과학철학 10 (2007) pp.45~73

진화론적 이타주의의 개념적 난점과 윤리학적 함축*†

정 상 모‡

진화론적 이타주의(altruism)는 자신의 적응도 희생을 통해 다른 개체(들)의 적응도를 상승시키는 행동 성향으로 정의된다. 이 정의에는 다음과 같은 역설이 숨어 있다: 이타주의는 비적응적이면서 적응적인 행동 성향이다. 이 역설의 해결을 위해 제시된 대표적인 이론들이 혈연선택론, 호혜 이타주의 이론, 그리고 집단선택론이다. 이 논문에서 필자는 먼저 각 역설의 해법에 나오는 이타주의 개념들을 분석하여, i) 그 의미를 명료히 하고, ii) 의미론적 순환이라는 하나의 공통적인 난점을 드러내고, iii) 그러한 분석의 결과가 갖는 윤리학적 함축을 살펴본다. 그 과정에서 이타주의의 일상적 의미와 생물학적 의미를 비교 분석하고, 일상적 또는 윤리적 이타주의의 생물학적 근거를 모색해 보는 작업도 병행한다.

【주요어】 생물학적 이타주의, 의미론적 순환, 개체선택론, 다-수준선택론, 도덕성

* 접수완료: 2007.11. 3 / 심사 및 수정완료: 2007. 12.25
† 이 논문은 2005년도 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(과제번호: A00140). 심사과정에서 유용한 지적을 해주신 심사자들께 감사드린다.
‡ 신라대 철학과 교수

1. 들어가는 말

진화론에서 이타주의는 한 개체의 적응도 희생을 통해 다른 개체(들)의 적응도 상승을 돕는 행동 성향으로 정의된다. '적응도 희생'은 그 개체의 생존과 재생산에서 손해를 본다는 뜻, 즉 '비적응적'이란 뜻이 된다. 그러면서도 그 형질은 어떻게든 자연에 의해 선택되어야, 즉 '적응적'이어야 한다. 그렇지 않으면 자연에 이타주의가 존재할 수 없기 때문이다. 따라서 이타주의는 다음과 같은 역설로 재정의 될 수 있다. 이타주의는 비적응적이면서 적응적인 행동 성향이다. 일찍이 다윈에게 심각한 위협이 되었던 이 역설에 대해 다윈 이래 여러 해법들이 제시되었다. 역설 해소의 주된 방법은 뒤에 나오는 "적응"의 외연을 개체가 아닌 유전자나 집단으로 바꾸든지, 앞에 나오는 "비적응"의 내포를 수단적 의미로 바꾸는 것이다. 이들 방법에 해당하는 대표적인 이론들이 혈연선택론, 집단선택론, 그리고 호혜 이타주의 이론이다. 앞의 두 이론은 이타주의가 개체로 볼 때는 부적응적이지만 개체가 가진 유전자나 개체가 속한 집단의 입장에서는 적응적이라고 보는 데 비해, 셋째 것은 개체의 행동이 겉으로는 이타적이지만 실제로는 최적의 적응도를 가져다주는 이기적 행동이라고 해석한다.

필자가 보기에 이타주의의 역설과 그 해법들에는 원래 일상 언어였던 '이타주의'가 생물학적 개념으로 재정의 되면서 생긴 의미론적 문제가 숨어 있다. 이 논문에서 필자는 역설의 해법으로 제시된 대표적인 이론들에 대한 분석을 통해 각 이타주의의 의미를 명료히 하고 숨겨진 의미론적 순환 문제를 드러내고자. 또한 그 결과의 윤리적 함축도 살펴본다. 그 과정에서 이타주의의 일상적 의미와 생물학적 의미를 비교 분석하고, 일상적 내지 도덕적 이타주의의 생물학적 근거를 모색해 보는 작업도 병행한다.

2. 일상적 이타주의와 진화론적 이타주의

'이타주의'(altruism)는 '타인'이라는 뜻의 라틴어 'alter'에서 유래하는 용어로 실증주의 창시자 오귀스트 콩트가 고안했다고 한다. 그는 우리가 다른 사람들의 관심 또는 "인간성의 더 큰 선"(greater good of humanity)에 봉사해야 하는 도덕적 의무를 지고 있다고 믿었다.[1] 그래서 그는 인간 행위의 목적과 동기는 자기 자신의 선(善)이어야 한다는 교설 또는 그러한 식으로 행동하려는 성향을 의미했던 '이기주의'(egoism)에 반대되는 개념이 필요했던 것이다. 동시대 영국인이었던 허버트 스펜서와 존 스튜어트 밀 등은 그 용어를 단지 다른 사람들을 위한다는 소극적인 단계를 넘어서 일반적인 사회 복지를 위한 개념으로 확장해 사용했다. 이처럼 일상적 내지 윤리적 이타주의는 타인의 선이나 이익의 추구를 의무와 올바름의 기준으로 삼는다. 그런데 이타적 행동은 행위자 자신에게도 만족감, 자기 존중, 의미의 충족 등과 같은 내적 이익을 가져다주기도 한다. 그런 이유로 이타적 행동도 결국은 이기적 행동이라고 주장할 수도 있다. 그러나 그렇게 되면 이기와 이타의 일상언어적 구분은 무의미해진다. 모든 의식적·자발적 행위가 다 이기적 행위로 되기 때문이다. 따라서 이기와 이타의 구분은 만족감 같은 내면의 이익보다는 물질적 이익과 같은 외재적 이익 관계를 기준으로 하는 것이 옳다.

곤충에서 사향소나 인간에 이르기까지 다양한 이타적 동물들이 관찰된 것은 오래 전의 일이었으나, 그것이 본격적으로 문제시 된 것은 다윈(C. Darwin)의 진화론이 성립하고서부터이다. 흡혈박쥐는 먹이 획득 실패로 죽을지도 모르는 동료에게 자신이 먹은 피를 나누어 준다. 새들은 동료가 둥지를 보호하고 새끼들을 먹이는 데 도움을 준다. 포식자를 발견한 긴꼬

1) 『실증주의 문답집』(*Catechisme Positiviste*)에서 꽁뜨는 그 이유를 다음과 같이 제시한다: "사회적 관점은 개인주의에 의존하는 권리 개념을 용인할 수 없다. 우리는 모든 종류의 의무 아래서 태어난다. 우리의 선조들에게, 우리의 후손들에게, 그리고 우리의 동시대인들에게 말이다. 탄생 이후 이들 의무들은 증가하거나 누적된다."(*Wikipedia*에서 재인용)

리원숭이는 죽음의 위험을 무릅쓰고 동료들에게 경고음을 낸다. 개미나 벌과 같은 진사회성 곤충의 불임 일꾼 개체들은 스스로의 자손을 생산하는 것도 포기하고 여왕과 그 새끼들을 돌보는 데 전적으로 헌신한다. 다윈의 진화론적 원리에 따르면 이기적 행동 성향을 가진 개체들만이 선택되고 진화하게 되어 있다. 따라서 이타적 행동은 심각한 반증 사례였다. 집단선택론은 다윈 자신이 이 문제의 해결을 위해 제시한 이론이었다.

