

사회적 선호이론 논쟁과 버논 스미스의 『도덕감정론』 해석*

황재홍**

요약

이 글은 인간 행동의 사회성을 설명하는 현대 이론들과 아담 스미스 『도덕감정론』의 연관성에 대해 살펴보고 있다. 사회적 선호이론 논쟁과 관련하여 사회적 선호이론가나 사회적 선호이론 비판자 모두 『도덕감정론』의 관점을 이어받고 있다고 주장하고 있으나 실제 이들 이론에는 『도덕감정론』의 중심 생각과는 일정한 차이가 존재함이 확인되고 있다. 현대 이론들의 아담 스미스 해석에 어떤 차이가 있으며 사회성을 설명함에 있어 아담 스미스의 핵심적 생각은 무엇인지 고찰한다. 그리고 버논 스미스의 『도덕감정론』 해석의 의미를 평가한다.

주제분류 : B030101

핵심 주제어 : 사회적 선호, 공감, 아담 스미스, 버논 스미스

I. 서론

최근 수십 년 동안 경제학에 나타난 근본적 변화 중 하나는 좁은 의미의 이기심 가정을 넘어서는 인간의 행동 동기에 대한 관심이라고 할 수 있다. 이는 많은 실험실 실험에 의해 촉발된 것이다. 최후통첩 게임, 독재자 게임, 신뢰게임, 공공재 게임 등의 실험에서 사람들은 전통적인 합리적 선택이론이 예측하는 바와 다른 행동들을 보였고 이러한 행동을 설명하고 해석하기 위한 여러 가지 시도들이 행동경제학, 실험경제학 등에서 행해지고 있다.

* 이 논문은 2015년도 목원대학교 연구년 지원에 의하여 연구되었음.

** 목원대학교 경제학과 부교수, e-mail: zootjhong@hanmail.net

그런데 이러한 흐름 내부에는 사회적 선호를 둘러싼 커다란 논쟁이 이어지고 있다. 사회적 선호 이론가들은, 위의 실험에서 나타나는 타인고려 행동은 인간이 이타심, 공정심, 상호성 등의 사회적 선호 또는 타인을 고려하는 선호를 가지고 있기 때문이라고 주장하며 이러한 행동을 설명하기 위해 전통적인 효용함수에 타인의 이득을 포함시키는 새로운 모델을 만들어냈다. 그러나 다른 한편에서는 전통적 이기심 가설에서 벗어나는 실험실 결과가 반드시 사회적 선호의 존재로 이어지는 것은 아니며, 평판효과 등 사실상 장기적인 의미의 이기심에 의해 설명될 수 있다는 주장이 제기되어 왔다.

이러한 논쟁에서 흥미로운 것 중 하나는 사회적 선호이론을 지지하는 진영과 비판하는 진영 모두 아담 스미스의 『도덕감정론』에 많은 관심을 보이며 자신들의 작업을 아담 스미스 이론과 연결시키고자 노력하고 있다는 것이다.¹⁾ 한편에서는 사회적 선호이론을 『도덕감정론』과 연결시키고 있는 반면 다른 한편에서는 『도덕감정론』을 근거로 사회적 선호이론을 비판하고 있다. 이와 관련하여 본 연구는 타인고려 행동을 설명하는 현대의 이론들에서 나타나는 아담 스미스 도덕이론과의 연관성을 확인하고, 이들 이론들이 행하고 있는 아담 스미스 해석에 어떤 차이점들이 존재하는가를 살펴보고자 한다.

본 연구는 이러한 문제들을 버논 스미스의 논의를 중심으로 살펴보고자 한다. 많은 현대 이론가들은 각자의 입장에서 자신들의 작업과 아담 스미스 이론의 연관성을 강조하고 있고, 아담 스미스와 현대이론의 연관성을 평가하는 문헌들은 현대 이론들 사이의 입장 차이에 대한 언급보다는 공통적으로 아담 스미스와 연결될 수 있는 부분들을 주로 서술하고 있다(Ashraf et al., 2005; Binmore, 1998; Gintis et al., 2005; Konow, 2012; Paganelli, 2009, 2011). 이와 달리, 버논 스미스는 아담 스미스의 저작을 적극적으로 자신의 이론과 연결시킬 뿐 아니라 아담 스미스의 저작을 근거로 사회적 선호이론을 비판함으로써 사회적 선호이론은 아담 스미스 이론에 부합하지 않는다는 주장을 명시적으로 제시하고 있기 때문이다. 따라서 사회적 선호이론 논쟁을 아담 스미스의 시각에서 평가함에 있어 버논 스미

1) 파가넬리(Paganelli, 2009)에 따르면, JSTOR 인용 건수가 1945과 55년 사이에 『국부론』이 1267건, 『도덕감정론』이 98건이었고 1995년과 2005년 사이에 『국부론』은 1804건, 『도덕감정론』 180건이었다. 『국부론』 인용 건수 증가율은 약 40%인 반면, 『도덕감정론』 인용 건수의 증가 비율은 거의 100%에 이른다.

스는 많은 논의거리를 제공하고 있다고 판단된다.

II. 버논 스미스의 아담 스미스 해석과 사회적 선호이론 비판

버논 스미스(Smith, 2010, 2012, 2013; Smith and Wilson, 2014)는 『도덕감정론』의 핵심원리가 사회적 행위의 규칙에 관한 공감적 근거라고 서술하고 있다. 그에 따르면, 아담 스미스에 있어 개인이란 상호 공감에서도 출되는 것과 독립적으로 정의될 수 없고 이 상호 공감은 인간 사회성의 핵심이 된다. 이러한 버논 스미스의 시각은, 심리가 행동의 영역에 들어올 때 이는 필연적으로 사회적인 것이며 사회적 맥락이 결여된 개인적 의사결정이라는 것은 없다는 주장으로까지 이어지게 된다. 인간은 타인의 악의적인 또는 호의적인 행위의 경험을 통해 그들의 행위를 읽고 판단하는 법을 배우게 되고 그 결과 타인도 같은 방식으로 자신의 행위를 판단하게 된다는 것을 알게 되는데 이러한 양방향 프로세스가 인간이 사회적으로 된다는 것의 핵심적 의미가 된다. 『도덕감정론』에 등장하는 공정한 관찰자는 이러한 양방향 프로세스의 은유적 상징이며 공정함이라는 단어에서 아담 스미스가 주목했던 것은 결과가 아닌 공정한 규칙을 지칭하는 것임을 버논 스미스는 강조하고 있다. 인간의 사회성은 규칙을 따르는 행동으로 나타나는데 행위자와 그 행위의 영향을 받는 타인이 주어진 맥락에서 그 행위의 적절성을 판단하게 되는 것이다. 『도덕감정론』에 나타난 사회성의 의미가 이와 같은 것이라고 할 때, 사회적 선호이론은 이러한 사회성의 핵심에 다가가지 못한 다는 것이 버논 스미스의 비판이다. 인간의 행동 동기는 사회적 선호이론의 주장대로 결과에 대한 효용에서 나오는 것이 아니라 주어진 맥락에서 행위의 적절성에 대한 판단에서 나온다는 것이다.

