로 표현했을 때, 죄수의 딜레마로 나타날 수 있다. 이렇게 실제 상황과 단절시켜 죄수의 딜레마를 볼 경우 선호의 배열이 딜레마를 가져온 것으로 볼 수 있다.

21) 특정 요소를 강조한 연구성과물의 예에 대한 보다 상세한 논의는 Friedman and Hechter(1988) 참고. 레비는 합리적 선택이론의 핵심은 행위자의 합리적 선택이라기보다는 오히려 그림 1의 기회비용에 해당하는 '희소성'(scarcity)과 그림 1의 게임의 규칙에 해당하는 '제도'(institution)이라는 두 요소로부터 오는 행위의 제약이라고 하면서(Levi, 1997), "제한된 선택"(choice-within-constraint; Ingram and Clay, 2000)에서 선택보다는 제약의 요소를 강조하고 있다. 제약의 요소를 너무 강조할 경우에 나타나는 위험성에 대해서는 후술할 것임.

러한 행위자는 자신의 효용을 극대화하는 것을 목표로 하고 있으므로, 이러한 개별 행위자의 행동의 집약결과인 거시적 현상은 대개의 경우 행위자에 의해 의도되지 않는 결과이다. 즉, 거시적 현상은 행위자의 의도를 넘어선 차원의 것이다(김용학, 1996, p.71). 물론 행위자는 어떤 결과가 나타날 것인지에 대해 어느 정도 인식은 하고 있을 것이다. 예를 들어 '죄수의 딜레마'에서는 비효율적인 결과가 초래될 것임을 추론 능력을 갖춘 합리적 행위자는 알 수 있다. 이 경우 행위자가 비효율적인 결과를 가져오려고 '의도'하였다고 할 수는 없을 것이나, 비효율적인 결과가 발생할 것은 알고 있다(그러나 게임의 구조상 이것을 피할 수 없다). 합리적 선택이론의 최대의 강점은 이 같이 거시적 결과(macro outcome)의 미시적 인과기제(micro causal mechanism)를 제공해준다는 데서 찾을 수 있다. 개별 합리적 선택이론에 의한 연구를 평가하는 데 있어 가장 중요한 요소의 하나가 바로 개인의 행동과 사회적 결과가 어떻게 연결되었는가 이다. 이 연결 과정이 너무나 직접적일 경우, 그리고 전혀 예기치 못한 결과가 없을 때, 미시적 기제를 제공한다는 의미에서의 합리적 선택 이론의 가치가 떨어질 수밖에 없다. 셸링의 주거분리 모델은 개인이 자기 이웃에 인종적으로 다른 사람의 비율이 일정 한도를 넘어서면 이사를 간다는 약한 가정 하에서 인종적으로 날카롭게 분리된 주거 형태가 나타난다는 것을 보여준다. 또한 죄수의 딜레마와 이와 연관된 집단행동의 논리(Olson, 1965)는 공통의 이익이 존재하더라도 이것이 항상 성취될 수 없다는 것을 보임으로써, 과거의 민중 봉기, 혁명, 이익집단의 활동 등의 현상에 대한 생각을 현저하게 변화시켰다. 형식이론 특히 게임이론을 이용한 연구에서는 개인의 행동과 (설명하고자 하는) 균형 결과 사이의 연계가 그렇게 명백하지 않다.<sup>22)</sup> 약간의 비합리성의 존재 가능성이 유한 반복된 죄수의 딜레마에서 일정한 협력을 창출해 낼 수 있다는 결과는 전혀 상식적이지 않으며, 또한 평판(reputation)의 중요성을 분명하게 보여준다.<sup>23)</sup> 친한 사람으로부터의 충고는 받아들여지 사이가 나쁜 사람의 충고는 무시하는 상식적 행동은 신호게임(signalling game)이 그 메커니즘을 분명하게 보여준다. 비형식적 합리적 선택이론을 사용하는 경우 인과기제를 분명하게 보여주지 못하거나 논리적으로 타당하지 않은 결론을 도출하기 쉬울 뿐만 아니라, 개인의 합리적 행동과 설명하고자 하는 현상과의 관계가 너무 직접적이라는 문제를 흔히 가진다. 이 경우 개인의 행동이 정말 그러했는가란 경험적 검증의 문제가 더욱 부각된다.<sup>24)</sup>

한편, 그림 1의 도식은 콜만의 논의에 따라 '거시에서 미시'(macro-to-micro), '개인에

22) 물론 어느 정도 명성 있는 학술지에 실린 연구성과물을 중심으로 말할 때 만이다.

23) 죄수의 딜레마와 관련한 논문은 Kreps et al.(1982), 그리고 같은 논리를 소위 '연쇄점 역설'(chain store paradox)에 응용한 논문은 Kreps and Wilson(1982)을 참고할 수 있다.

24) 주 17의 논의 참고.

서 행동'(individual-action), '미시에서 거시'(micro-to-macro)의 3 요소로 구성된 것으로 파악할 수 있다(Coleman, 1990, Chapter 1). 이런 측면에서 보면, 전체 이론이 다를 수 있을 정도로 간단해야 한다는 측면에서, 또 '의도적 행동'(purposive action)의 의미를 정확히 한다는 의미에서 합리적 선택이론의 '개인에서 행동'이라는 구성요소는 될 수 있는 한 단순해야 하고, 이런 이유에서 합리적 선택이론에서는 효용극대화를 가정한다고 볼 수도 있다(Coleman, 1990: pp.18-19). 이렇게 합리적 선택이론을 파악할 경우, 합리적 선택이론 내의 합리성은 현실의 불완전한 '합리성'을 단순화한 것이라고 볼 수 있다.

## 2. 합리적 선택이론의 형식적 구조

합리적 선택이론의 구조에 대한 지금까지의 논의가 합리적 선택이론에서 합리성의 의미를 그렇게 분명히 하지 못했다면, 합리적 선택이론의 구조를 의사결정이론(decision theory)과 게임이론(game theory)의 시각에서 살펴보는 것은 합리성의 의미를 분명히 할 뿐 아니라, 그 한계를 어느 정도 밝힐 수 있다. 엘스터는 합리성을 '모수적 합리성'(parametric rationality)과 '전략적 합리성'(strategic rationality)으로 구분하는데(Elster, 1986), 모수적 합리성은 의사결정의 상황에서 나타나는 합리성이고, 전략적 합리성은 한 행위자의 행동뿐만 아니라 다른 행위자(들)의 행동도 그 상황의 결과에 영향을 미치는 상호의존의 결정상황(interdependent decision situation), 즉 게임상황에서 나타나는 합리성이다. 합리적 선택이론가는 현실의 관심문제를 특정한 의사결정상황 혹은 게임상황으로 추상하여 모델을 구성한다. 개개의 합리적 선택이론은, 비록 그것이 비형식적(informal) 이론이라 하더라도, 풀어야 할 특정 의사결정 혹은 게임이론적 문제를 가지고 있다고 볼 수 있는 것이다. 문제를 푸는 것은 실제로는 연구자이지만, 모델 속의 행위자가 그 문제를 푸는 것으로 간주된다. 문제를 풀기 위해서는 문제가 풀릴 수 있는 성격을 갖추어야 하고, 모델 속의 행위자 역시 문제를 풀 능력을 갖추어야 한다. 따라서 주어진 문제의 틀 내에서 문제를 풀기 위해서는 행위자의 목표(선호)가 일정한 일관성을 가지고 있어야 하고, 충분한 계산능력을 갖추어야 하므로 행위자는 고도의 합리성을 가지고 있는 것으로 간주된다. 예를 들어 험린은 신고전경제학(neoclassical economics)의 대부분의 연구가 욕구의 형식적·실질적 내용과 관련하여 개인의 선호가 (1) 반사성(reflexivity), (2) 완전성(completeness), (3) 전이성(transitivity), (4) 연속성(continuity), (5) 불포만성(non-satiation), (6) 볼록성(convexity)을 만족하며, 모든 욕구 혹은 '이익(interest)'이 효용(utility)이라고 불리는 동일 단위로 계량화될 수 있다(commensurability)는 가정을 한다

고 한다(Hamlin, 1986).<sup>25)</sup> 이러한 가정들은 미분 등 수학적 도구를 활용하여 의사결정의 상황에서 (효용)극대화의 문제를 풀기 위한 것이라고 볼 수 있다.