생물의 선택과 진화를 적응도로 정의하는 (진화)생물학에서는 한 유기체가 자신의 적응도를 희생하여 다른 유기체의 적응도를 높이는 는 방식으로 행동할 경우 이타적이라고 규정한다. 적응도는 기대자손수로 측정된다(정상모 2007 참조). 한 유기체가 이타적으로 행동할 경우, 그는 자신의 기대자손수를 감소시키는 대신 다른 유기체의 그것을 증가시킨다.[2] 이처럼 생물학적 이타주의는 단지 적응도에 입각해서만 정의되기 때문에 일상적인 의미의 이타주의와 몇 가지 점에서 중요한 차이가 난다. 우선 윤리적 내지 일상적 이타주의는 주로 인간에게만 적용되는 '원인론적'(etiological) 개념, 즉 그 정의에 의도적 원인이 들어 있는 개념이다(Rosenberg 1992, 20). 그래서 사회성 곤충의 헌신적 행동은 이타적이라기보다는 본능적 행동으로 간주된다. 그에 비해 생물학적 이타주의는 동기나 의도와는 무관하고 단지 적응도에 미치는 영향만으로 판정된다.[3] 따라서 설사 의도나 동기가 이타적이라고 해도 적응도에 반영이 되지 않는다면, 그 행동은 이타적이 아니다. 이와 관련된 또 다른 차이로, 일상적 이타주의가 절대적 개념인데 비해 진화론적 이타주의는 상대적 개념이다. 전자는 의도나 동기와 같은 행위자

2) 이타주의 정의의 예: "한 동물이 자신의 이익을 희생하여 자신의 후손이 아니라 그 종의 다른 성원들의 이익을 증대시키는 방식으로 하는 행동"(Hamilton 1964), "다른 개체의 기대 번식 성공률을 증대시키고 행위자 자신의 그것은 감소시키는 행동"(Kitcher 1998), "자기 자신의 적응도를 희생하여 타 개체의 그것을 높이는 행동"(Rosenberg 1992, Sober and Wilson 1998)

3) 그래서 해밀턴은 이타적 행동을 유발하는 유전자 G를 가정하고 다음과 같이 말했다. "적자생존의 원리에도 불구하고 유전자 G가 전파될지를 결정하는 궁극적인 척도는 행위가 그 행위자에게 이로움을 주느냐 하는 점이 아니라, 그 행위가 유전자 G에게 이로움을 주느냐 하는 점이다, 이것은 행위의 평균적인 결과가 유전자 전체에서 유전자 G의 비중을 높여주는 경우이다."(Hamilton 1963, 354-355)

의 "내재적 특성"(intrinsic property)인 반면, 후자는 둘 이상의 개체들 간의 상대적 적응도의 의해 결정되는 외재적 특성(관계)이기 때문이다(Sober and Wilson 1998, 460). 이런 점들을 기초로 해서 일상적 이타주의와 진화론적 이타주의를 대비하면 다음과 같다:

<표1> 일상적 이타주의와 진화론적 이타주의[4)]

	일상적 이타주의	진화론적 이타주의
정의	개체 자기비용 희생 + 타개체(들) 이익 증가	개체 적응도 희생 + 타개체 적응도 증가
적용 대상	(주로) 인간	모든 생물
동인	마음(의도, 욕구)	본능 또는 유전자(유전 형질)
속성	타인 배려적 욕구	적응도 계산적 욕구
척도	행위자의 심적 욕구 상태	행위가 산출하는 적응도 분포
행동 결과	개체의 외재적 자기 비용 소모 (=자기 이익 감소)	개체 자기 적응도 감소 = 재생산율 감소 + 자손들의 미래 이익 감소
	타개체(들) 외재적 이익 증가	타개체(들) 적응도 증가 = 타개체 재생산율 증가 + 타개체 자손들 미래 이익 증가

3. 역설의 해법들 분석

다윈주의는 생물의 진화가 서로 경쟁하는 개체군 내에서 적응도 상에 우월적 지위를 가진 개체들이 자연에 의해 선택되는 메커니즘을 통해 이루어진다는 간단한 원리를 중심으로 한다. 이 원리의 핵심은 선택이 개체 단위(수준)에서 이루어지며, 선택의 기준인 개체의 적응도가 자손생산력을 뜻하며, 그리고 자연은 자신의 적응도를 높이는 행동 성향을 보이는 개체, 즉 이기적 개체를 선호한다는 점이다. 그래서 생물학적 이타주의 논제의 다음

4) 선행 연구(정상모 2004, 95)의 초기 버전 참조.

과 같이 표현된다. “자연은 어떻게 개체 차원에서는 정의상 도태될 수밖에 없는 이타주의를 선택할 수 있는가?” 아래 소절에서 이 역설에 대한 대표적인 해법으로 제시된 혈연선택론, 호혜 이타주의 이론, 그리고 집단선택론에 나타난 이타주의 개념을 분석해 보기로 한다.[5]

1) 혈연 이타주의

혈연선택(kin selection)이란 자신의 직접적인 재생산 기회를 희생하고 대신 유전자의 일부를 공유한 혈연의 자손증식을 도움으로써 간접적으로 자신을 재생산하는 진화 방식을 말한다. 주지하듯이, 할대인이 제시한 혈연선택이라는 아이디어를[6] 하나의 획기적인 설명 모델로 확립한 이는 해밀턴이다(Hamilton 1964). 그는 혈연선택이 성립하려면 형질의 유전과 진화를 개체가 아니라 유전자의 관점에서 바라보아야 한다는 사실을 간파하고 ‘포괄적응도’(inclusive fitness)라는 새로운 개념을 고안했다. 포괄적응도는 개체의 적응도에다 그것과 유사한 유전자를 갖고 있는 혈연들 각각의 적응도를 모두 합한 것으로, 관련 형질을 유발하는 유전자 복제본의 총량을 나타낸다. 따라서 혈연선택론은 적응도를 혈연집단 차원에서 계산한다는 점에서 기존의 개체 중심의 선택론과 구분된다.