버논 스미스는 자발적 신뢰게임과 비자발적 신뢰게임의 결과를 비교하면서, 사회적 선호이론이 맥락의 변화에 따른 행동의 변화를 설명할 수 없지만 『도덕감정론』에는 이를 설명할 수 있는 내용들이 담겨 있음을 지적하고 있다(Smith, 2012). 자발적 신뢰게임은 다음과 같이 설계된다. 행위자 A는 두 가지 중 하나를 선택하게 된다. 자신과 상대방이 모두 20달러((20,

20), 앞에 것이 행위자 A의 몫, 뒤에 것이 행위자 B의 몫)씩 받는 것을 선택하거나 아니면 행위자 B가 (25, 25)와 (15, 30) 중 하나를 선택할 수 있도록 선택권을 넘긴다. 호모 이코노미쿠스에 입각한 전통적인 경제학은 행위자 A가 (20, 20)을 선택할 것으로 예상한다. 왜냐하면 행위자 B는 당연히 (15, 30)을 선택할 것이라고 보기 때문이다. 그러나 위 실험의 결과는, 행위자 A 27명 중 17명이 (20, 20)을 선택하는 대신 행위자 B에게 선택권을 넘겼고 선택권을 넘겨받은 17명의 행위자 B 중 11명이 (15, 30)이 아닌 (25, 25)를 선택한 것으로 나타났다. 사회적 선호이론에 따르면, 이러한 결과가 나타난 것은 사람들이 타인을 몫까지 고려하는 사회적 선호를 가지고 있기 때문이라고 해석할 수 있다. 그러나 『도덕감정론』에 따르면 행위는 결과가 아닌 적절성에 대한 판단에 의한 것이다. 버논 스미스는, “인간은 칭찬을 원할 뿐 아니라 칭찬받을 만함을 갈망한다. 비난을 두려워할 뿐 아니라 비난받을 만함도 두려워한다.”는 『도덕감정론』의 문장을 사회적으로 도출된 동기 공리(Motivation Axiom)라 칭한다. 이는 인간이 자제심을 통해 자신의 행동을 통제하여 더 나은 행동을 추구하려고 노력함을 의미한다. 더 나아가 “우리는 공정한 관찰자가 우리의 행동을 관찰한다고 상상하는 것 같이 우리 자신의 행동을 관찰하려고 노력한다.”는 문장을 인용하면서 여기서 공정함이란 “페어플레이를 위반하지 않으려 한다”는 것을 의미함을 강조한다. 즉 선택에 있어 이 행동이 공정한가 반칙인가를 물어보고 있음을 의미한다는 것으로 해석하고 있다. 공정함은 적절한 처신(conduct)에 관한 것이지 결과의 효용에 관한 것이 아니라는 지적이다. 인간이 자신을 위한 행동을 하려는 성향을 가지고 있지만 이러한 자기애(self love)적 행동은 규칙에 의해 제한되는 것이다. 이와 같은 방식으로 버논 스미스는, “적절한 동기에서 나오는 선행만이 보상을 요구할 수 있을 것으로 보인다. 왜냐하면 이러한 행동만이 감사의 대상으로 승인되며 관찰자의 공감적 감사함을 유발시킬 수 있다”는 『도덕감정론』의 문장을 행위 판단에 관한 명제 1로 규정하고 있다.²⁾ 위 실험에서 행위자 A가 행위자 B에게 선택권을 넘기는

2) 버논 스미스가 『도덕감정론』에서 뽑아낸 명제 2, 3은 다음과 같다.

명제 2: 부적절한 동기에서 나오는 악의적 행동만이 처벌의 대상이 된다. 왜냐하면 이러한 행동만이 분개의 대상이 되기 때문이다.

명제 3: 선행은 자발적인 것이다. 그것은 강요될 수 없다. 그리고 그것이 부족하다고 해서 처벌되지 않는다.

것은 자발적인 선행이고 칭찬받을 만한 행동이며 감사의 대상이 되는 행동이다. 이 행동은 보상을 받을 수 있으나 이것이 행동의 주요 동기는 아니며 오히려 보상이 주어진다면 이는 페어플레이에 대한 인정을 의미한다.

버논 스미스는 이 실험을 약간 변형한 두 번째 실험의 결과를 통해, 사회적 선호이론은 이러한 과정과 맥락의 중요성을 이해하는데 한계가 있으며 과정과 맥락의 이해가 왜 중요한지를 설명하고 있다. 첫 번째 실험에서 행위자 A에게 두 가지 선택권이 있었던 것과 달리 비자발적 신뢰게임 실험에서는 행위자 A가 (20, 20)을 선택할 수 없고 무조건 행위자 B에게 선택권을 넘겨야 하는 상황이다. 이 경우 행위자 B는 행위자 A에 대해 어떤 자애심도 느끼지 못하며 또한 감사함을 느끼지 않는다. 따라서 자발적 신뢰게임과 비교할 때 더 많은 행위자 B가 (25, 25)가 아닌 (15, 30)을 선택하는 결과가 나타난다. 이 실험 결과는 행동의 암묵적 규칙과 의도를 포함하는 맥락을 중요시 하는 『도덕감정론』의 주장과 일치한다는 것이 버논 스미스의 해석이다. 사회적 선호이론에 따르면 사람들은 타인의 몫까지 고려하는 사회적 선호를 가지므로 자발적 신뢰게임에서나 비자발적 신뢰게임에서나 행동의 변화가 없어야 하는데 이러한 실험 결과는 사회적 선호이론의 예상과 어긋나는 것이 된다.

이와 같이 버논 스미스는 인간의 사회성을 이론화하는데 있어 『도덕감정론』으로 돌아가자고 주장하고 있다. 사회적 선호이론이 호모 이코노미쿠스에 기초한 표준경제학 모델의 약점을 드러내는데 큰 기여를 했다고 인정할 수 있지만, 사회적 선호이론은 필연적으로 신고전학파의 구제에 이르게 될 것이라고 그는 주장한다. 신고전학파의 대한 자신의 비판은 신고전학파가 공리주의적 근거를 갖는다는 점이며, 공리주의적 근거에서 벗어나야 한다는 관점에서 볼 때 사회적 선호이론은 자신보다 신고전학파 전통을 더 용인하고 있다는 주장이다.

Ⅲ. 사회적 선호이론과 『도덕감정론』

아담 스미스에 있어 공감을 통한 도덕적 승인은 결과에 대한 효용에 기인하는 것이 아니며 따라서 사회적 선호를 효용함수에 삽입하는 방식은 『도덕

감정론』을 올바르게 해석하는 것이 아니라 버논 스미스의 주장은 기존의 아담 스미스 해석과 맥락을 같이 하는 것으로 보여진다. 많은 아담 스미스 연구자들이 도덕적 행동에 대한 아담 스미스의 설명은 공리주의와 대립되는 것으로 보아야 한다는 해석을 제시하였다(Martin, 1990; Sugden, 2002; Witztum and Young, 2013). 흄과 스미스의 여러 가지 공통점에도 불구하고, 효용이 도덕성의 근거라는 흄의 주장에 반대하여 효용은 도덕적 승인의 근거가 될 수 없다고 한 아담 스미스의 대응은 잘 알려져 있다(Smith, 1976, pp.188-189). 사람의 마음씨나 성격과 마찬가지로 어떤 사물도 개인이나 사회의 행복을 증진시킬 수 있는데 만약 도덕적 승인의 원천이 이러한 행복을 증진시키는 것이라면 인간 뿐 아니라 잘 고안된 빌딩에 대해서도 도덕적 승인을 느낄 것이라고 아담 스미스는 주장하였다. 또한, 어떤 성향의 유용성은 도덕적 승인의 첫 번째 근거가 될 수 없다고 주장했다. 승인의 감정은 효용의 인식과는 구분되는 적절성의 감정을 포함한다.

우리가 어떤 사람의 판단을 그 유용성이 아니라 진실과 현실에 대해 올바르게 정확하고 동의될 만한 것으로서 승인한다. ... 어떤 자질들의 효용이라는 생각은 단순히 사후적인 사고이지 우리의 승인을 위해 첫 번째로 추천되는 것이 아니다(Smith, 1976, p.20).