의사결정 상황은 게임 상황에서 게임의 참가자(player)가 소위 '자연'(nature)을 제외하고는 한 명뿐인 경우로 생각할 수 있으므로,<sup>26)</sup> 이하에서는 합리성이 문제의 틀, 즉 모델 속에서의 합리성이라는 것을 게임이론을 사용하는 경우에 나타나는 일반적인 특징을 들어 살펴보자. 주어진 현상을 게임으로 구성하기 위해서는 우선 게임의 참가자, 그들의 선호(효용함수), 선택 가능한 대안, 나타날 수 있는 결과, 가지고 있는 정보, 게임의 진행 순서 등등을 우선 설정해야 한다. 즉, '게임의 규칙'(rules of the game)을 상정해야 한다. 이러한 현실에 대한 추상(abstraction)이 끝난 후에 나타나는 모델 속의 인간은 고도의 합리성(즉, 단순성)을 가지게 된다. 우선 참가자의 선호는 '기대효용이론'(expected utility theory)이 적용될 수 있도록 전이성, 완전성을 만족하는 등의 충분한 일관성을 갖추어야 하며,<sup>27)</sup> (기대)효용을 계산하고 베이즈의 규칙(Bayes' Rule)에 따라 새로운 정보를 처리하여 믿음을 수정할 수 있는 고도의 계산력을 갖춘 존재가 되는 것이다. 게임이론적 모델에서의 참가자의 합리성은 게임이론적 분석의 가장 중요한 도구라고 할 수 있는 '균형'(equilibrium) 개념에 궁극적으로 반영된다. 여러 균형 개념 중 가장 기본적인 개념이 '내쉬균형'(Nash equilibrium)인데, 내쉬균형에서는 모든 게임의 참가자들이 다른 모든 게임의 참가자들의 전략에 대해 자신의 기대효용을 극대화하는 전략을 선택하고 있다.<sup>28)</sup> 내쉬균형을 포함한 여러 균형개념이 게임에서 어떤 결과가 나타나는지를 보여주는 개념이라면, 균형이 존재하지 않는다는든지 복수의 균형이 존재하는 것은 합리적 선택이론에서의 합리성 개념이 '비결정적'(indeterminate: Elster, 1986)임을 보여준다.

25) 또한 햄린은 이 효용이 (1) 사적(personal), (2) 전체적(global) (3) 통시적(intertemporal), (4) 기대적(expectational), (5) 결과론적(consequentialist), (6) 행동특정적(act-specific)인 성격을 가지고 있다고 한다. 효용의 이러한 측면은 당면한 모델 속의 문제 자체의 풀이와는 관계가 없으며, 모델 속의 문제풀이와 현실의 문제를 연관시켜 해석할 때 나타난다고 생각된다. 모델 속의 문제 풀이는 통상 이러한 성격을 전부 만족시킨다. 예를 들어 이타적 동기가 포함되어도 행위자가 자기의 효용을 극대화한다는 점에서 사적이라 할 수 있고, 실제 어떤 의사결정의 규칙을 택할 것인지를 알기 위한 모델 내에서는 규칙 자체의 선택이 행동의 선택이 되는 것이다. 통시적이라는 표현 역시 실제 개인의 전 생애에 걸친 것이 아니라 할지라도 분석의 대상이 되는 기간을 통하여 모델 내에서는 통시적이 된다.

26) 그렇다고 하여 의사결정 모델이 게임모델보다 풀기가 더 쉽다든지 유용성이 더 떨어진다고 일률적으로 말할 수 있는 것은 아니다. 게임모델의 복잡성은 한계가 있어 별다른 수학적 기법을 요구하지 않는 경우가 많으나 의사결정 모델은 오히려 고도의 수학적 기법을 사용하는 경우가 많다. 상호의존의 상황일지라도 행위자의 수가 많은 경우, 게임모델보다는 의사결정모델을 구성하는 것이 더 유용할 수도 있음은 물론이다.

27) 기대효용이론에 대해서는 Luce and Raiffa(1957) 참고.

28) 내쉬균형에 대한 보다 상세한 설명 및 기타 여러 균형 개념에 대해서는 정준표(2000) 및 Morrow(1994) 참고.

게임이론적 모델 속의 행위자가 고도의 합리성을 가진 존재로 파악되는 반면, 이렇게 게임모델을 구성하기 위해 외부적으로 주어진 것이라고 간주되는 선호와 초기의 믿음이 과연 합리적인가를 따지지는 않는다. 선택 가능한 대안 및 실현 가능한 결과의 집합 역시 추상화된다. 이 과정에서 위에 논한 철학적·규범적 의미에서의 합리성에 대비되는 비합리성은 제거되고 모델 속의 행위자는 고도의 합리성을 갖게 되는 것이다. 물론 현실의 인간이 어느 정도는 목적을 달성하기 위해 상당히 효율적인 수단을 선택하는 합리성을 가지고 있지 않다면 이러한 모델의 유용성은 미미하다고 할 것이다.<sup>29)</sup> 사이몬은 수재보험 등 보험구입행위는 신고전파 경제학자들의 예측(수재를 입을 객관적 확률에 의거한 비용·편익분석)과는 달리 수재를 입었다든지 그러한 경험을 가진 사람을 아는 사람들에 의해 주로 행해진다면서 어떤 환경에서 현실의 특정 측면에 주의를 기울이게 되는지가 매우 중요하다고 하면 극대화(maximization)에 기반한 합리적 선택을 비판하고 있으나(Simon, 1986), 이러한 행위자의 주관적 인식을 이미 수용한 모델에서 극대화 대신에 만족화(satisfying)를 지향하는 것은 득보다는 오히려 잃는 것이 많다. 만족화의 문제는 그것이 잘 정의되어 있지 않고 예측이 예리하지 못하다는 데 있다. 이미 현실의 추상화를 통하여 행위자의 '비합리성'을 일정 부분 반영한 연후에 또다시 만족화라는 기준을 도입할 필요는 없는 것이다. 합리적 선택, 특히 게임이론적 모델이 최소한 이론적 수준에서 상당한 발전을 한 이유 중 하나가 분명한 연산의 규칙이 있다는 것이다(cf. Coleman, 1990, pp.18-19).

### 3. 합리적 선택이론의 한계

합리적 선택이론의 한계는, 물론 두 가지가 서로 밀접하게 연결되어 있지만, 설명하고자 하는 현상과 이를 추상화한 모델의 관계라는 측면에서 파악될 수도 있고, 모델 내부에서 문제를 풀 도구 자체가 문제가 있다는 측면에서 파악될 수도 있다.