혈연집단에서는 개체적응도보다 포괄적응도가 더 높은 경우가 있을 수 있다. 이 경우 해당 개체의 최선의 진화적 생존 전략은 개체적응도가 아니라 포괄적응도를 최대화하는 방향으로 행동하는 것이 된다. 즉 자신의 개체적응도가 포괄적응도보다 높을 경우에는 통상적인 이기적 개체로 행동하지만, 포괄적응도가 개체적응도를 초과할 경우에는 이타적으로 행동하는 것이 최선이다(Hamilton 1964; 2001). 이타주의의 진화를 예측하는 데 사용

5) 이타주의 문제와 그 해법에 관한 개괄적 논의는 최근의 여러 자료에서 참고 할 수 있다. 특히 최종덕(2004)과 장대익(2006)은 그것에 관한 훌륭한 정리를 담고 있다.

6) “한 개체의 후손과 가까운 혈연들의 생존에 유리한 한, 이타적 행동들은 일종의 다윈적 적응이다.” 그러한 경우 적어도 혈연에 대한 이타적 행동들은 그 행동의 주체에게는 해롭거나 심지어 치명적일 수도 있지만, 자연 선택에 유리하다(Haldane 1932, 131; Rosenberg 1992에서 재인용).

되는 "해밀턴 규칙"(Hamilton's Rule)은 가장 간단한 모델에서 rb >c로 표현된다. 여기서 b는 이타적 행동으로 인한 혈연의 적응도 이익분(자손증가분), r은 이타주의자와 혈연간의 유전적 근연도, c는 이타적 행동에 따르는 이타주의자의 적응도 손실분(자손감소분)을 나타낸다. 이 규칙은 이타적 행동으로 입는 행위자 자신의 적응도 손실분보다 포괄적응도의 혈연 요소가 더 클 때 이타주의가 선택됨을 의미한다.[7] 혈연선택론은 이타주의가 개체 또는 개체적응도로는 볼 때는 부적응적이지만 개체가 가진 유전자 또는 포괄적응도의 입장에서는 적응적이라고 간주함으로써 이타주의의 역설을 해소하고 있다.

20세기 생물학계의 획기적 발견의 하나로 간주되는 혈연선택론은 오늘날 이론적으로는 물론 경험적으로도 널리 인정되고 있다. 이 이론은 동물들이 비혈연보다 혈연들에게 더 이타적일 것이라는 점과 이타성의 정도가 클수록 혈연관계가 더 가까울 것이라는 예측에 잘 들어맞았다. 여러 종의 새, 다람쥐, 일본원숭이 등 많은 입증 사례들도 관찰되었다. 특히 사회적 곤충들에서 관찰되는 반수성(haploidy) 유전시스템은 극적인 예이다. 암컷들은 자신의 자손보다 자매들과 더 유전적으로 가깝다. 그래서 자신의 새끼를 낳는 대신 여왕의 재생산을 도와 자매의 수를 늘이는 것이 형질 유전에 더 유리하다.

혈연선택론에서 이타주의는 과연 어떤 의미를 가질까? 혈연집단에서 어떤 개체가 "이타적으로 행동한다"는 것은 "자신의 비용을 들여 혈연의 이익을 증진시킨다"는 말이고, 더 정확하게는 "자신의 포괄적응도를 극대화하는 식으로 행동한다"는 말이다. 한편 후자는 "자신과 동일한 형질을 가진 후손들의 번성에 가장 유리한 방식으로 행동한다"는 말이다. 결국 혈연선택론에서 말하는 이타주의란 실제로는 '혈연집단내'라는 특정 환경에서

---

7) 이타주의자들과 이기주의자들의 평균적인 포괄적응도는 각각 다음과 같다.

$$w(A) = x - c + rb \qquad w(S) = x$$

여기서 $x - c + rb > x$ (= $rb > c$)일 때, 즉 이타적 행동이 행위자 자신에게 주는 손실보다 포괄적응도의 혈연 요소가 더 클 때, 이타성은 선택된다(Hamilton 1964; 2001, Sober 1993).

다윈 원리에 충실한 개체가 택하는 적응 방식, 즉 이기적 행동 방식을 말한다. 따라서 "A가 이타적이다"는 말은 "혈연집단의 성원으로서 A는 이기적이다"는 말과 같다. 이타주의가 적응도로써 정의되는 한, 이는 당연한 결론이다. 문제는 이 과정에서 '이타성'의 뜻이 '자신과 동일한 형질의 진화(후손들의 수)를 극대화하려는 행동 성향'으로 바뀌고, '자신과 동일한 형질' 속에 '자신과 동일한 형질의 진화를 극대화 하려는 형질', 즉 '이타적 형질'도 있어야 한다는 점이다.[8)]

그 이유는 다음과 같다. 예컨대 자신과 동일한 형질이 파란 눈의 형질이라고 해 보자. 그러면 이타적 개체들을 통해 선택되는 형질은 파란 눈의 형질과 그것의 진화를 극대화 하려는 형질의 두 가지이어야 한다. 따라서 이타성의 진화라는 관점에서 보면 개체들간의 근연도는 파란 눈의 형질과 그것의 진화를 극대화 하려는 형질, 즉 이타적 형질의 합으로 측정되어야 하며, 개체들간의 '동일성'(유전적 근연도)에는 이타적 형질의 요소도 포함되어야 하는 것이다. 그렇기 때문에 '자신과 동일한 형질의 진화를 극대화 하려는 행동 성향'으로 정의되는 '이타성' 개념은 그 내포에 다시 '이타성'이 존재하는 중첩(nested) 관계를 갖게 된다. 이타성이 이타성에 의존하는 의미론적 순환이 발생한다는 말이다. 따라서 혈연이타주의 이론에서 이타성의 의미는 불완전한(열린) 상태로 남는다.

2) 호혜 이타주의

혈연 이타주의는 오직 혈연관계에서만 진화할 수 있다는 한계를 갖는다. 비혈연관계에 있는 개체들에서 진화할 수 있는 새로운 이타주의 설명 모델로 제시된 것이 호혜 이타주의(reciprocal altruism) 이론이다. 기본 아이디어는 간단하다: 한 유기체의 이타적 행동에 대해 상대로부터 최소한 그 이상의 보답이 장차 돌아올 것이라고 기대된다면, 그는 이타적으로 행하는

8) 혈연 이타성에서 혈연성은 단지 이타성 관련 유전자의 공유 정도를 확인하는 손쉬운 징표(indicator)일 뿐이라는 점에 유의하자(Hamilton 1975, Sober and Wilson 1998).