아담 스미스에 있어, 도덕적 판단은 효용이 아니라 일상의 경험에서 느끼는 감정에서 나오는 것이다. 일상의 상호작용 중 어떤 것이 동의될 만한 것이고 어떤 것이 승인될 만한 것인지에 대한 인식을 형성시키는 것이다.

이러한 해석에도 불구하고 『도덕감정론』의 유명한 첫 문장(인간은 단지 그것을 바라보는 즐거움 이외에 아무 것도 얻는 것이 없을 지라도 타인의 행복을 필요로 한다)은, 자신의 효용함수에 타인의 이익을 포함시키는 방식으로 사회적 행위가 설명될 수 있다는 주장의 근거로 사용되어 왔다.³⁾ 그리고 전통적인 합리적 선택이론의 틀을 유지하면서 이타적이고 협조적 행위를 설명하려는 경제학 내에서의 시도들이 『도덕감정론』의 첫 문장과 형식적으로 일치하는 이러한 접근법을 채택해왔었다. 그 대표적인 사례가 베커라

3) 버논 스미스(Smith, 1998)도 『도덕감정론』의 영향을 본격적으로 인정하기 이전에는 이 문장을 들어 아담 스미스가 사회적 선호이론가들과 같은 공리주의적 사고를 하고 있었다고 서술하였다.

고 할 수 있다. 베커(Becker, 1976) 역시 자신의 효용함수에 타인의 소비까지 포함시키는 효용함수($U_i = U_i(X_i, X_j)$)를 통해 이타적 행위를 설명해 왔다. 사회적 선호이론가들이 사회적 선호를 자신에게 배당되는 물질적 자원 뿐 아니라 타인에게 배분되는 물질적 자원에 대한 관심으로 정의하고 있으며(Fehr and Fischbacher, 2002), 사회적 선호이론의 접근법이 신고전학파의 합리성 가정을 유지하면서 효용함수의 형태를 바꾸는 것(Fehr and Schmidt, 2006; Burnham, 2013)이라고 서술되고 있다는 점을 볼 때, 형식논리적인 측면에서 베커와 사회적 선호이론은 유사한 접근을 하고 있다. 그러나 이타적 행동에 관한 이러한 ‘자기중심적인’ 접근은 아담 스미스가 명시적으로 비판해왔던 것이며 공감에 근거한 아담 스미스의 접근과는 분명한 차이를 보인다고 볼 수 있다. 공감에 의한 행동은 자신을 타인의 상황에 대입하는 것이지 자신의 입장에서 타인의 즐거움을 판단하는 것이 아님을 분명히 하고 있다. 타인의 상황을 나에게 발생한 것으로 상상하는 것이 공감이라면 공감은 이른바 ‘이기적 공감’이 될 수 있다. 그러나 아담 스미스는 다음과 같은 주장을 펼치고 있었다.

공감은 어떤 경우에도 이기적인 원리로 간주될 수 없다. 내가 너의 슬픔에 공감할 때 나의 감정은 자기애에 근거하고 있는 것으로 보일 수 있다. 왜냐하면 이 공감은 당신의 문제를 나의 문제로 인식함으로써 내가 당신의 처지에 있다면 내가 그 상황에서 어떻게 느끼게 될지를 상상하는 것에서 생겨나기 때문이다. 그러나 공감이 상상적 입장 변화에 의한 것이기는 하지만 이러한 상상적 입장 변화는 나에게 일어났다고 가정하는 것이 아니라 내가 공감하는 사람에게 일어났다고 가정하는 것이다. 당신이 아들을 잃은 것을 내가 위로하려고 할 때 너의 슬픔에 들어가기 위해 나는 내 아들이 죽었다면 얼마나 슬플까를 상상하는 것이 아니라 내가 정말 너라면 얼마나 슬플까를 생각하는 것이다. 환경과 입장만 바꾸어보는 것이 아니라 사람과 성격도 바꾸어 보는 것이다. 그러므로 나의 슬픔은 너에게 달려 있는 것이며 그러므로 이는 전혀 이기적인 것이 아니다 (Smith, 1976, p.317).

IV. 평판과 『도덕감정론』

그렇다면 사회적 행위에 대한 사회적 선호이론 비판자들의 설명은 아담 스미스의 시각에 얼마나 다가가고 있는가?

사회적 선호이론 논쟁과정에서, 전혀 모르던 타인과의 일회성 상호작용에서도 사회적 행동을 보이는 것은 사회적 선호에 기인하는 것이라는 사회적 선호이론에 대해 여러 가지 반론들이 제기되었다. 가령 게일(Gale et al., 1995)의 경우 최후통첩게임에서 공정한 제안의 존재, 그리고 낮은 제안에 대한 거부를 이기심에 근거한 학습모델을 통해 설명하고 있다. 낮은 금액의 제안을 거부하는 것은 응답자에게 그렇게 큰 손해가 아닌 상황이므로 제안을 거절하는 것이 손해라는 것을 느리게 학습하게 된다는 것이다. 반면 제안을 거부당하는 것은 제안자에게는 큰 손실이므로 낮은 금액을 제한하는 것이 자신에게 이득이 되지 않는다는 것을 빨리 학습하게 된다는 것이다. 따라서 부분게임완전균형으로 수렴이 매우 느리게 진행되며, 이들의 시뮬레이션은 수 천 번의 반복이 있어야 표준경제학이 예측하는 균형점에 도달하게 된다는 것을 보여주고 있다.

또한 예상보다 높은 수준의 협력이라는 실험 결과는 정보가 완전해도 사람들이 실험실 상황을 실제와 구분하지 못해서 발생한다는 해석도 제기되었다(Binmore, 2006; Burnham and Johnson, 2005). 실제 생활에서는 실험실과 달리 반복적 상호작용이 발생하고 있고 이것에 익숙해져 있는 사람들이 일회성 상호작용이 일어나는 실험실에서도 일상생활의 행동 패턴을 그대로 보여준다는 해석이다. 이러한 해석을 지지하는 실험결과들도 보고되고 있는데, 가령 실험참가자가 게임에서 사용하는 컴퓨터 화면에 눈 모양의 그림을 넣는다거나, 게임에 영향을 미치지 않는 사람들과 상호작용을 하게 되면 협력의 수준이 달라지는 것으로 나타났다. 이렇듯 단순하고 작은 상황 변화가 실험실 밖의 상황에 대한 대응과 같은 행동을 초래하고 있음을 보여주고 있는 것이다. 또한 지역적, 문화적 차이에 따라 실험결과가 달라진다고 보고되고 있는데 이는 실험결과가 일상생활과 연관되어 있다는 것을 보여주는 것으로 해석될 수 있다.