#### <균형 개념의 문제>

멍크(Munck, 2001)는 합리적 선택이론의 한 갈래로서의 게임이론을 이론적 측면과 모

29) 합리적 선택에 의한 설명이 적실성을 가지는 경우에 대한 논의가 이와 관련이 있다. 그린과 샤피로의 용어로는 '부문화된 보편주의'(segmented universalism)에 해당하는 것으로 예를 들어 선택지의 수가 비교적 작고 결과가 상당히 중요하며 개인이 결과에 영향을 미칠 가능성이 상당히 높을 때 합리적 선택에 의한 설명이 성공적이기 쉽다는 것 등이다(Green and Shapiro, 1994, pp.27-28; Taylor, 1996).

델이라는 측면으로 나누어 분석하면서 이론적 측면의 문제로서 앞의 2절에서 언급한 기대효용이론의 사용, 균형의 복수성 문제, 게임의 규칙과 관련된 문제 등 3 가지를 들고 있다. 이 3 가지 중 필자의 생각으로는 모델 내부의 문제는 오직 균형의 복수성 문제뿐인 것 같다.<sup>30)</sup> 의사결정의 상황에서는 이런 문제가 생기지 않지만,<sup>31)</sup> 게임상황에서는 다수의 균형이 존재할 수 있다는 사실은, 멩크의 말을 빌리면, “기대효용모델의 적용이 적절하다고 판단되는 상황에서도 게임이론은 완전한 설명을 필연적으로 제공하지는 않는다.” 필자는 균형의 복수성이 게임이론의 한계라는 주장을 대체로 수긍하지만, (1) 현실 자체가 예측 혹은 설명이 불가능할 경우도 있으며 또한 (2) 균형의 복수성이 합리성외의 다른 측면에 관심을 기울이게 한 공헌도 무시할 수 없다고 본다.

만일 게임의 ‘해’(solution)를 분석자가 생각할 때 합리적이라고 가정한 이 게임의 참가자들이 택할 가능성이 매우, 매우 큰(very, very likely to choose) 전략조합(strategy profile)이라고 정의한다면(Kreps, 1990a, p.392), 어떤 전략조합이 게임의 해가 되기 위한 필요조건(necessary condition)이 바로 그 전략조합이 내쉬균형이어야 한다는 것이 된다.<sup>32)</sup> 모든 게임에서 더 나아가 특정 게임의 원천(source)이 된 현실에서 이러한 해가 꼭 존재한다고 볼 수는 없는 것이다.<sup>33)</sup> 한편, 유일한 (내쉬)균형이라 하여 무조건적으로 게임의 해로 간주할 수는 없지만,<sup>34)</sup> 일단 균형이 하나밖에 없는 경우는 이를 해로 간주하면 복수의 균형이 존재할 경우 어떻게 해가 존재할 수 있는지가 논의의 대상이 된다.<sup>35)</sup>

30) 물론 균형이 존재하지 않는 경우도 있을 수 있다. 게임이론 자체의 문제로는 그밖에 ‘전략’(strategy) 및 ‘혼합전략’(mixed strategy) 등 게임이론의 기초적 개념 자체의 의미가 생각보다 불분명하다는 점을 들 수 있다. 이에 대해서는 Rubinstein(1991) 참고.

31) 최선의 선택이 여러 개 있더라도 이들 모두가 행위자에게 동일한 효용을 줄 수밖에 없다.

32) 여기서 전략조합이란 게임의 모든 참가자들의 전략을 질서있게 배열한 것이다. 만일 내쉬균형이 아닌 어떤 전략조합이 해라고 가정해보자. 다른 참가자들이 그 전략조합이 규정하는 전략을 선택할 것이 어떤 참가자에게 명백하다면 (즉, 굉장히 가능성이 크다면), 합리적인 그 참가자는 자신의 효용을 극대화하는 다른 전략을 선택할 것이고, 따라서 적어도 문제가 되는 전략조합에 규정된 이 참가자의 전략이 이 참가자에 의해 사용될 것이 다른 참가자에게는 명백할 수가 없으며, 그 전략조합이 내쉬균형이 아닌 한 이런 참가자는 항상 존재한다. 따라서 내쉬균형은 해가 되기 위한 필요조건이다(정준표, 2000).

33) 필자는 본문에서의 해에 해당하는 “게임을 하는 유일하게 합리적인 방법”(uniquely rational way to play the game)이 모든 게임에서 항상 존재하지는 않는다는 입장이다. 유일하게 합리적인 방법이 존재한다는 입장에서는 “합리성에 대한 공통지식”(common knowledge of rationality: CKR)뿐만 아니라 행위자의 믿음이 서로 상충되지 않을 것(consistent alignment of beliefs: CAB)을 필요로 한다. CKR은 모든 게임의 참가자가 (도구적으로) 합리적이며 (a), 이 사실(a)을 모두 알고 있으며 (b); 모두가 합리적이라는 사실을 알고 있다는 사실(b)을 모두가 알고 있고 (c); c를 모두가 알고 있고 (d); …… 라는 식으로 무한히 계속될 수 있는 것을 말한다. CAB는 대략 동일한 정보를 가진 합리적인 게임의 참가자는 서로 다른 사고의 과정을 거칠 수 없다는 것에 해당한다. 보다 상세한 내용에 대해서는 Hargreaves Heap and Varoufakis, 1995 참고.

34) 이에 대한 예로는 Kreps, 1990a, pp.416-417 참고. 물론 이 예는 “가장 큰 수를 말하라”라는 게임을 포함하는 매우 인위적인 것이다.

소위 ‘균형의 정교화’(refinement of equilibrium) 외에도 사전협상(preplay negotiation), 참가자들의 과거의 경험(experience), 관습(convention), 참가자의 기대를 일치시키는 눈에 띄는(prominent) 초점(focal point)의 존재 등이 복수의 균형이 있는 경우에도 게임의 참가자들이 어느 특정의 균형 속의 전략을 선택하게 하는 이유가 될 수 있다.<sup>36)</sup> 게임이론의 비결정성(indeterminacy)으로 나타나는 합리성의 한계를 이와 같이 보충하려는 노력은 소위 ‘부분적 보편주의’(partial universalism: Green and Shapiro, 1994, pp.26-27)로 나타난다. 합리성 외에도 사회적 관습, 규범, 아이디어, 문화적 특성 등등의 다른 요소들도 고려되어야 주어진 현상에 대한 적절한 설명을 할 수 있다는 (필자가 보기에는 새삼스러운) 인식이 뒤에 언급하듯이 신제도주의의 한 분파라고 할 수 있는 소위 ‘합리적 선택 제도주의’(rational choice institutionalism)의 등장 배경이 된다.<sup>37)</sup> 물론 비게임이론적 요소(non-game theoretic factors)를 도입하는 것이 게임이론의 일관성을 궁극적으로 해칠 것이라는 우려도 있지만(Munck, 2001), 이 또한 게임이론이 무엇이나에 대한 견해에 따라 다르게 해석될 여지도 있으며, 뒤에 논의할 소위 ‘게임의 규칙’이 어디서 오는지의 문제와 비교해 볼 때 그렇게 중요한 문제는 아니라고 본다. 한편, 게임의 해의 존재가 사회적 관습, 규범, 과거의 경험, 초점 등에 의존한다는 것은 인간의 합리성이 ‘합리적이길 의도하나 거기에 미치지 못하는’(intendedly rational, but only limitedly so) ‘제한된 합리성(bounded rationality)’이며 ‘회고적(retrospective) 합리성임을 보여준다고 할 수도 있다.<sup>38)</sup>