것이 적응적으로 유리하고, 따라서 그 행동 방식은 선택된다. 호혜 이타주의는 혈연관계일 필요도 또 동종의 일원일 필요도 없는 개체들간에 성립하는 이타성이다. 이 이론을 처음 제시한 트리버스(Trivers 1971)는 게임이론을 이용해서 상호협동적 행동을 위한 수학적 설명 모델을 만들었다. 잘 알려져 있듯이, 단판 죄수의 딜레마 상황에서는 타자로부터 제공되는 이익은 받고 그들에게 이익을 주지는 않는 것 즉, 이기적으로 행동하는 것이 최선의 전략이기 때문에, 협동은 일어날 수 없다. 그러나 반복된 딜레마 상황에서는 사정이 달라진다. 최종 예측이 불가능할 경우 장기적으로는 서로가 협동적일 때 개체에 돌아가는 손익배당이 최적화될 수 있다. 나의 가려운 등을 내가 직접 긁는 것보다는 믿음직한 동료의 등을 긁어주고 그 보상으로 그가 나의 등을 긁게 하는 것이 더 이익이다. 동료가 예뻐서가 아니라 자신의 등을 긁는 것이 동료의 등을 긁는 것보다 더 힘들기 때문이다.

트리버스의 모델은 액슬럿을 통해 본격적인 진화론적 게임 모델로 발전하는데, 그는 반복된 죄수의 딜레마와 같은 진화적 상황에서 한 유기체가 취할 수 있는 최적의 전략이 이른바 '팃포탯'(tit-for-tat: TFT)전략임을 입증했다. TFT전략이 "진화론적으로 안정적인 전략"(ESS: evolutionarily stable strategy)임을 이론적으로 밝힌 것이다.[9] 이 전략은 게임의 첫판에서는 상대와 협력하고, 다음 판부터는 상대가 그 전 판에서 한 대로 따라하는 것이다. 래포포트는 이 모델을 다수가 참여하는 복합적 상황으로 확대하여 TFT전략이 그런 상황에서도 ESS가 될 수 있음을 보여주었다(Rapoport 1991).

그런데 TFT전략은 다음과 같은 조건 아래서만 성립한다. 첫째, 게임의 횟수가 두 선수에게 알려져서는 안 된다. 횟수를 알 경우 최종 게임은 단판 죄수의 딜레마 상황과 같아서 두 선수 모두 비협조적(이기적)일 것이고, 그 계산은 앞 게임으로 연쇄적으로 파급되어 협동이 불가능하게 된다. 둘

9) 그가 고안한 진화론적 모델 게임의 결과에 따르면, TFT빈도가 1에 접근하는 집단에 있어서는 어떤 다른 변칙적 전략도 그것을 대신하지 못했으며(Axelrod and Hamilton 1981), 모든 게임에서 TFT는 지배적 전략이 될 때까지 확장되었고, 또 일단 확립되면 어떤 다른 전략도 그것을 이겨낼 수 없었다고 한다(Axelrod 1984).

째, 장기적 이익배당이 단기적 이익배당보다 더 많아야 한다. TFT가 현재의 무임승차 기회를 희생하여 미래의 이익을 얻는 전략이기 때문이다. 셋째, 선수들이 서로를 알아보고 앞 게임을 기억할 수 있어야 한다. 게임마다 정확한 손익계산이 필수이기 때문이다(Rosenberg 1992). 이런 조건이 충족되면 게임은 지속되고, 그 어느 하나라도 충족되지 않으면 신속히 붕괴된다. 게임 참여의 동기가 자기이익의 최적화이기 때문이다. 지속 기간에 비례하여 참여자 전체 집단은 번성한다.

호혜성이 협동적 이타주의 확산의 중요한 메커니즘임을 보여주는 경험적 증거들은 많다. 작은 집단을 이루어 사는 흡혈박쥐들은 흡혈 실패가 흔하고 이틀 이상 굶으면 죽는다. 상호협력은 생존에 필수불가결하다. 피를 나눈 시기가 최근인 짝들간에 피나누기 가능성이 더 높다는 관찰도 좋은 입증 사례이다(R. Wilson 2003). 열대 산호초에 사는 작은 물고기들은 큰 물고기들의 입이며 아가미에 있는 찌꺼기와 기생충을 먹이로 하고 산다. 청소와 먹이 제공을 통한 공생관계이다. 그런데 청소 중에 큰 물고기가 적의 공격을 당하면 큰 물고기는 바로 대응을 하지 않고 청소물고기가 입 밖으로 안전하게 나갈 때까지 기다린다. 그곳에 와서 장차 더 많은 청소서비스를 받을 수 있다는 기대 이익 때문에 큰 물고기들은 얼마간의 위험을 감수하고서라도 청소부들을 돌보는 것이다(Trivers 1985). 인간의 사회의 다양한 협동도 좋은 예이다.

일반적인 구분법 의하면 자아와 타아 모두에게 이익이면 협동, 자아에게 손해 타자에게는 이익이면 이타, 타자에게 손해 자신에게 이익이면 이기, 자타 모두에게 손해면 원한이나 앙심(spite)이다 (Wilson & Dugatkin 1992). 따라서 호혜 이타주의는 협동(cooperation)에 속한다. 더 정확하게 그것은 합리적 또는 지연된 이기주의이고, "자기-이익을 위한 경제적 교류"이다(같은 곳, 32). 지금의 협동이 장차 더 큰 이익을 얻으려는 욕구에서 나오기 때문이다. 따라서 호혜 이타주의 이론에서 "이타적으로 행동한다"는 말은 "미래의 더 큰 이익을 위해 상대에게 이익을 주는 식으로 행동한다"는 말이고, 더 정확하게는 "(호혜적 집단에서) 자신의 형질을 남기는 데 가장 유리한 방식으로 행동한다"는 말이다. 이 말은 다시 "(호혜적 집단

에서) 자신의 유전 형질을 가진 후손들의 번성에 가장 유리한 방식으로 행동한다"는 말이 된다. 결국 여기서 말하는 이타주의란 '호혜 집단(또는 관계) 내'라는 특수한 환경에서의 다윈 원리에 충실한 개체의 적응 방식, 즉 이기적 행동 성향을 말한다. 따라서 "A가 이타적이다"는 말은 "호혜집단의 성원으로서 A는 이기적이다"는 말과 같다.