이러한 사회적 선호이론 비판은 진화심리학에 의해서도 이루어졌다(Price, 2008; Price et al., 2002; Tooby et al., 2006). 사람들이 협

조적으로 행동하는 것은 문화적 집단선별과 유전자-문화 공진화의 과정을 거치면서 친사회적 유전자가 환경에 적응한 결과라고 주장하는 사회적 선호이론과 달리, 부적응/미스매치이론은 과거 소규모집단에서의 행위 습성이 현재에도 여전히 남아 있기 때문에 현대의 대규모 집단에서는 이기적 행동이 자신의 생존가능성을 높일 수 있음에도 불구하고 사람들이 잘못(즉 이타적으로) 행동하고 있다는 주장을 제시하고 있다. 즉 인간의 협조적 행위는 인간이 환경에 제대로 적응하지 못한 결과라는 것이다.⁴⁾ 인간은 과거 오랜 동안 소규모 사회에서 생활하면서 상호작용은 주로 근친 간에 반복적으로 이루어져왔고, 상대적으로 익명적인 타인과의 일회적인 교류는 거의 없었다고 진화심리학자들은 주장한다. 소규모 사회에서 생존을 위해서는 구성원들 간에 협조는 필수적이었고, 이 과정에서 인간은 그러한 환경에 적응하여 여기에 적합한 심리 메커니즘을 길게는 수백만 년 동안 진화시켜 왔다. 사회적 선호이론가들이 자신들의 주장의 근거로 제시하는 실험에서, 피실험자가 처한 환경은 인간의 인지 메커니즘이 진화한 환경과는 아주 다르기 때문에 피실험자들의 행동은 적합성 차원에서 부적응적 행태를 낳게 된 것이다. 따라서 그러한 실험 결과는 단지 과거 인간이 진화해온 조건과 다른 실험실 환경에 피실험자들이 제대로 적응하지 못했다는 것을 보여줄 뿐이라는 주장이다.

그러나 이렇게 인간의 모든 행동 동기를 이기심으로 환원하고, 인간에게서 관찰되는 사회적 행위 역시 장기적 이익을 얻기 위한 전략적 행위로만 보는 이러한 시각 역시 아담 스미스의 견해와는 큰 차이를 보여주고 있다.

아담 스미스는 분명 『도덕감정론』에서 인간은 타인에게 인정받기를 원하고 자신의 평판을 높이려는 성향을 가진다고 서술하고 있으나 타인을 고려하는 행위와 장기적 이기심에 기초한 전략적 행위를 동일시하지는 않았다. 아담 스미스는 타인에 대한 감정과 자기에는 모두 덕스러운 것으로 보았지만 덕에 대한 사랑이 인간 본성의 최고 열정이며 자기 자신의 평판, 영광에 대한 사랑은 그보다 못한 덕이라고 주장하였다. 덕에 대한 사랑과 영광에 대한 사랑은 비슷한 면이 있지만 가장 큰 차이는 사람들이 칭찬을 받을 수

4) 다른 동물들에게도 이러한 부적응은 관찰된다. 예를 들면, 재갈매기의 둥지에 크기와 색깔이 다른 알을 넣어 놓아도 어미 새는 계속해서 그 알을 품는다(Price, 2008). 이는 어미 새가 달라진 환경을 제대로 인지하지 못하고 예전과 동일하게 행동하는 부적응의 예라고 할 수 있다.

없는 경우에도 그것이 옳다고 생각되면 시행한다는 것이다. 이러한 아담 스미스의 시각은 모든 행동 동기를 이기심으로 보는 맨더빌의 시각과 구분되며 아담 스미스는 자신과 맨더빌의 차이를 분명히 하고자 했다(Smith, 1976, pp.306-314). 아담 스미스는, 덕과 악덕을 구분하는 에피쿠루스 학파, 스토아 학파, 허치슨 등과 달리 맨더빌은 그렇지 않다고 평가하고 있다. 아담 스미스의 시각에서 볼 때 맨더빌의 가장 큰 오류는 모든 감정을 다 악덕이라고 한 점이다. 그래서 타인에 대한 감정을 모두 다 허영이라고 생각했고 이런 생각에 기초하여 '개인의 악덕이 사회의 공익'이 된다는 주장에 이르게 되었다는 것이다. 맨더빌의 시각에서 보자면 아름다운 예술에 대한 관심, 인간생활의 향상에 대한 관심 등이 다 사치이고 과시이다. 따라서 세상에 진정한 덕은 없으며 이는 모두 사기이고 기만이 되는 것이다. 이에 비해 아담 스미스는 자기에 역시 때로는 덕스러운 행동이며 평판과 명성을 추구하는 행동을 허영이라고 볼 수 없다고 주장한다. 이러한 아담 스미스의 도덕이론은 홉스나 맨더빌과 같이 이기심을 유일한 행동 동기로 보는 자아중심주의와 구분되며, 평판을 허영으로 보지 않는다는 점은 많은 아담 스미스 연구자들의 공통된 해석이다(Clark, 2009; Hanley, 2010; Kahlil, 2001; Martin, 1990).

V. 버논 스미스와 『도덕감정론』

3, 4절의 논의에 비추어 볼 때 실험실과 사회에서 관찰되는 인간의 사회적 행동을 사회적 선호이론으로 설명하는 것도, 평판효과나 장기적 이기심에 의해 설명하는 것도 아담 스미스의 도덕이론에 완전하게 부합하지 않는다는 점을 확인할 수 있다. 이 두 가지 설명방식이 아담 스미스의 『도덕감정론』에 잘 부합하지 않는다는 지적은 버논 스미스에 의해서도 행해졌다. 버논 스미스는 전통적 합리적 선택이론으로 설명되지 못하는 실험실 실험의 결과를 해석하기 위한 두 가지 시도가 있었음을 지적하면서 그것이 각각 사회적 선호이론과 상호성이론에 해당된다고 보았다. 이는 모두 효용극대화 모델이 부정되는 결과에 대한 사후적 합리화였을 뿐이라는 평가다. 사회적 선호함수는 실험 데이터에 일치하는 대답을 할 수 있는 효용함수를 삽입함

으로써 진정한 사회성을 설명하는데서 도망쳤고, 상호성이론은 상호성이라는 이름 자체로 설명이 되는 것이라고 상상하면서 사회적 행위의 이유를 더 깊이 이해하는 방법의 부족을 겪게 되었다고 평가하였다. 아담 스미스의 개념과 비교할 때 현대 경제학의 상호성, 생물학에서의 상호적 이타심의 개념은 과도하게 단순화된 기계적 환원일 뿐이라고 주장했다(Smith, 2010, 2015).

이러한 비판을 통해 버논 스미스는 『도덕감정론』에 근거한 새로운 행동 모델을 제시하고 있다. (Smith & Wilson, 2014, p. 6)

$$a_i(\text{Propriety}|C) = \alpha_i(C)(PR) + \beta_i(C)(PR) \cdot (PW) + \gamma_i(C)(PW) + \delta_i(C)$$

행위자 i 에 의해 행해지는 행동 a_i 는 주어진 상황(C)에서 적절하다고 판단될 수 있어야 하는데 여기에는 사회적 칭찬(PR)과 칭찬받을만함(PW) 그리고 사회성 지표와 무관한 자기애(self love)적 요소들($\delta_i(C)$)이 영향을 미치게 된다.⁵⁾ 버논 스미스는 『도덕감정론』의 주장을 받아들여 인간은 자기애에 기초한 행동을 우선적으로 고려하지만 그 오만함은 제어될 수 있으며, 인간의 본성은 칭찬과 칭찬받을만함을 추구하고 비난과 비난받을만함을 피하고자 하는 성향을 지녔다고 서술하고 있다. 칭찬과 칭찬받을만함은 연결되어 있지만 독립적인 것이다. 만약 칭찬을 받더라도 그것이 썩 칭찬받을만한 것이 아니었다고 판단된다면 큰 즐거움과 만족감을 가져다주지 못한다.

버논 스미스의 이러한 서술은 앞서 언급된 두 그룹들에 비해 아담 스미스 도덕이론에 훨씬 더 가까이 다가가 있는 것으로 보인다. 그러나 아담 스미스의 관점에서 실제 실험 결과들을 해석하는 과정에서 중요한 괴리들이 나타나고 있다. 앞서 사회적 선호비판에 사용한 신뢰게임의 맥락을 변경하여 실험자조차 피실험자 개개인이 누구인지 확인할 수 없도록 익명성이 높아진 이중은폐 실험에서 행위자 B의 이기적 행동이 증가하는 것((25,25)가 아

5) PR 과 PW 는 0 또는 1의 값을 갖는다. 어떤 행위가 칭찬이나 칭찬받을만한 가치가 있을 때 1, 그렇지 않을 때 0이 된다. 그리고 $\alpha_i, \beta_i, \gamma_i$ 는 적절함을 결정함에 있어 PR 과 PW 의 가중치를 부여하는 함수이다.