35) 필자는 적절한 이유를 제시하지 않은 한 균형이 여러 개 존재하더라도 합리적 행위자들은 이 균형 중에 하나를 선택할 것이니 균형이 아닌 결과는 예측에서 제외할 수 있다는 견해에 동의하지 않는다. 필자는 행위자의 초합리성(hyper-rationality)을 근거로 CAB를 가정하는 것은 적절한 이유가 될 수 없고, 예를 들어 게임이론의 목적이 현실세계의 반복되는 혹은 반복될 수 있는 규칙성(regularity)을 이해·설명하는 것이다(Rubinstein, 1991; van Damme, 1998)라는 것은 적절한 이유가 될 수 있다고 본다. 명크는 균형의 복수성이라는 게임이론의 한계에 대한 이론가들의 반응을 ‘순수주의자 입장’(purist position)과 ‘실용주의자 입장’(pragmatist)으로 나누고 있는데(Munck, 2001), 명크가 말하는 순수주의자보다 더 순수한 견해가 바로 무조건적으로 균형만이 게임의 결과가 될 수 있다는 위의 견해이다.

36) 이에 대한 보다 상세한 설명은 Kreps(1990b), 정준표(2000) 등을 참고.

37) Yee(1997)는 이런 움직임을 물질적이 아니라 규범, 아이디어, 믿음 등의 정신적 요소를 강조한다는 점과 이 움직임이 엄청난 유행을 타고 있다는 점에서 “Ideational Tempest”라 부른다.

38) Kreps, 1990b, Chapter 6. 그는 더 나아가 균형의 정교화에서 응용되는 것과 같이 연역(deduction) 및 전망(prospaction)을 적용하여 균형의 복수 문제를 해결하는 것 역시 무엇보다도 회고적이고 제한된 사고의 예라고까지 주장한다(p.153). 이러한 견해에 따르면, 결국 합리적 선택이론의 (모델 내의) 합리성은 현실에서의 인간의 제한된 합리성을 단순화시킨 것이 된다. 박효중(1994)은 민주화 과정에 관한 합리적 선택이론의 적실성을 분석하면서, 합리적 선택이론이 다만 “보고 싶은 현상만 골라서 보는데” 효과적이며 유의미하다는 점에서 방법을 나름대로 ‘제한된 합리성’(bounded rationality)과 ‘만족화’(satisfying)의 합정에 함몰되어 있는지 모른다”고 하나, 위 견해에 따르면 실제 그 방법론에서 다루는 행위자 역시 이미 제한된 합리성과 만족화의 함정에 빠져 있는 것이다.

### <기대효용이론의 문제>

이제 기대효용이론의 사용과 관련된 문제를 살펴보자. 모델 내부라는 관점에서 보면 기대효용이론의 사용은 강점이지만 약점이 아니다.<sup>39)</sup> 기대효용이론을 사용하지 않는 경우를 생각해 보면, 모든 결과(outcomes)뿐만 아니라 이런 결과들을 발생할 확률과 함께 고려한 최소한 수많은 '복권'(lottery)들에 대한 선호를 확인하고 나서야 의사결정 혹은 게임상황에서 어떤 결과가 나올지 알 수 있을 것이다. 전망이론(prospect theory), 제한된 합리성에 기반한 만족화(satisfying)는 이 경우 과연 구체적으로 어떤 지침을 내릴 것인지 알기 어렵다. 결국 기대효용이론을 사용하는 것은 복잡한 문제를 단순화하고, 보편적으로 적용할 기준을 제시해 주는 것으로 모델 내부적 관점에서는 엄청난 강점이라고 할 수 있다. 제한된 합리성 및 전망이론 등은 기대효용이론에 대한 비판으로서는 어느 정도 타당성을 인정받을 수 있으나, 취급이 용이하지도 않으며 누적적 이론 발전에도 상당한 제약이 따라 대안으로서의 가치는 제한적일 수밖에 없다.<sup>40)</sup>

한편, 모델과 현실과의 관련 측면에서 볼 때, 이미 상당한 추상을 거쳐 모델을 구성한 마당에 과연 세세한 심리적 요소까지를 다시 반영할 필요가 있는지는 의문이다. 틀짜기(framing)의 효과를 예를 들어 이를 설명해보자. 다음 예는 기대효용이론의 문제점을 분명히 보여준다(Tversky and Kahneman, 1986):

문제 3: 다음과 같은 두 개의 동시적인 결정을 해야 한다. 두 결정을 모두 검토하고, 선호하는 대안에 표시하라.

결정 (1): 다음 중에서 선택하라:

- A. 240 달러를 확실히 얻음(150명 중 84%가 선택)
- B. 1000 달러를 얻을 확률이 25%이고 아무 것도 얻지 못할 확률이 75%임(16%)

결정 (2): 다음 중에서 선택하라:

- C. 750 달러를 확실히 잃음(13%)
- D. 1000 달러를 얻을 확률이 75%이고 아무 것도 얻지 못할 확률이 25%임(87%)

문제 4: 다음에서 택하라:

- A와 D: 240 달러를 얻을 확률이 25%이고 760달러를 잃을 확률이 75%임(86명 중 0%)
- B와 C: 250 달러를 얻을 확률이 25%이고 750달러를 잃을 확률이 75%임(86명 중 100%)

39) 합리적 선택이론에서 베이즈의 규칙 사용에 대해서도 유사한 논의가 적용될 수 있다.

40) 전망이론은 1979년에 제시되었다(Kahneman and Tversky, 1979). 필자는 사실 이 이론이 얼마나 누적적인 발전을 이루었는지 알지 못한다.



문제 3에서 응답자의 73%가 A와 D를 한 묶음으로 하여 선택하였으나, 문제 4에서는 문제 3에서 3%의 응답자만이 택한 B와 C의 묶음을 전부 A와 D에 비해 나은 것임을 인식하여 전부 B와 C를 택한 것이다. 문제는 과연 합리적 선택 모델의 현실에 대한 추상화의 정도가 이런 정도로 세밀한 것인가에 달려있다. 많은 합리적 선택 모델은 이미 언급한 대로 거시적 현상을 설명하기 위한 것이고, 이런 모델에서 가능한 결과 및 선택할 수 있는 대안은 현실에 비해 엄청나게 단순화되는 것이다. 레인은 인식적 과부하(cognitive overload)와 심리적 스트레스가 심한 상황에서 합리성의 적용은 문제가 있다고 한다(Lane, 1996). 필자는 합리적 선택이론이 잘 적용될 수 있는 분야가 있고 그렇지 않은 분야가 있다는 ‘부문화된 보편주의’와 아울러 효용극대화의 합리성만으로는 충분한 설명이 되지 못할 수 있다는 ‘부분적 보편주의’를 수용하는 입장이지만,<sup>41)</sup> 합리적 선택 모델은 객관적 현실(objective reality)이 아니라 행위자에게 인식된 주관적 현실(subjective reality)을 모델 속에 반영한다고 본다면,<sup>42)</sup> 레인이 말하는 상황에서도 합리적 선택이론이 항상 적용 불가능하다고 보기는 어렵다고 할 것이다.