이 결론 또한 대단할 것 없다. 문제는 여기서도 의미론적 순환이 발생한다는 점이다. 호혜적 협동에서 진화하는 이타성은 지금의 작은 이익을 자신과 마찬가지로 이타적인 형질을 가진 상대에게 주고 나중에 더 큰 이익으로 돌려받는 방식의 행동 성향이다. 간단히 '이타성'은 '이타성을 가진 상대와는 협력하는 행동 성향'으로 정의된다. '이타성'의 의미가 '이타성'에 의존하는 것이다.

3) 이타주의 일반

이제 논의를 혈연관계도 아니고 호혜관계도 아닌 일반적인 개체들간의 이타주의로 확장해 보자. "지난 30여 년간 [진화생물학] 논의의 동력(engine)이 되어왔던 것이 집단선택이라면, 이타주의라는 주제는 그 연료였다"(Kerr and Godfery Smith 2002, 478)고 할 만큼, 이타주의 문제와 집단선택은 밀접한 관련이 있다. 이타주의의 역설이 집단 차원에서는 손쉽게 해소되는 것처럼 보이기 때문이다. 집단선택은 "선택이 한 계통 집단에 속하는 단위체인 둘 또는 그 이상의 구성원에게 작용하는 것"을 말한다(Wilson and Dugatkin 1992, 718) 고전적 집단선택론은 다윈이 제시한 것이다. 그는 이타적 성원들로 구성된 집단이 이기적 성원들로 구성된 집단보다 자연선택에 더 유리하기 때문에, 집단 전체 차원에서 이타적 형질이 진화해왔다고 생각했다. 사회성 곤충들은 불임 개체들이 자신들의 생식적 이익을 희생하여 집단내의 가임 성원들의 이익을 도모함으로써(Darwin 1859), 고등한 사회성 동물인 인간은 구성원들이 상호협력함으로써, 집단 전체의 적응도를 높이는 방식으로 이타적 형질을 진화시킨다고 보았다(Darwin 1871, 82-84). 개체가 아니라 그것이 속한 집단을 선택 단위로 본

것이다.

이후 오랫동안 이타주의는 다윈식 집단선택의 결과로 간주되다가, 60년대 초 조류학자 윈-에드워즈(Wynne-Edwards 1962)에 의해 재개된 논의는[10] 윌리엄스의 논증과 매이나드 스미스의 증명에 의해 결정타를 입고 폐기된다.[11] 그 후 90년대에 소버와 윌슨이 새롭게 들고 나온 집단선택론이 이른바 "형질(trait group)집단선택론" 또는 "다-수준(multi-level)선택론"이다(Sober 1993, 1998; Sober and Wilson 1998).

소버와 윌슨은 생물 개체군을 구성하는 많은 요소들은 각기 고유한 "선택압"(selective force)을 가지며, 그것들이 서로 충돌하거나 보완하는 역동적인 관계 속에서 최종 선택이 일어나기 때문에, 형질의 진화는 그러한 관점에서 분석해야 한다고 주장했다. 이타주의 문제 또한 그러한 관점에서만 적절히 해결될 수 있으며, 혈연선택론이나 호혜 이타주의 이론은 단지 제한된 의미의 이타주의만을 보여줄 뿐이라고 주장했다(1998, 330~31) 그들이 제시하는 이타주의 진화의 메커니즘은 다음과 같다. 우선 여러 "형질집단"들로 구성된 하나의 개체군이 있다. 형질집단은 "어떤 형질과 관련하여 서로서로의 적응도에 영향을 주는, 그러나 그 집단 밖의 적응도에는 영향을 주지 않는 일군의 개체들"이다(1998, 92). 한 형질집단 내에서는 당연히 이기적인 개체들이 점차 증가하여 결국에는 모든 집단에서 이타적 개체들

---

10) 그는 홍뇌조 개체군이 환경의 수용능력 이하로 그 수를 자가조절한다는 사실을 발견하고, 그것에 대해 이타주의적 설명을 제시했다. 즉 각 개체가 자신의 생식을 조절(=희생)함으로써 집단의 생존을 강화한다고 보았다. 이 설명은 그러나 그 전에 랙스(Lacks 1954)의 핀치새 부화 연구를 통해 이미 내려진 전혀 다른 해석에 비해 별다른 설득력을 얻지 못했다. 랙스는 핀치새가 양육능력 이하의 적은 수로 새끼를 낳고 양육하는 것은 집단의 이익을 위한 이타적 자기통제가 아니라 새끼의 생존율을 높이기 위한 이기적 전략의 일종으로 간주했다.

11) 윌리엄스 논증의 핵심은 사고의 경제 원리다: 이타적 현상이라고 규정되는 것들이 개체선택론과 집단선택론에 의해 동일하게 설명될 수 있고 높은 수준의 선택론보다 낮은 수준의 선택론이 더 근본적이라면, 집단선택론에 의한 설명은 번거롭고 낭비적이다. 그는 또 집단에 유리한 형질은 개체군 내에서 급속히 사라진다고 주장했다. 매이나드 스미스는 실험을 통해 이타적 개체들이 "내부로부터의 전복"(도킨스의 용어)에 의해 결국 사라지고 만다는 사실을 보여주었다(Maynard Smith 1964).

이 사라질 것이다. 이타주의는 오직 집단내 개체들간에 적응도 차이가 있을 때 발생하는 "집단-내 선택"과 동일한 개체군에 속하는 집단들 사이에 적응도 차이가 있을 때 발생하는 "집단-간 선택" 사이의 역동적인 선택압 충돌의 결과로 진화한다. 그 충돌에서 이타주의가 진화할 필요충분조건은 다음과 같다(Sober and Wilson 1998, 26, 29):

① 이타적 유형의 비율이 상이한 둘 이상의 집단이 집단군을 형성한다.
② 이타적 개체가 많은 집단은 적은 집단보다 더 적응적이다.
③ 각 집단은 서로 고립되어 있지만, 그 자손들은 주기적으로 서로 섞이거나 새로운 집단을 형성할 때 서로 경쟁한다.
④ 집단을 형성할 때 이타적 개체들은 유유상종한다.(이상 필요조건)
⑤ 집단간 적응도 차이(이타주의를 선호하는 힘)가 집단내 개체간 적응도 차이(이기주의를 선호하는 힘)를 상쇄할 정도로 충분히 강해야 한다.(충분조건)

이런 조건에서 상대적으로 많은 이타적 개체들을 포함한 집단의 번식률 증가로 인해 집단군 전체에서 볼 때 이타적 개체의 증가 비율이 이기적 개체의 그것을 앞지르게 되고, 번식 후 각 집단이 해체되고 다시 새로운 집단을 구성하는 과정을 여러 세대 거치면서 장기적으로 이타성은 진화한다.[12] 그래서 소버는 "이타적 형질이란 그것을 소유한 개체에게는 해롭지만 그것이 나타나는 집단에게는 유리한 형질"이라고 말한다(Sober 1993, 90).