닌 (15,30)을 선택하는 것)으로 실험결과가 나타나고 있다. 이에 대해 버논 스미스는, 사적이고 비밀적인 성격이 증가한 상호작용의 상황은 공정한 관찰자가 자신의 몫을 증가시키는 선택을 보다 쉽게 승인할 수 있는 경우라고 서술하고 있다. 또한 단일은폐의 상황에 비해 이중은폐의 상황은 칭찬받을만함도 감소하고 비난받을만함도 감소한다는 설명을 덧붙이고 있다 (Smith and Wilson, 2014, p.19). 버논 스미스도 아담 스미스가 칭찬과 칭찬받을만함을 구분했음을 언급하고 있음에도 불구하고 이러한 설명은 버논 스미스가 생각하는 칭찬받을만함과 아담 스미스가 생각하는 칭찬받을만함의 개념에 차이를 보여주고 있다. 아담 스미스에 있어 덕이 완전하게 발현되는 때는 칭찬이나 보상과는 무관한 경우이기 때문에 칭찬을 좋아하고 영광을 추구하는 자는 이류로 취급된다.

현명한 사람은 칭찬받을 만함이 없는 곳에서 받는 칭찬에 별 즐거움을 느끼지 못하는 반면, 칭찬을 받지 못한다는 것을 알고 있어도 자신이 칭찬받을 만한 것이라고 생각한 것을 시행할 때 최상의 기쁨을 느낀다 (Smith, 1976, p.117).

명성, 영광에 대한 사랑은 ... 현명한 사람에게 가치가 없는 것은 아니지만 그는 때로 이를 무시하고 경멸한다. ... 이 경우 자기 승인은 타인의 승인을 필요로 하지 않는다. 자기 승인만으로 충분하고 만족한다 (Smith, 1976, p.117).

이런 점들에 근거할 때 아담 스미스는 상호적 보답이나 평판에 대한 기대 없는 선행이 들어설 자리를 만들려고 노력했다고 해석할 수 있다. 그러나 이중은폐 신뢰게임에서 버논 스미스는 타인이 자신을 정확하게 인식할 수 있는가의 정도에 따라 사회적 행위의 정도가 달라지는 것을 칭찬받을만함, 비난받을만함과 연결시키고 있음을 확인할 수 있다. 그런데 이는 앞 절에서 언급된, 실험실에서 나타나는 사회적 행동은 평판이나 장기적 이득을 기대한 것이라는 설명으로 되돌아 갈 수 있음을 의미한다. 이는 아담 스미스의 영향력을 본격적으로 인정하기 이전의 시기에 버논 스미스가 실험실에서 관찰되는 사회적 행위에 대해 가지고 있었던 기본적인 시각이었다.

협력에 대한 가장 포괄적인 이해는 실험주체들이 사회교환 경험을 실험실로 가지고 와서 게임이론의 추상적 원리들을 자동적으로 따르지 않는다는 가설에서 나온다고 나는 생각한다. ...

많은 사람들은 일회성 게임의 실험결과를 가지고 이기심 가설은 기각되어야 하며 따라서 타인고려 선호를 가정해야 한다고 주장한다. 그러나 반복적 상호작용에서 사람들은 타인고려 행동을 하기 위해 타인고려 선호를 꼭 가져야 하는 것은 아니다. 아담 스미스가 이야기한대로 ‘친절은 친절의 어머니’이며 달라이 라마가 이야기한 것처럼 ‘자신을 위하는 현명한 방법은 타인의 후생을 위해 일하는 것이다.’ 이 타인을 향한 친절함과 타인의 후생을 위해 일하는 것은 장기적인 자기만족 -너의 평판은 가치 있다- 과 일치하는 행동의 세련된 형태이다. 따라서 사람들이 좁은 의미의 이기심 모델이 예측한 결과와 일치하지 않는 실험결과를 보고 타인의 이익까지 고려하는 공리주의적 선호를 지지할 수도 있지만, 주체들의 의사결정 프레임이 일회적 게임이라는 게임이론적 개념과 일치한다는 암묵적 보조가설을 부정할 수 있다... 이러한 사회적, 주체들이 일상의 경험을 실험실로 가지고 오는 경향을 보여주는 창은 일회성 게임에서 맥락의 변화를 살펴보는 것이다. 맥락은 과거 경험을 담고 있어서 주체들이 일회적 게임을 일상의 반복적 게임과 분리된 게임으로 인지하는지를 확인할 수 있다(Smith, 2008, pp.201-202).

버논 스미스는 실험실에서 타인고려 행동이 관찰된다고 해서 이것이 이기심에 반하는 것이 아니며 처벌을 피하고 평판을 유지하기 위한 바램에서 나오는 것이라는 시각을 견지해왔다. 이러한 생각은 코스미데스, 투비 등 진화심리학의 영향을 받은 것임을 버논 스미스는 명시적으로 서술하고 있다 (Hoffman, McCabe and Smith, 2008, p.411).

이에 반하여 아담 스미스는 칭찬을 받지 못하거나 처벌의 위협이 없을 때에도 옳은 일을 할 수 있다고 주장하고 있으며 이것이 칭찬과 칭찬받음만함의 구분 속에서 나타나고 있다. 이는 사회적 행위를 장기적인 이기심, 평판 등에 의해 설명하려는 시도에 대한 경고라고 할 수 있다. 사회적 선호이론가인 긴티스는 *Moral Sentiments and Material Interests*에서 이 책의 아이디어는 아담 스미스의 지적 유산의 연장선에 있다고 서술하였다. 사회 선호이론 비판자들의 주장과 같이 이타심이 잘못 적용된 이기심이라면 문명화된 사회에서 나타나고 있는 발전된 협력들이 붕괴될 것이라고 주장했다

(Gintis et al., 2005). 사회적 선호이론은 평판효과나 장기적 이기심과는 구분되는 사회적 선호의 존재를 확인하고자 한다는 점에서, 이 지점에서는 사회적 선호이론을 비판하고자 했던 버논 스미스보다 사회적 선호이론이 아담 스미스의 칭찬받을만함의 개념에 더 다가가고 있는 것으로 보인다.

버논 스미스가 평판이나 장기적 이기심에 근거하여 실험 결과를 설명하는 입장에 가깝다고 하는 점은 아담 스미스와 맨더빌의 관계를 바라보는 버논 스미스의 시각에서도 드러난다. 앞 절에서 언급된 바와 같이 아담 스미스는 인간의 모든 행동 동기를 이기심으로 환원하려는 맨더빌과의 차별성을 드러내고자 했으나, 버논 스미스는 아담 스미스와 맨더빌의 공통점에 더 주목하였다. 버논 스미스는 2008년 저서에서 개인의 이기적인 행동이 의도하지 않게 사회적 공익을 가져다준다는 메시지가 스코틀랜드 계몽주의 철학자들에게서 공통적으로 나타난다고 주장하고 있다. 여기서 맨더빌을 언급하면서 아담 스미스가 맨더빌의 영향을 많이 받았다고 서술하고 있다. 그리고 2010년 논문에서는 '개인의 악덕이 사회적 공익'이 된다는 맨더빌의 명제가 『도덕감정론』에도 나타난다고 주장하였다.