#### <게임의 규칙 문제>

게임이론으로 본 합리적 선택이론의 가장 큰 한계(?)는 ‘게임의 규칙’이 외부적으로 주어져 있다고 가정하는 것이다. 게임이론적 모델을 구성하기 위해서는 게임의 참가자, 그들의 선호(효용함수), 선택 가능한 대안, 나타날 수 있는 결과, 가지고 있는 정보, 게임의 진행 순서 등등을 우선 결정해야 하는 것이다. 이러한 게임의 규칙은 어디에서 오는가? 게임이론이 게임이론적 모델을 구성하는데 있어서는 이리이러한 요소가 필요하다는 것을 알려주고, 게임이론적 연구결과들은 연구자가 새로 만들 모델이 풀릴 수 있는 문제인가에 대한 대강의 아이디어를 제공하지만, 게임의 규칙을 구체적으로 어떻게 만들 것인가에 대한 지침은 제공하지 않는다. 그러나, 이것은 아래에 논의하는 바와 같이 게임이론의 한계가 아니라 모든 형식모델의 한계라고 볼 수 있다. 한계에 의문부호를 찍은 이유가 여기에 있다.

41) 부문화된 보편주의를 수용하는 경우에 ‘적용영역의 구체화’(domain specification)란 문제가 생기는데, 명크는 여기에 대해 이론가들 사이에 합의가 없음을 지적하고 있다(Munck, 2001). 필자의 생각으로는 표면상 대립된 견해는 적용의 목적에 대한 견해 차이의 반영이라고 생각되며, 경험과학의 측면에서의 적용영역에 대해서는 상당한 합의가 이미 이루어져 있다고 생각되나(cf. Green and Shapiro, 1996) 적용의 가능성이 작은 영역에서도 창조적 적용의 가능성이 있으므로 적용 가능영역을 미리 한정할 필요는 없다고 본다.

42) 루빈스타인은 게임이 세상의 물리적 규칙(physical rules)을 묘사하는 것이 아니라면, 게임이론적 모델은 참가자에게 ‘의미가 있다고’(relevant) 인식되는 요소만을 반영하여야 한다고 말한다(Rubinstein, 1991).

명크는 이미 언급한 대로 게임이론을 게임이론의 이론 부분과 모델 부분으로 나누어 게임이론을 비판적으로 고찰하는데(Munck, 2001), 사실 게임의 규칙을 상정하는 것이 바로 (형식)모델을 구성하는 것에 해당하므로 이 이분법이 과연 적절한지는 의심스럽다. 게임이론적 모델이 형식모델(formal model)이라는 사실에서 여러 가지 장점이 나온다. 형식모델에서는 그 모델의 가정, 결론 및 가정에서부터 결론의 도출과정을 명백하게 제시해야 하며, 분명한 논리에 의해 결론이 가정으로부터 도출되어야 한다. 따라서 형식모델에서는 암묵적 가정 및 불명확한 논리를 배제할 수 있고, 가정이 달라지면 정확히 어떤 식으로 결론이 변경되는지를 알 수 있다. 이러한 형식모델의 사용이 지식의 축적에 기여하는 바가 크다는 것은 다시 말할 필요가 없을 것이다.<sup>43)</sup>

그러나 형식모델에서 사용되는 연역적 논리(deductive logic) 그 자체로 세상의 진실을 발견하게 하지는 않는다. 중요한 것은 그 모델의 가정이 얼마나 현실과 부합하느냐이다. 결국 게임이론적 모델이 형식모델이라면, 게임의 규칙은 모델의 가정이다. 게임이론적 모델이 현실에 대한 설명력을 가지려면 결국 게임의 규칙을 잘 설정해야 한다. 이는 게임이론의 문제가 아니라 연구자의 문제이다. 모든 이론 혹은 모델은 현실 혹은 현실의 일부를 이해·설명하기 위해 현실로부터 '추상'(abstraction)한 것이다. 추상은 실질적으로 중요하다고 생각되는 부분만을 추출하는 것이다. 추상화의 정도가 너무 심해도 현실에 대한 의미 있는 분석이 될 수 없고, 추상화의 정도가 너무 약한 경우에도 모델의 복잡성만 더해지지 현실의 이해에 도움이 되지 않는다. 주어진 상황에 대한 게임의 전개형 혹은 전략형을 구성하는 것 역시 추상화를 필요로 한다. 게임이론적 모델은 게다가 그 게임을 풀 수 있기 위해서는 다른 모델보다 더 단순해야 할 경우가 많다. 참가자를 누구로 할 것인가? 누가 언제 움직이며, 각 참가자에게 가능한 행동은 무엇인가?<sup>44)</sup> 참가자의 몫은 어떻게 설정할 것인가? 불완전정보를 도입한다면 어떤 형식으로 도입할 것인가? 이런 근본적 질문에 대답하기 위해서는 게임이론적 지식도 필요하지만 주어진 상황에 대한 깊이 있는 지식도 필요하다.<sup>45)</sup>

43) 형식모델의 특징 및 그 장점에 대해서는 김재한, 1994 참고. 명크는 지식의 축적이 과연 합리적 선택이론가들이 주장하는 만큼 이루어지고 있는가에 회의를 표한다(Munck, 2001). 여기서 문제는 '지식'의 의미가 무엇인가에 달려있다. 합리적 선택이론가들은 예를 들어 모델이 불완전정보(incomplete information)를 가정할 경우 완벽·완전정보(perfect and complete information)를 가정한 경우와 어떻게 달라지는가를 알게 될 경우 지식의 축적이 이루어진다고 보며, 이런 측면에서는 분명히 지식의 축적이 가능하다. 또한 지식 축적의 주장은 다른 접근방법을 사용할 때와 비교하여 이해되어야 할 것이다.

44) 어떤 행동을 택할 경우 강력한 제재를 높은 확률로 받는다고 할 때, 이 행동을 아예 불가능하다고 보고 행위자가 택할 수 있는 대안에서 제외할 수도 있고, 이러한 행동의 선택 가능성과 그에 따른 처벌의 강도와 확률을 모델 속에 포함시킬 수도 있다. 즉, 제2장 그림 1의 제도적 제약과 기회비용의 구분 역시 이미 구성된 모델 내에서나 분명해지는 것이다.

위의 논의가 주로 현실을 경험적 측면에서 총체적으로 설명하고 이해하기 위한 모델에 관한 것이라면, 형식모델은 이와는 좀 다른 목적에서 구성될 수도 있으며 필자의 생각으로는 지금까지의 합리적 선택이론의 명성은 오히려 여기에 기인한 바가 크며, 합리적 선택이론가들이 자신의 모델에 대한 경험적 검증을 잘 하지 않는다는 비판 역시 여기서 온다. 모델은 크게 보아 '근사치로서의 모델'(models as approximations)과 '캐리커처로서의 모델'(models as caricatures)로 나누어질 수 있다.<sup>45)</sup> 앞의 설명이 근사치로서의 모델에 관한 것이었다면, 캐리커처로서의 모델은 복잡한 현실의 특정 측면만을 부각하여 이를 분석한다. 근사치로서의 모델의 목표가 취급가능성(tractability)이 허용하는 한 현실을 될 수 있는 한 왜곡시키지 않는데 비해, 캐리커처로서의 모델은 때로는 고의적으로 현실을 왜곡하기도 한다. 예를 들어 스펜스는 학위를 얻는 이유가 교육을 받는 것이 그 자체 직업을 수행하는데 도움이 되어서가 아니라 단순히 자신의 능력에 관한 신호(signal)라는 가정을 하고 이러한 가정의 함의를 분석한다(Spence, 1974). 울슨의 집단행동의 논리 역시 캐리커처로서의 모델로 볼 수 있다. 합리적 선택이론가들이 합리적 선택이론의 경험적 적실성의 한계에 대해 반박하면서 '다른 조건이 변하지 않은 한'(ceteris paribus)이라는 가정을 자주 드는 것도 바로 이 때문이다. 근사치로서의 모델보다는 캐리커처로서의 모델이 엘스터(Elster, 2000)가 강조하는 거시적 현상의 미시적 메커니즘의 발견에 기여하였음은 물론이다.