다-수준선택론의 이론적 가능성은 상당 정도 확립되었다. 입증 사례도 풍부하다고 알려져 있다. 암컷편향성비, 기생자독성감소(reduced virulence in disease organisms, Sober & Wilson 1998, 35-50), 뇌충(brain worm)의

12) 예컨대 개체 수가 각각 100이며, 이타적 개체의 비율이 각각 20%, 80%인 두 무성생식 <집단1>과 <집단2>로 구성된 개체군을 가정하자. 기저 적응도 10, 타개체의 이익 5, 그리고 이타적 개체의 손실을 1이라고 하자. 그러면 <집단1>의 이타성 개체와 이기적 개체의 평균 적응도는 각각 9.96과 11.01이 되고, <집단2>의 그것들을 각각 12.99와 14.04가 된다. 결과적으로 개체군 전체에서 애초 50.0%였던 이타적 개체들의 비율이 다음 세대에서는 51.6%로 증가한다. 집단의 이합집산이 주기적으로 계속된다면, 세대를 통해 이타주의는 진화하게 된다(Sober and Wilson 1998, 24-5).

자기희생(같은 책, 27-30) 등의 현상이 대표적이다.[13]

여기까지의 논의만 보면 다-수준선택론에서 이타주의는 앞서 봤던 두 이타주의와 사뭇 다르다. 즉, 가) 이타적 행동이 혈연관계도, 호혜관계도 아닌 성원들로 된 집단에서 일어나고, 나) 행위자 자신이든 자신의 유전자에게든 최대의 적응적 이익을 남기려는 목적이 직접적으로는 없으며, 다) 이타적 개체의 희생이 소속 집단의 다른 성원들 전체에 돌아갈 뿐 그 자신에게는 돌아오지 않는다. 이런 점에서 그것은 이타주의의 일반적 의미에 가장 충실하다. 소버와 윌슨에 따르면 그러한 이타적 형질이 진화하는 것은 독특한 집단 구조와 변화 형태로 인한 우연한 결과 때문이다. 여기서 "이타적으로 행동한다"는 말은 "자신의 비용을 들여 집단 내 타 개체들의 (적응적) 이익을 증진시키는 방식으로 행동한다"는 말이고, 더 정확하게는 거기다 덧붙여 "이익이 증진된 타 개체들이 많은 후손을 낳고 그 때문에 증가된 다른 이타적 개체들이 조건에 나타난 독특한 집단 구조와 변화 형태 덕분에 우연히 선택되는 식으로 행동한다"이다.

여기까지로 볼 때, 다-수준선택론에서의 이타주의는 분명 다윈적 이기주의가 아니다. 생물학적 이타주의를 이런 식으로 볼 수 있게 된 것은 적응도를 개체의 행동(성향)과 집단의 구조 및 변화 형태로 분리한 발상 덕이다. 그러나 이 참신한 발상에도 근본적인 문제가 있다. 필자는 그것을 집단 개념 그 자체의 임의성과 유유상종이라는 조건에서 찾는다. 절을 바꾸어 살펴보자.

4. 개체와 집단

다-수준선택론에서 설명된 이타적 형질의 진화는 개체선택론의 입장에

13) 비율이 높은 암컷, 독성이 약한 병균, 뇌충 등은 모두 집단내에서 불리한 적응도를 보인다. 그러나 그들이 있는 집단 자체는 그것을 상쇄하고 남는 높은 적응도를 보이기 때문에 이타적 형질은 진화한다. 소버와 윌슨(1998, ch.1)에 상세한 기전이 잘 나와 있다.

서도 단순명료하게 설명될 수 있다.[14] 그러나 소버와 윌슨에 따르면 개체선택론적 설명은 "평균화의 오류"(the fallacy of averaging)를 범한다(Sober and Wilson 1998, 31). 그것은 집단 내 개체간에 일어나는 선택과 집단간에 일어나는 선택이 엄연히 구분되어야 하는데도, 그런 구분 없이 개체의 적응도를 집단을 가로질러 평균화 해버리는, 그리하여 진화하는 모든 것을 하나로 뭉쳐버리는 그런 오류이다. 과연 이 주장은 옳은가?

다-수준선택론은 선택이 유기체를 규정하는 다양한 수준에서 상호역동적으로 진행된다고 보는데 비해, 개체선택론은 생존과 번식의 기본 단위인 개체 수준에서 선택이 일어날 수밖에 없으며 유전자나 집단과 같이 다분히 임의적인 요소는 개체의 환경으로 간주해야 한다고 본다. 다-수준선택론에서는 적응도를 유전자, 개체, 집단 등의 요소로 분해하는 데 비해, "광의의 개체주의"(broad-sense individualism)와 같은 최근의 세련된 개체주의 이론은 적응도를 자기요소와 환경요소로 분해한다.[15] 집단은 선택이 작용하는 단위가 아닌 개체의 '환경'의 일부로 간주되고, 선택은 빈도의존적 개체선택으로 통일된다. 다-수준선택론이 집단간 이타주의자 적응도가 집단내 이기주의자 적응도를 능가할 때 이타주의가 진화한다고 보는데 비해, 개체선택론은 이타주의자 적응도의 자기요소는 이기주의자의 그것보다 작지만 환경요소가 그 이상으로 이기주의의 그것보다 클 경우에 이타주의가 진화

14) 앞 절 각주 12에서 다-수준선택론의 수학적 모델로 주어진 <집단1>과 <집단2>의 경우를 개체주의적으로 설명해 보자. <집단1>의 이타적 유형 20은 각각 9.96의 후손, <집단2>의 이타적 유형 80은 각각 12.99의 후손을 갖는다. 평균하면 12.38이다. <집단1>의 이기적 유형 80은 각각 11.01의 후손을, <집단2>의 이기적 유형 20은 각각 14.04의 후손을 갖는다, 평균하면 11.62의 자손을 갖는다. 이타적 유형이 이기적 유형보다 평균적으로 더 높은 적응도를 갖는다. 그리하여 단순한 개체선택에 의해서 이타적 형질이 진화한다(Sober and Wilson 1998, 32).