이런 버논 스미스의 서술들을 볼 때, 버논 스미스는 전통적인 합리적 선택이론으로 설명되기 어려운 실험실 결과를 해석하는 두 가지 이론, 즉 사회적 선호이론과 상호성이론이 모두 만족스럽지 못하다고 비판하고 있고, 사회적 선호이론은 『도덕감정론』의 핵심에 다가가지 못했다고 비판하고 있지만 상대적으로 평판이나 장기적 이기심에 의존하는 상호성이론과 『도덕감정론』의 차이에 대해서는 덜 주목하고 있는 것이 확인된다. 나아가 상호성이론과 『도덕감정론』은 훌륭한 보완관계로 인식하고 있는 것으로 보인다. 버논 스미스의 직접적인 언급은 아니지만, 실험실에서 나타나는 협력적 행동을 설명하기 위해 『도덕감정론』과 당대의 문헌들을 적극적으로 활용하는 버논 스미스의 작업을 높이 평가하고 있는 파가넬리의 다음과 같은 언급은 이기심과 사회적 행동 사이의 관계에 대한 버논 스미스와 그와 동료들의 시각을 잘 드러내고 있다.

인간이 왜 그리고 어떻게 협력하고 신뢰하며 산업사회에서 번영하고 있는가에 대한 보다 완전한 그림을 만들어내기 위해 공감, 도덕적 상상력, 칭찬받을 만함에 관한 아담 스미스의 생각을 실험경제학의 결과에서

도출되는 설명들과 결합시킬 것을 제안한다. 그렇다고 이기심에 기초한 전통적인 설명의 중요성이 감소된다고 하는 것은 결코 아니다. 다른 어떤 것을 이기심과 동등한 위치에 두려는 것도 아니다. 이기심은 여전히 인간 협력의 중요 동력이다. 타인고려 선호나 도덕감은 이기심을 대체할 수 없다. 다만 이기심에만 기초하여 이방인 사이의 협력을 설명하는 방식에 존재하는 빈 공간을 메울 수 있을 것이다(Paganelli, 2011, p.247).

VI. 공감, 감정이입, 동료의식

위의 서술에서 나타나듯이 사회적 선호이론 진영이든 이를 비판하는 진영이든 각각 아담 스미스와의 연관성을 주장하고 있으나 각자에게 비춰지는 아담 스미스 도덕이론의 모습은 큰 차이를 보이고 있다. 이러한 차이가 발생하는 이유 중 하나는 아담 스미스가 사용하는 공감(sympathy)의 의미가 여러 가지로 해석될 수 있기 때문으로 보여진다.

인간은 바라보는 기쁨만으로도 타인의 행복에 관심을 갖는다는 『도덕감정론』의 표현을 근거로 공감이 이타심, 자비심 등으로 해석된다면 이는 자신의 효용함수에 타인의 물질적 이득이 포함되는 방식으로 표현될 수 있다. 그리고 전통적인 합리적 선택이론의 틀에서 이타적 행위를 설명할 수 있다. 이러한 시각이 베커 등의 이론에 적용되어 왔고 현재 사회적 선호이론에서 사용하는 모델도 이러한 방식을 따르고 있다고 볼 수 있다.

반면 빈모아의 경우 공감에 관한 아담 스미스의 복잡한 개념을 감정이입(empathy)이란 개념으로 대체하자고 주장하였다(Binmore, 1998, p.12). 아기에 대한 엄마의 사랑과 같이, A가 B의 목적을 확인하고 B의 후생이 A의 효용함수에 포함될 때 A는 B에 공감한다고 말할 수 있다면, 이와 대비하여 A가 B의 관점에서 사물을 바라보기 위해 B의 입장에 서본다면 A는 B에 감정이입을 한 것이라고 말할 수 있다고 빈모아는 정리하고 있다. 이러한 감정이입의 경우 A는 B의 후생을 돌볼 생각 없이 자신을 B와 동일시할 수 있다. 가령 총싸움에서 이기려고 목적으로 상대방의 움직임에 예측해 보기 위해 상대방의 입장에서 생각해 보는 것을 예로 들 수 있다. 이러한 정의에 따르면 공감의 대상은 타인의 후생이지만 감정이입의 목적은

이해가 된다. 이러한 개념 정리를 통해 빈모아는 소위 '아담 스미스 문제'⁶⁾의 해법은 공감을 감정이입으로 재해석하는데 있다고 주장한 바 있다 (Binmore, 1998, pp.367-369). 빈모아에 있어 이 때 감정이입 선호는 개인적인 취향이나 선택과는 구분되어야 하는 사회적 가치이다. 개인적 선호에 따른 상호작용에서 도출되는 결과가 여러 개일 경우 그 여러 균형 중 하나가 선별될 때 감정이입적 선호가 사용되는 것이라고 그는 생각했다. 이러한 감정이입의 선호는 문화적 가치를 담고 있는 것으로 공정한 관찰자라는 은유를 통해 나타난다. 그러나 이는 개인의 선호나 취향이 아닌 사회적 규범이 된다. 이러한 해석에 따라 경우 아담 스미스의 문제가 해결되며 인간 행동을 이기심에 기초한 호모 이코노미쿠스의 개념으로 이해하는데 아무런 문제를 발생시키지 않는다고 빈모아는 주장하고 있다. 이러한 주장은, 실험실에서 관찰되는 사회적 행위들은 타인의 효용이 자신의 효용에 영향을 미치는 형태의 공감에 의한 사회적 선호로 인하여 나타나는 결과가 아니라 협동이 필요한 현실에 적응되어 있는 실험참가자들이 현실과 실험을 구분하지 못한 결과, 게임의 구조를 정확히 이해하지 못한 결과이며 이러한 혼돈이 사라지면 실험실에서 사람들의 사회적 행위는 감소한다는 해석으로 이어지게 된다. 사회적 선호이론가들은 사람들이 공정하게 행동하는 것을 좋아하기 때문에 공정하게 행동한다는 무의미한 동어반복을 하고 있는 것이며 사회규범과 선호의 구분을 흐리고 있다고 빈모아는 주장하였다(Binmore, 2010).

석덴(Sugden, 2002)은 『도덕감정론』에 등장하는 공감 개념은 위와 같은 의미의 공감, 감정이입을 넘어서며, 아담 스미스 공감 개념의 진정한 특징은 동료의식(fellow feeling)이라는 용어로 나타낼 수 있다고 주장하였다. 이는 감정의 조응을 의미한다. 이것은 타인의 기쁨을 보고 자신도 기뻐지고 타인의 슬픔을 보고 자신도 슬퍼지는 메커니즘 이외에 그 감정이 기쁨이든 슬픔이든 타인이 자신과 동일한 감정을 공유하고 있다는 사실로부터 즐거워질 수 있는 메커니즘을 말하는 것이다. 그리고 아담 스미스 체계에서 이러한 감정의 조응이 도덕에 기초한 승인과 연결된다. 이러한 승인의 심리적 메커니즘은 상호작용 하는 집단 내에서 감정의 적절성에 대한 규범을 이

6) 아담 스미스의 두 저작, 『도덕감정론』과 『국부론』 사이의 (비)일관성에 대한 논쟁을 지칭하는 용어로 이에 대한 자세한 소개와 정리는 Montes(2003)를 참조.

끌어 내는 경향을 만들어내므로 타인의 승인을 원하는 사람들은 자신의 감정을 지배적인 규범에 맞추어 가게 되는 것이다.