합리적 선택이론에 대한 비판으로는 이외에도 여러 가지가 있다. 합리적 선택이론의 약점으로 흔히 행위자의 목표 혹은 선호를 주어진 것으로 간주하는 점을 든다. 그러나, 필자는 경우에 따라서는 선호의 변화를 모델 속에 포함할 수 있다고 보나 그것은 원하는 결과를 얻기 위해 자의적으로 사용될 수밖에 없다는 것을 강조하고 싶다. 한편, 합리적 선택이론에서는 모델 속에서의 결과에 대한 선호는 변하지 않지만, 대안에 대한 '유인된 선호'(induced preference)는 변화할 수 있음을 알아야 한다. 사실 구성주의(constructivism)적 시각 혹은 역사제도주의(historical institutionalism)적 시각에서 합리적 선택이론의 외생적으로 주어진 선호에 대한 가정을 비판하는 경우가 많으나, 그들의 실제 연구에서 과연 외생적으로 주어진 선호에 해당하는 것이 없는 것인지는 의문이다.<sup>47)</sup> 행위자의 선

45) 이런 측면에서 보면 상황에 대한 깊은 지식(in-depth knowledge)과 아울러 연역적 논리를 강조하는 소위 "Analytic Narrative"(Bates et. al., 1998)는 새삼스러운 것이 아니다. 이것이 새삼스럽게 느껴지는 이유는 지금까지의 합리적 선택이론이 투표, 선거 등 게임의 규칙이 비교적 분명히 드러나는 매우 구조화된 상황에 적용되거나 혹은 본문에 언급되는 "캐리커처로서의 모델"에 관심을 기울였기 때문이다.

46) Gibbard and Varian, 1978. 사실 그들은 모델과 응용모델(applied model)의 구분을 먼저 하고 난 후에 응용모형을 본문에서와 같이 두 가지로 나눈다. 따라서 필자가 모델이라는 것은 그들의 응용모형을 의미한다.

호 외에도 사용할 수 있는 전략 등 게임의 규칙이 고정되어 있어 변하지 않는다는 가정 역시 자의적 결과를 도출하는 것을 막는다는 점에서 위와 마찬가지로의 논의를 전개할 수 있다. 현실의 인간은 때로는 전혀 생각도 않은 전략을 사용하기도 하고, 규칙(rules)을 준수하기도 하지만 규칙을 파괴하고 새로운 규칙을 만들기도 한다. 규칙이 불명확하거나 규칙이 변화하는 이러한 유동적 상황에서는 결국 게임이론적 모델을 구성하기가 어렵고, 더구나 어떤 일이 일어날지 예측하기는 정말로 어렵다고 할 것이다. 현실에 일정한 규칙성(regularity)이 없는 경우, 합리적 선택이론은 사후적 설명(post hoc explanation)을 가능하게 할지언정 예측을 가능하게 하지는 않는다는 것이 필자의 견해다.<sup>48)</sup> 그러나 과연 이것이 합리적 선택만의 문제인지는 의문이다. 기타 모델 구성에 있어서 사소한 차이가 그 결론에 있어 상당한 차이를 가져올 수 있다는 것도 문제이다(Kreps, 1990b, Chapter 5).

#### IV. 결 론

합리적 선택이론은 80년대에는 국제관계, 90년대에는 비교정치, 그리고 요즈음은 역사적 현상에까지 그 적용영역을 넓혀가고 있다.<sup>49)</sup> 소위 합리적 선택 제도주의라 부르는 일련의 흐름이 바로 그것이다. 이러한 추세에 배경으로는 크게 3 가지를 들 수 있다. 첫째, 세상의 변화다. 1980년대와 1990년대를 거치면서 아프리카, 중동, 발칸반도에서 일어난 일련의 사태는 문화적 요소의 중요성을 부각시켰고, 민주화의 불안정한 제3의 물결 역시 합리적 선택이론을 과거의 미국을 중심으로 한 선진국의 고도로 제도화된 상황에만 분석을 한정하도록 놔두지 않았던 것이다(Bates, de Figueiredo, and Weingast, 1998). 둘째, 그린과 샤피로(Green and Shapiro, 1994)를 필두로 하여 합리적 선택이론의 경험적 적실성에 대한 비판이 계속적으로 제기되었다는 점이다. 마지막으로, 이미 언급한 바와 같이 합리성 외에도 사회적 관습, 규범, 아이디어, 문화적 특성 등등의 다른 요소들도 고려

47) 국제관계 분야이긴 하나, 이와 관련된 논의로는 Clark(1998)과 Sterling-Folker(2000) 참고.

48) 물론 규칙성은 생각지도 않은 측면에서 발견될 수도 있다. 루빈스타인은 색깔을 구분할 수 있는 한 살짜리 자기 딸에게 청색, 적색, 녹색의 3 직육면체를 무작위로 배열하고 딸이 과연 일관성 있는 선택을 하는가를 알아보기 위한 실험을 약 12번을 했는데, 딸의 선택에 아무런 일관성이 없다고 생각했으나, 결국 딸이 색깔이 아니라 왼쪽, 중간, 오른쪽이라는 위치 선택에 있어 항상 왼쪽을 선택하는 엄청난 일관성을 보였다는 것을 깨달았다고 한다(Rubinstein, 1991). 친한 사람으로부터의 충고는 받아들여지지 않으나 나쁜 사람의 충고를 무시하는 현상도 규칙성의 한 예이고, 이미 언급한 대로 이는 신호게임이 그 매커니즘을 잘 밝혀주고 있다.

49) 국제관계에서 합리적 선택, 특히 게임이론의 적용에 대한 논의로는 정준표(2000), 비교정치의 분야에서 합리적 선택이 어떻게 사용되고 있는지에 대한 논의로는 Bates(1997), 역사적·거시적 현상을 둘러싼 합리적 선택이론의 적용의 성공여부에 대한 논쟁으로는 Elster(2000)와 Bates et al.(2000) 참고.

되어야 주어진 현상에 대한 적절한 설명을 할 수 있다는 인식이 널리 퍼지게 되었다는 것이다.<sup>50)</sup> 이미 언급한 바와 같이 합리적 선택을 실제로 적용하는데 있어서는 행위자의 선호, 믿음, 실현 가능한 결과의 집합, 선택 가능한 대안, 행동의 순서 등등을 결정해야 한다. 이러한 이론 혹은 모델 외부에서 결정되는 많은 요소가 있기에 합리적 선택이론은 오히려 사회학, 심리학 등 다른 분야의 연구성과를 그 이론 속에 받아들일 여지가 있으며 (cf. 김용학, 1996, pp.73-76), 그런 측면에서는 이러한 합리적 선택 제도주의의 흐름을 일단 긍정적으로 받아들일 수 있다. 그러나, 이러한 흐름에서 우려되는 현상은 짜 맞추기 식의 논리전개와 그렇고 그런 이야기를 할 가능성이 그만큼 커질 수 있다는 것이다.<sup>51)</sup> 신제도주의에서 제도(institution)의 개념은 확장에 확장을 거듭하여 공식적 규칙뿐만 아니라 규범, 아이디어, 문화 등등 모든 정신적 요소를 포함하는 방향으로 가고 있고, 경우에 따라서는 이러한 요소들이 행위자의 효용(선호)에 반영되기도 하고, 유일한 균형을 얻기 위해 부가적으로(additively) 균형의 선택에 이용되기도 한다. 문제를 축소해 형식모델을 사용하기 편한 곳은 이를 사용하고, 그것이 어려울 경우는 비공식적 이야기로 적당히 얼버무릴 가능성도 항상 존재한다. 제도의 제약을 강조하다보니 궁극적으로 개인의 선택에 남는 것은 별 의미 없는 것일 수도 있다. 비교방법(comparative method)과 형식이론의 장점을 취해 가장 좋은 종합(synthesis)을 이룩하려는 목적이 오히려 거꾸로 가장 나쁜 종합이 될 가능성이 있는 것이다(Mahoney, 2000).