15) 간단히 말해, "대부분의 진화는 번식집단내에 있는 개체들간의 이기적인 재생산 경쟁으로부터 발생한다"(Dugatkin and Reeve 1994, 107)는 견해로, 두갓킨과 리브 및 스티렐리(Sterelny 1996), 그리고 그들의 뒤를 이어 케르와 갓프레이-스미스(Kerr and Godfrey-Smith 2002)가 더욱 세련된 개체주의적 시각을 제시하고 있다. 케르와 갓프레이 스미스는 개체주의적 접근과 다-수준적 접근을 일반화하여 각각 "배경적(contextual) 접근"과 "집합적(collective) 접근"이라고도 부른다(2002, 481).

한다고 본다.

소버와 윌슨의 주장에도 불구하고 최근의 정교한 비교 연구의 결과들을 보면 개체주의와 다-수준주의간의 논쟁이 어느 한쪽의 일방적 승리로 끝나는 것은 가능하지 않아 보인다. 수학적 동등성 증명이 정교하게 이루어지면서 최근의 경향은 복수 시각의 존재를 인정하는 다원주의(pluralism) 쪽이다.[16] 소버와 윌슨도 양자의 수학적 동등성은 인정한다. 그러나 그들은 이타주의가 진화하는 인과적 과정을 '실제적'(real)으로 반영하는 것은 다-수준선택론뿐이라고 주장한다(Sober and Wilson 1998, 50~54). 개체주의 쪽의 반론도 만만치 않다. 반론의 핵심은 집단을 (개체에 비견되는) 적응도의 객관적 운반체 또는 상관자로 보기가 곤란하다는 점이다. 그 근거는 두 가지다. 하나는 스티렐리(Sterenly 1996, 583)가 제시한 것으로, 집단이 개체의 환경의 일부로 충분히 간주될 수 있지만 그 역은 아니라는 사실이 후자가 전자보다 더 근본적임을 함축한다는 주장이다. 다른 하나는 '집단' 개념 또는 그 실재 자체가 개체에 비해 대단히 임의적이라는 점이다. 두 번째가 첫 번째 근거의 보다 근본적인 원인이므로 아래에서 그것에 집중한다.

'집단'의 임의성의 한 근거는 집단의 작용이 집단내 개체들간의 상호작용에 의해 영향을 받는다는 사실이다(R. Wilson 2003). 예컨대 밀가루 벌레는 동족살해(cannibalism) 비율과 개체군 밀도에 의해 집단의 크기와 이주비율이 각각 영향을 받고, 애기장대의 잎 영역은 식재 밀도에 의해 영향을 받는다. 실험 조건에서 이들 각 쌍의 특성들은 서로 엉켜있어서 집단선택에 의해 둘이 함께 선택된다. 즉 높은 개체군 크기와 낮은 동족살해율 그리고 높은 이주율과 높은 개체군 밀도가 함께 말이다. 반대로 개체 수준에 속하면서도 집단적 또는 다-수준적 특성을 갖는 형질로부터 독립적이지 못한 특성들도 있다. 예컨대 일벌이 생의 대부분을 이타적으로 사는 데는 한 가지가 아니라 불임성, 여왕벌과의 근연도, 그리고 동료들로부터의 벌칙과

16) 두 모델간 동등성에 관해서는 Dugatkin and Reeve 1994(특히 p. 120의 그림1)를 참조할 수 있다. 거기서 그들은 양자간의 일대일 대응 관계(mapping)를 상세히 보여주고 있다. Reeve 2000, Kerr and Godfrey-Smith 2002도 좋은 참고가 된다.

같이 적어도 세 가지의 복합적인 요인이 있다(Maynard Smith 2002). 불임성과 유전적 근연도는 개체적 특성에 그리고 벌칙은 집단적 특성에 속한다.

진화생물학자 웨이드(M. J. Wade)가 말한 "간접유전효과"(indirect genetic effects)는 선택 단위로서의 집단 개념의 모호성을 보여주는 또 다른 증거이다. 간접유전효과란 "개체 자신의 유전자가 아니라 그것의 사회적 동반자들의 유전자 변화로부터 초래되는 표현형적 또는 적응도 변화"를 말한다(R. Wilson 2002, 547). 연구에 따르면 표현형 발현과 적응도가 환경변수의 값에 항상 의존하는데, 간접유전효과의 경우 선택에 미치는 영향이 극단적일 수 있을 정도로 그 영향력의 폭이 크다고 한다. 좋은 예가 식물의 키다. 식물의 키가 적응도에 미치는 영향은 부분적으로는 그것을 둘러싸고 있는 이웃 나무들의 키에 의해 결정된다. 그리고 그와 같은 개체의 유전적 환경 변화는 다시 그것의 표현형의 적응도 값을 신속히 변화시킬 수 있다. 주위 나무들의 키가 클수록 키 큰 형질의 나무가 선택된다. 마치 동종의 집단인 냥 말이다.

결론적으로, '집단내'와 '집단간'의 명확한 구분이 곤란할 정도로 생물집단은 존재론적으로 불안정한 실재이며, 그에 따라 집단 개념도 다분히 잠정적이고 임의적이다. 개체나 집단을 통해 선택되는 특성들은 물론 두 수준을 가로질러 선택되는 특성들을 산출하는 과정들 자체도 분리불가능하게 "뒤엉켜"(entwined)있어서,[17] 특정 선택론이 진화적 변화에 인과적으

17) 이 결론은 일벌, 미어캣, 그리고 거미에 관한 연구에서 메이나드 스미스가 언급한 다음 결론과 부분적으로 일치한다:

> 어떤 경우들에 동일한 과정이 두 개의 수학적으로 동등한 모델, 즉 상황적 모델과 집합적 모델로 기술 될 수 있다는 사실은 집단선택을 둘러싼 논쟁이 세상이 실제로 어떠한가에 대한 것이 아니라 단어의 사용에 관한 논쟁이었다는 점을 암시할지도 모른다. 확실히 많은 의미론적 문제가 있어 왔다. 집단선택이라는 명칭이 어떤 집단의 선택적 생존과 멸종에서 발생하는 집단-수준의 적응(예: 성)의 존재로부터 개별 유기체의 적응도가 부분적으로 동일한 종의 다른 성원들과의 상호작용 중에 결정되는 경우에 이르기까지 모든 것을 의미하는 데 사용되어 왔다(Maynard Smith 2002, 527).