아담 스미스의 공감 개념이 가진 이러한 여러 가지 의미 중 첫 번째 의미와 두 번째 의미가 사회적 선호이론과 사회적 선호이론 비판자들에 의해 각각 강조된 반면 버논 스미스의 경우 사회적 선호이론이나 상호성 이론과 달리 아담 스미스의 공감 개념에 대한 석텐의 해석과 유사하다. 그리고 첫 번째 의미를 강조하는 해석에는 매우 비판적인 반면, 두 번째 의미를 강조하는 입장에 대해서는 덜 비판적일 뿐 아니라 버논 스미스 자신도 이 입장에서 있다.

VII. 맺음 말

인간 행동의 사회성이라는 것은 사람들 사이의 상호작용을 통해 적절하다고 합의된 규범에 의해 행동하는 것이며 이것이 아담 스미스가 『도덕감정론』에서 말하는 사회성의 핵심이라는 버논 스미스의 해석은 사회적 선호이론이나 상호성 이론에 비해 『도덕감정론』의 핵심에 더 다가간 것은 분명해 보인다. 그러나 사회적 선호이론은 전통적인 공리주의적 신고전학과 이론의 부활이라는 비판은 평판 등 장기적 이기심에 근거하는 상호성이론에도 똑같이 적용되며, 버논 스미스 자신도 인정하듯이⁷⁾ 일부 버논 스미스의 저작에도 적용될 수 있는 비판으로 보여진다. 베커 식의 효용함수가 공리주의적 합리적 선택이론을 따르는 것이라고 한다면 평판 등 장기적 이기심에 기초한 의사결정 모델 역시 공리주의적 합리적 선택이론이 아니라고 보기 어려

7) 버논 스미스가 과거 자신의 상호성 이론의 한계를 스스로 지적한 사례는 여러 곳에서 찾을 수 있다.

“나는 이 단락이 내(와 나의 동료)가 예전에 신뢰게임에서의 행동을 설명하던 방식-즉, 『도덕감정론』의 영향이 아직 크지 않았던 때의 ‘상호성’에 의한 설명-의 변화를 구성하는 것임을 강조하고 싶다(Smith, 2012, p.4, 각주 9).”

“(신뢰게임 결과에 대한) 한 가지 대응은 이러한 협력이 상호성 즉 사실상의 교환에서 나오는 것이라고 주장하는 것이다. 그리고 이것은 (1998년 본인 논문 포함) 『도덕감정론』과 『국부론』의 연결을 이해하는 핵심이라고 주장하는 것이었다. 그러나 상호성은 단지 경험적 규칙성, A가 협력할 것을 제안하고 B가 받아들이는 교환을 위한 이름에 불과하다. 『도덕감정론』은 예측적 모델을 제공한다(Smith, 2013, p.8).”

을 것이다. 현재까지 나타난 버논 스미스의 논의를 종합하면, 평판 등 장기적 이기심에 기초하면서 사람들의 상호작용을 통해 만들어지는 규칙이나 규범을 통해 과도한 이기심이 제어되고 주변 사람들에게 적절하다고 승인받을 수 있을 정도의 행동이 나타난다는 주장을 통해 버논 스미스는 기존의 단순 상호성이론을 발전시키고 있다고 판단된다. 버논 스미스의 적극적인 아담 스미스 해석이 이와 같은 단순 상호성 이론의 세련화에 기여하는 것은 분명하나 단순 상호성 이론의 기본 틀을 유지하는 것이라면, 사회적 선호이론이 공리주의적 신고전학파의 부활이라는 버논 스미스의 비판은 자신에게 돌아오는 부메랑이 될 수 있다. 이러한 사회적 선호이론 비판이 자기 자신에게 돌아오는 것을 피하기 위해서는, 그리고 『도덕감정론』과 상호성이론의 차이를 인정한다면 상호성 이론과의 보다 근본적인 단절이 필요하며, 일회성 게임에서 나타나는 실험 결과를 해석하는 기존 자신의 관점이 부분적으로 수정되어야 할 것으로 보인다. 적절성 규범에 대한 강조에도 불구하고 일회성 게임에서 나타나는 사회적 행동의 설명은 결국 평판 혹은 장기적 이기심으로 귀결되고 있기 때문이다. 동료의식으로서의 공감의 의미를 강조한 석덴(Sugden, 2002)의 경우 『도덕감정론』에 담겨진 공감의 의미는 합리적 선택이론 틀에서 다루어지기 어려운 것이라고 명시적으로 주장한 바 있다. 여기서의 합리적 선택이론의 틀이란 베커 식의 효용함수 뿐 아니라 빈모아가 감정입의 개념을 다루는 방식까지 포함된다. 반베르그(Vanberg, 2008)의 경우도 도덕적 행위의 설명을 위해 합리적 선택이론에서 규칙 따르기(rule following) 이론으로의 전환이 필요하다고 주장한 바 있다. 버논 스미스 역시 사회적 행동이 적절성에 관한 규칙을 따른 행동임을 강조한다는 점에서는 석덴이나 반베르그와 차이가 없다. 그러나 구성적 합리성과 생태적 합리성⁸⁾에 관한 버논 스미스의 논의에서 나타나듯이 버논 스미스에 있어 합리적 선택(구성적 합리성)과 규칙따르기(생태적 합리성)는 인간의 경

8) 구성적 합리성과 생태적 합리성에 관한 버논 스미스의 정의는 다음과 같다.

“개인 또는 조직에 적용되는 구성적 합리성은 다른 대안적 행동보다 더 나은 것이라고 판단되는 행동을 분석하고 처방하기 위해 이성을 의도적으로 사용하는 것을 포함한다. 제도에 적용되면 구성주의는 더 바람직한 성과를 달성하기 위해 규칙 체계를 의도적으로 설계하는 것을 포함한다. 생태적 합리성은 관행, 규범, 진화된 제도적 규칙의 형태로 출현한 질서를 의미한다. 개인의 행동을 지배하는 것으로 이는 문화적, 생물학적 유산의 일부이며 인간의 상호작용에 의해 만들어지지만 인간에 의해 의식적으로 고안된 것은 아니다(Smith 2008, p.2).

제적 혹은 사회적 행위를 설명함에 있어 상호 보완적인 것이지 대립적인 것은 아니었다.⁹⁾

버논 스미스는 인간의 사회성 연구에 있어 『도덕감정론』으로 돌아가자고 주장했다. 버논 스미스가 강조한 사회성은 사회적 선호이론 논쟁에서 나타나는 이기심-이타심의 이분법적 구도가 『도덕감정론』의 논의를 계승하는데 적합하지 않다는 것을 보여주는데 분명한 기여를 하였다. 그러나 인간의 사회적이고 도덕적 행위는 합리적 선택이론의 틀 내에서 논할 수 없는 것인지, 그리고 『도덕감정론』에 등장하는 도덕적인 행위 주체는 주어진 규칙이나 규범을 따르는 존재일 뿐 도덕적 자율성을 지니지 못한 수동적 주체인가에 대해서 버논 스미스가 어떤 입장을 가지고 있는 것인지가 더 분명해져야 한다. 심리가 행동의 영역에 들어올 때 이는 필연적으로 사회적인 것이며 사회적 맥락이 결여된 개인적 의사결정이라는 것은 없다는 버논 스미스의 주장은 위의 질문에 예스라고 대답하고 있는 것처럼 보여진다. 하지만 버논 스미스가 사회적이고 도덕적인 행위를 설명함에 있어 합리적 선택이론을 완전히 버린 것인지는 분명하지 않다. 아담 스미스는 공리주의자가 아니며, 사회적 선호이론은 공리주의적 신고전학파의 부활이라는 버논 스미스의 주장이 명확한 설득력을 확보하기 위해서는 이에 대한 대답이 필요하다.