끝으로 지금까지의 논의를 간략히 종합하고, 합리적 선택이론에 관해 유의할 점을 몇 가지만 지적해보자. 합리적 선택이론에서의 합리성은 (1) 다분히 정의의 문제이며, 합리적 선택이론이 경험적 적실성을 가지려면 다양한 인간 행동의 (주관적) 동기를 상황에 맞게 도입하는 것이 필요하며 (2) 합리성만으로는 모델 내에서도 유일한 결과를 도출하기 어려우며 (3) 현실 속의 인간의 불완전한 합리성을 단순화시킨 개념이며 (4) 현실세계에서의 규칙성을 담보로 하여 예측을 가능하게 하는 개념이다. 합리적 선택이론의 유용성으로는 무엇보다도 그 적용범위가 대단히 넓다는 것을 지적할 수 있다. 어느 정도 합리적인 행위자를 인정할 수 있는 상황은 모두 잠재적으로 합리적 선택이론적 틀을 적용할 수 있다. 따라서 사회과학 전체를 통하여 합리적 이론의 틀을 적용할 수 있고, 그런 만큼 한 분야에서의 연구성과를 쉽게 다른 분야에 변형·적용할 수 있게 된다. 한편, 이미 언급한

50) 예를 들어 미국정치에 관해서는 더 이상 새로운 것을 내놓을 것이 없어, 다른 분야에 관심을 가지게 되었다는 식의 지식사회학(sociology of knowledge)적 이유는 여기서 언급하지 않고자 한다.

51) “분명히 ‘경제적’ 이론들의 상시적 위험은 모든 것을 단순히 재서술함으로써 ‘설명’하게 될 수 있다는 것이다. 그럴 경우 경제적 이론들은 저것이 아니라 이것이 일어날 것이라는 예측에는 전혀 소용이 없게 된다”라는 배리의 경고는 아직도 유효한 것이다(Barry, 1978, p.33).

대로, 많은 경우 합리적 선택의 유용성은 형식모델의 사용에서 잘 드러난다. 그러나, 특정 상황을 설명 혹은 예측하기 위해서는 우선 제대로 된 추상에 의한 모델 구성이 필요하다.<sup>52)</sup> 형식모델은 분석자가 활용해야 할 도구이지, 분석자에게 모자라는 상황에 대한 전문성을 보충해 주지는 않는다는 것을 명심해야 할 것이다. 도구가 연구자가 해야 할 일이 무엇인지를 결정하는 것이 아니라 연구자가 해야 할 일이 그에 유용한 도구를 결정해야 할 것이다. 물론 형식모델의 엄격성(rigor)의 요구 때문에 다룰 수 있는 문제의 범위에 제한이 있을 수 있음을 항상 인식해야 한다. 형식모델을 비판하는 목소리 중에 실제 현실의 세계에서 중요한 문제를 다루지 않고 분석적으로 흥미 있는 질문에 관심을 집중한다는 것이 있다는 점은 유의할 필요가 있다.

마지막으로 합리적 선택이론에 별다른 전문가가 따로 없다는 것을 강조하고 싶다. 형식모형을 제대로 활용하기 위해서는 실제 어느 정도의 훈련이 필요한 것은 사실이지만 흔히 생각하고 있는 것처럼 엄청난 수학적·논리적 능력을 요하는 것은 아니다. 정리의 증명에는 간혹 고도의 수학이 필요할지 모르지만 특정상황을 분석하는 것은 대체로 고등학교 수준 이상의 수학을 넘지 않는다. 또한 비형식적 합리적 선택이론의 적용 혹은 합리적 선택이론에서 나온 결과의 적용은 합리적 선택이라고 불리지 않더라도 대부분의 연구에서 수시로 이루어지고 있음을 인식해야 할 것이다.

## 참고문헌

- 강원택, 2000, 「지역주의 투표와 합리적 선택: 비판적 고찰」, 『한국정치학회보』 제34집 2호: 51-67.  
 김용학, 1996, 『사회구조와 행위』, 서울: 나남.  
 김재한, 1994, 「한국정치외교연구의 형식모형적 접근의 현황」, 김웅진 편, 『방법, 방법론과 한국정치학』, 서울: 한울.  
 박효종, 1994, 「민주화와 합리적 선택」, 『한국정치학회보』 제28집 1호: 298-330.  
 정준표, 1993, 「집단행동과 게임이론」, 『한국정치학회보』 제27집 2호: 109-139.  
 정준표, 2000, 「국제정치경제와 게임이론」, 여정동·이종찬 편, 『현대국제정치경제』, 서울: 법문사.  
 조기숙, 1996, 『합리적 선택: 한국의 선거와 유권자』, 서울: 한울.

52) 원래 필자는 이 논문에서 국내의 합리적 선택이론을 적용한 연구성과물을 비판적으로 검토하고자 하였으나, 지면 관계상, 또 솔직히 말하면 괜한 논쟁만 일으킬 것 같아 이는 생략하였다. 전반적으로 보면, 우선 형식적 모델을 사용한 연구가 그렇게 많지 않으며, 그나마 필자의 생각으로는 학계에 별로 기여하지 못했던 것 같다. 기존의 모델을 그대로 다른 상황에 적용한 경우가 많아 형식모델의 발전에 공헌했다고 볼 수도 없고, 그나마 그 모델을 설명하고자 하는 현상에 적용하는 것이 타당하다는 것을 설득력 있게 제시하지 못하였다고 판단되기 때문이다.

- 조기숙, 1997, "지역주의 논쟁: 비판이론적 시각에 대한 비판", 『한국정치학회보』 제31집 2호: 203-232.
- 최종원, 1995, 「합리성과 정책연구」, 『한국정책학회보』 제4권 2호: 131-160.
- 황수익, 1989, 「사회행위의 의도되지 않은 결과, 기능적 설명, 그리고 개인주의적 설명」, 『한국정치학회보』 23: 372-384.
- Barry, Brian, 1978, *Sociologist, Economists and Democracy*, Chicago: University of Chicago Press.
- Bates, Robert H., 1997, "Comparative Politics and Rational Choice: A Review Essay." *American Political Science Review* 91(3), 699-704.
- Bates, Robert H., de Figueiredo, Rui J. P. Jr., and Barry R. Weingast, 1998, "The Politics of Interpretation: Rationality, Culture, and Transition," *Politics & Society* 26(4): 603-642.
- Bates, Robert H., Greif, Avner, Levi, Margaret, Rosenthal, Jean-Laurent, and Barry R. Weingast, 1998, *Analytic Narratives*, Princeton: Princeton University Press.
- Bates, Robert H., Greif, Avner, Levi, Margaret, Rosenthal, Jean-Laurent, and Barry R. Weingast, 2000, "The Analytic Narrative Project," *American Political Science Review* 94(3), 696-702.
- Chong, Dennis, 1996, "Rational Choice Theory's Mysterious Rivals," in Friedman, Jeffrey, ed., *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, New Haven: Yale University Press.
- Clark, William Roberts, 1988, "Agents and Structures: Two Views of Preferences, Two Views of Institutions," *International Studies Quarterly* 42, 245-270.
- Coleman, James S., *Foundations of Social Theory*, Cambridge, M.A.: Belknap Press of Harvard University Press.
- Davidson, Donald, 1979, *Essays on Actions and Events*, Oxford: Oxford University Press.
- Elster, J. 1988. "The Nature and Scope of Rational-Choice Explanation," in Ullmann-Margalit, E., ed., *Science and Reflection* 3, 51-65.
- Elster, J. 2000, "Rational Choice History: A Case of Excessive Ambition," *American Political Science Review* 94(3), 685-695.
- Elster, J., 1986. "Introduction" in Elster J., ed., *Rational Choice*, N.Y.: New York University Press.
- Friedman, Debra, and Michael Hechter, 1988, "The Contribution of Rational Choice Theory to Macrosociological Research," *Sociological Theory* 6: 207-224.
- Friedman, Milton, 1953, "Methodology of Positive Economics," in Friedman, Milton, *Essays in Positive Economics*, Chicago: University of Chicago Press.
- Gibbard, Allan, and Hal R. Varian, 1978, "Economic Models," *Journal of Philosophy* 75(11): 664-677.
- Green, Donald P. and Ian Shapiro, 1994, *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*, New Haven: Yale University Press.
- Green, Donald P. and Ian Shapiro, 1996, "Pathologies Revisited: Reflections on Our Critics," in Friedman, Jeffrey, ed., *The Rational Choice Controversy: Economic Models of Politics*

- Reconsidered*, New Haven: Yale University Press.
- Hamlin, A.P. 1986, *Ethics, Economics, and the State*, Brighton: Wheatsheaf Books.
- Hardin, Russell, 1995, *One for All*, Princeton, New Jersey: Princeton University Press
- Hargreaves Heap, Shaun P., 1989, *Rationality in Economics*, Oxford: Blackwell.
- Hargreaves Heap, Shaun P., and Yanis Varoufakis, 1995, *Game Theory: A Critical Introduction*, New York: Routledge.
- Hollis, M., and Lukes S. (eds.) 1982. *Rationality and Relativism*. Oxford: Blackwell.
- Ingram, Paul, and Karen Clay, 2000, "Choice-within-Constraints New Institutionalism and Implications for Sociology," *Annual Review of Sociology* 26: 525-546.
- Kahneman, Daniel, and Amos, Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47: 263-91.
- Knudsen, Christian, 1983, "Equilibrium, Perfect Rationality and the Problem of Self-reference in Economics," in Mäki, Uskali, Gustafsson, Bo, and Christian Knudsen, eds., *Rationality, Institutions and Economic Methodology*, New York: Routledge.
- Kreps, D. M., and R. Wilson. 1982. "Reputation and Imperfect Information," *Journal of Economic Theory* 27: 253-279.
- Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson. 1982. "Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27: 245-252.
- Kreps, David M., 1990a, *A Course in Microeconomic Theory*, New York: Oxford University Press.
- Kreps, David M., 1990b, *Game Theory and Economic Modelling*, New York: Oxford University Press.
- Lane, Robert E., 1996, "What Rational Choice Explains," in Friedman, Jeffrey, ed., *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, New Haven: Yale University Press.
- Levi, Margaret, 1997, "A Model, a Method, and a Map: Rational Choice in Comparative Historical Analysis," in Lichbach, Mark Irving, and Alan S. Zuckerman, eds., *Comparative Politics: Rationality, Culture, and Structure*, Cambridge: Cambridge University Press.
- Levi, Margaret, 1998, *Of Rule and Revenue*, Berkeley: University of California Press.
- Luce, Duncan R., and Howard Raiffa, 1957, *Games and Decisions*, New York: John Wiley.
- Mahoney, James, 2000, "Rational Choice Theory and the Comparative Method: An Emerging Synthesis?" *Studies In Comparative International Development* 35(2): 83-94.
- Morrow, James D., *Game Theory for Political Scientists*, Princeton, New Jersey: Princeton University Press, 1994.
- Munck, Geraldo L., 2001, "Game Theory and Comparative Politics," *World Politics* 53: 173-204.
- Olson, M., Jr. 1965. *The Logic of Collective Action*. Cambridge, M.A.: Harvard Univ. Press.
- Ostrom, Elinor, 1990, *Governing the Commons: The Evolution of Institutions for Collective Action*, New York: Cambridge University Press.
- Ostrom, Elinor, 1991, "Rational Choice and Institutional Analysis: Toward Complementarity,"



- American Political Science Review* 85(1): 237-243.
- Rubinstein, Ariel, 1991, "Comments on the Interpretation of Game Theory," *Econometrica* 59(4): 909-924.
- Schelling, T. C. 1978. *Micromotives and Macrobehavior*. New York: Norton.
- Simon, H. A. 1985. "Human Nature in Politics: The Dialogue of Psychology with Political Science." *American Political Science Review* 79: 293-304.
- Simon, H. A., 1986, Rationality in Psychology and Economics, in Hogarth, Robin M., and Melvin W. Reder, eds., *Rational Choice: the Contrast between Economics and Psychology*, Chicago: University of Chicago Press.
- Spence, M., 1974, *Market Signalling: Information Transfer in Hiring and Related Screening Processes*, Cambridge: Harvard University Press.
- Sterling-Folker, Jennifer, 2000, "Competing Paradigms or Birds of a Feather? Constructivism and Neoliberal Institutionalism Compared," *International Studies Quarterly* 44, 97-119.
- Taylor, Michael, 1996, "When Rationality Fails," in Friedman, Jeffrey, ed., *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, New Haven: Yale University Press.
- Tversky, Amos, and Daniel Kahneman, 1986, "The Framing of Decisions and the Psychology of Choice," in Elster J., ed., *Rational Choice*, N.Y.: New York University Press.
- van Damme, Eric, Interviewer, 1998, "On the State of the Art in Game Theory: An Interview with Robert Aumann," *Games and Economic Behavior* 24: 181-210.
- Yee, Albert S., 1997, "Thick Rationality and the Missing "Brute Fact": The Limits of Rationalist Incorporations of Norms and Ideas," *Journal of Politics* 59(4): 1001-1039.

# The Concept of Rationality in the Rational Choice Theory

Joon Pyo Jung

There is much confusion about the concept of rationality in the rational choice theories. This paper tries to clarify the concept of rationality in the rational choice theories by comparing it with the concept of rationality in the philosophical and normative discussions. It is an instrumental rationality and what is rational in a theory depends more less on how it is defined in that theory. It is a simplification of the limited and incomplete rationality in the real world. And in many cases, it alone is insufficient to explain the phenomenon in question. What is rational in a given theory is embodied in the expected utility theory, Bayes' rule, and ultimately in the equilibrium concept used in the theory. Rational choice theories cannot have explanatory power without adequately representing how the actors subjectively perceive the situation. This means that a rational choice theorist needs in-depth knowledge about the situation in addition to the ability to model the situation and get the solution.

**Key Words :** rational choice theory, game theory, rationality, bounded rationality, expected utility, rational choice institutionalism