투고 일자: 2016. 5. 27. 심사 및 수정 일자: 2016. 6. 15. 게재 확정 일자: 2016. 6. 21.

◆ 참고문헌 ◆

- 이규상 (2012), “큰 그림을 찾아: 경제현상과 (실험)경제학자들이 하는 일에 대한 버논 스미스의 진화론적 해석” 『사회경제평론』, 39호, pp.7-46.
- Lee, K. S. (2012), “Looking for a ‘Big Picture’: Vernon Smith’s Evolutionary Interpretation of Economic Phenomena and What (Experimental) Economist Do”, *Social Economic Review*, 39, pp.7-46. (Written in Korean)

9) 버논 스미스 이론 체계에서 구성적 합리성과 생태적 합리성이 갖는 의미와 역할은 이규상(2012)에 상세히 서술되어 있음. 구성적 합리성과 생태적 합리성 사이의 관계를 바라보는 버논 스미스와 반베르그의 시각차에 대해서는 황재홍(2012)을 참조.

- 황재홍 (2012), “두 가지 합리성과 자유주의: 하이에크와 버논 스미스”, 『한국경제학보』 19권 2호, pp.245-259.
- Hwang, J. H. (2012), “Two Forms of Rationality and Liberalism: Hayek and Vernon Smith”, *Korean Journal of Economics*, 19(2), pp.245-259. (Written in Korean)
- Ashraf, N., Camerer, C. and G. Lowenstein, (2005), “Adam Smith, Behavioral Economist”, *Journal of Economic Perspectives*, 19:3, pp.131-145.
- Becker, G. (1976), “Altruism, Egoism and Genetic Fitness: Economics and Sociobiology”, *Journal of Economic Literature*, 14:3, pp.817-826.
- Binmore, K. (1998), *Game Theory and Social Contract, Vol. 2: Just Playing*, MIT Press.
- _____ (2006), “Why do People Cooperate?,” *Politics Philosophy Economics*, 5:1, pp.81-96.
- _____ (2010), “Social Norms or Social Preference?,” *Mind and Society*, 9, pp.139-158.
- Burnham, T. C. (2013), “Toward a Neo-Darwinian Synthesis of Neoclassical and Behavioral Economics”, *Journal of Economic Behavior and Organization*, 90S, pp.S113-S127.
- Burnham, T. C. and Johnson, D. D. P. (2005), “The Biological and Evolutionary Logic of Human Cooperation”, *Analyse & Kritik*, 27:1, pp.113-135.
- Clark, H. (2009), “Adam Smith and Neo-Darwinian Debate over Sympathy, Strong Reciprocity and Reputation Effect”, *Journal of Scottish Philosophy*, 7:1, pp.47-64.
- Fehr, E. and Fischbacher, U. (2002), “Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives”, *Economic Journal*, 112, pp.C1-C33.
- Fehr, E. and Schmidt, K. M. (2006), “The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories”, in Kolm, S. and Ythier, J. M. (ed.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Elsevier.
- Gale, J., Binmore, K. and Samuelson, L. (1995), “Learning To Be Imperfect: The Ultimatum Game”, *Games and Economic*

- Behavior*, 8, pp.56-90.
- Gintis, H., Bowles, S., Boyd, R., and E. Fehr, (2005), *Moral Sentiment and Material Interests: The Foundations of Cooperation in Economic Life*, MIT Press.
- Hanley, R. (2010), "On Vernon's Smith", George Mason University Workshop Manuscript.
- Hoffman, E., McCabe, K., and Smith, V. (2008), "Reciprocity in Ultimatum and Dictator Games: An Introduction", in *Handbook of Experimental Economics Results*, Vol 1, edited by Plott, C. & Smith, V., North Holland.
- Khalil, E. (2001), "Adam Smith and Three Theories of Altruism", *Louvain Economic Review*, 67:4, pp.421-435.
- Konow, J. (2012), "Adam Smith and the Modern Science of Ethics", *Economics and Philosophy*, 28:3, pp.333-362.
- Martin, M. (1990), "Utility and Morality: Adam Smith's Critique of Hume", *Human Studies*, XVI:2, pp.107-120.
- Montes, L. (2003), "Das Adam Smith Problem: Its Origin, The Current Debate, and One Implication for Our Understanding of Sympathy", *Journal of History of Economic Thought*, 25:1, pp.63-90.
- Paganelli, M. (2009), "Smithian Answers to Some Puzzling Results in the Experimental Literature", in *Elgar Companion to Adam Smith*, Young, J. ed. EE.
- _____ (2011), "The Same Face of the Two Smiths: Adam Smith and Vernon Smith", *Journal of Economic Behavior & Organization*, 78:3, pp.246-255.
- Price, M. E. (2008), "The Resurrection of Group Selection as a Theory of Human Cooperation", *Social Justice Research*, 21, pp.228-240.
- Price, M. E., Cosmides, L. and Tooby, J. (2002), "Punitive Sentiment as an Anti-Free Rider Psychological Device", *Evolution and Human Behavior*, 23, pp.203-231.
- Smith, A. (1976), *The Theory of Moral Sentiments*, Raphael, D. & Macfie, A. (eds), Oxford: Clarendon Press.
- Smith, V. (1998), "The Two Faces of Adam Smith", *Southern Journal*

- of Economics*, 65:1, pp. 1-19.
- _____ (2008), *Rationality in Economics: Constructivist and Ecological Forms*, Cambridge University Press.
- _____ (2010), "What Would Adam Smith Think?", *Journal of Economic Behavior & Organization*, 73, pp.83-86.
- _____ (2012), "Adam Smith on Humanomic Behavior", *Journal of Behavioral Finance & Economics*, 2:1, pp.1-20.
- _____ (2013), "Adam Smith: From Propriety and Sentiments to Property and Wealth", *Forum for Social Economics*, 42:4, pp.283-297.
- _____ (2015), "Adam Smith: Homo Socialis, Yes; Social Preference, No; Reciprocity Was to Be Explained", *Review of Behavioral Economics*, 2:1-2 pp.183-193.
- Smith, V. and Wilson, B. (2014), "Fair and Impartial Spectators in Experimental Economics Behavior", *Review of Behavioral Economics*, 1:1, pp.1-26.
- Sugden, R. (2002), "Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow Feeling", *Economics and Philosophy*, 18:1, pp.63-87.
- Tooby, J., Cosmides, L., and Price, M. E. (2006), "Cognitive Adaptations for n-person Exchange: The Evolutionary Roots of Organizational Behavior", *Managerial and Decision Economics*, 27:2-3, pp.103-129.
- Vanberg, V. (2008), "On the Economics of Moral Preferences", *American Journal of Economics and Sociology*, 67:4, pp.605-628.
- Witztum, A. and Young, J. (2013), "Utilitarianism and the Role of Utility in Adam Smith", *European Journal of History of Economic Thought*, 20:4, pp.572-602.

Social Preference Controversy and Vernon Smith's Interpretation of the Theory of Moral Sentiments

Jaehong Hwang*

Abstract

The purpose of this paper is to consider the relationship between modern theories of social preference/behaviour and Adam Smith's works. Social preference theorists and the critic of social preference theories respectively insist that their theories succeed the view of the theory of moral sentiments. However, there are some fundamental differences between their theories and the theory of moral sentiments. This paper discusses what causes these different interpretations on Adam Smith and what is Adam Smith's core idea on sociality. It also evaluates the Vernon Smith's interpretation of the theory of moral sentiments.

KRF Classification : B030101

Key Words : social preference, sympathy, Adam Smith, Vernon Smith

* Associate Professor, Department of Economics, Mokwon University,
e-mail: zootjhong@hanmail.net