로 책임이 있는 세세한 부분이나 과정을 “사실적으로”(realistically) 구분해 내는 것은 논리적으로 불가능하다. “그 부분에 해당하는 요소가 독립적이고 분리가능한 세계의 부분으로 미리 존재하지 않기 때문이다”(R. Wilson 2004, 545~4 6).

소버와 윌슨은 뇌충과 점액종 바이러스의 경우를 들어 이 주장을 반박할 수 있다. 두 경우 모두 숙주(개미와 토끼)의 각 개체를 형질 집단의 분명한 경계로 간주할 수 있기 때문이다. 그들은, 스스로를 희생하는 뇌충이 자기 집단내 타개체들에게만 이익을 주며, 그 덕에 잘 번식한 그 집단은 다음 단계(소의 간)에서 뇌충의 희생이 없어서 덜 번식한 다른 집단들과 경쟁하고, 다시 개미별로 아집단을 형성하는 방식으로 이타성이 진화하는, 분명한 다-수준선택 과정이 성립한다고 주장한다. 문제는 이 주장이 각 개미 속에 오직 한 집단의 뇌충만 있다는 가정하에서만 성립한다는 점이다. 그러나 이 가정은 입증되지 못했다. 그들 스스로 지적했듯이, 뇌충에 대한 연구는 70년대 위클러(W. Wickler)와 윌슨(D. S. Wilson) 이래 그 가정을 확인해 줄 후속 연구가 없다(Sober and Wilson 1998, 30). 한편 점액종 바이러스의 독성 약화가 다-수준선택의 결과라는 그들의 주장은 선택압이 집단내 요소와 집단간 요소의 두 지배적인 요소로 대별될 수 있다는 가정하에서만 성립한다. 그러나 로버트 윌슨의 최근 연구에 의하면, 그 가정 역시 입증되지 못했다. 점액종 바이러스의 진화는 바이러스의 독성, 숙주(=집단)의 특성, 전염 방법, 기후, 심지어 인간의 개입 등과 같은 다양한 요소에 의해 복합적으로 결정된다. 이 경우, 선택압은 집단내와 집단간으로 구분하기보다 개체요소와 환경요소로 구분하는 편이 더 적절하다. 집단간 요소는 다양한 비개체적 요소중의 하나일 뿐이기 때문이다. 심지어 번식이 매우 빨라, 감염된 숙주의 피부에 많은 감염 부위를 남기지만(그래서 전염성을 높인다) 내부 장기에는 손상을 주지 않는 그런 바이러스, 즉 독성은 높지만 치명성은 낮은 바이러스도 존재할 수 있다. 이런 경우 형질의 선택에서 집단요소가 하는 역할은 전무하다(R. Wilson 2004). 결론적으로, 집단이 개체에 비견하는 지배적인 선택압을 갖는다는 소버와 윌슨의 주장은 여전히 설득력이 부족하다.

5. 이타주의 일반 다시 보기

소버와 윌슨의 이론이 집단 개념의 임의성 문제로부터 자유롭지 못함을 보여주는 또 하나의 중요한 분석이 있다. 그 이론의 핵심인 집단의 선택압이 실제로는 없다는 길던후이스(Gildenhuys 2003)의 분석이다. 그에 따르면, 소버와 윌슨의 수학적 모델에는 "집단적응도"(group fitness)라는 말이 반복적으로 나오지만, 정작 그 말의 수학적 분석에서는 그 '값'이 없다. 실제로 각주 12의 모델에서 볼 수 있듯이, 적응도의 집단적 요소는 전적으로 각 아집단내의 이타적 및 이기적 개체군의 적응도로써 계산되고 있다. "집단간 선택압"을 형성할 <집단1>과 <집단2>의 적응도는 어디에도 없다. 그의 지적처럼 소버와 윌슨 모델에서 실제로 집단선택이 일어나는 곳은 개개의 개체군들(individual populations)간이다. 그것도 이타주의를 위한 선택이 아니라, 이타적 개채들을 포함하고 편향된(이타적 개체들의 유유상종) 아집단으로 주기적으로 나누어지는 개체군을 위한 선택이다. 이타적 개체를 포함하지 않거나 포함해도 편향된 아집단으로 주기적으로 나누어 지지 않는 개체군은 도태된다. 따라서 그들의 모델은 "적어도 개체군의 일부가 이타주이자일 때, 어떻게 편향된 아집단의 형성이라는 개체군-수준의 특성이 선택될 수 있는가" 하는 데 대한 모델, 즉 "특정 성향을 가진 개체군선택 모델"이지 "이타주의를 위한 집단선택 모델"은 아니다(같은 논문 46~7; 필자의 강조). 집단이 적응도의 운반체가 아니라는 말이다.

이 결론은 이타주의 진화를 개체와 집단간의 선택압 상충 탓보다는 이타적 개체가 처한 특별한 환경 탓으로 보는 것이 더 적절할 수 있음을 함축한다. 그 특별한 환경은 소버와 윌슨이 내세운 이타주의 진화의 조건들이다. 그리고 필자가 보기에 그 조건들 속에도 이 함축을 뒷받침 하는 요소가 숨겨져 있다. 앞 절에서도 말했듯이, 소버와 윌슨의 이타주의는 특별한 종류의 구조에 의해 성립한다. 한 집단군내에서, 높은 이타적 개체 비율을 가진 집단에서 더 많이 번성한 개체들이 그렇지 못한 집단들에서 번성한 개체들과 함께 섞여 주기적으로 아집단을 형성하고, 그때 이타적 개체

들은 유유상종 한다. 이 과정을 주기적으로 반복하는 그런 구조 말이다. 핵심은 유유상종이다. 이 과정에서 비이타적 개체들보다 더 많은 이타적 개체들이 이타적 행동으로부터 혜택을 입을 수 있게 된다. 이타적 개체의 희생이 다른 이타적 개체들에게 최대한 돌아가게 함으로써 집단군 전체를 통해 이타주의가 진화할 수 있는 구조인 것이다.

이타주의의 진화를 가능하게 하는 핵심 계기가 유유상종이라는 사실은 다-수준 선택론에서의 이타주의도 결국은 자신과 유사한 형질을 가진 다수의 다른 개체들을 통해서 선택되고 진화하는 혈연 이타주의와 질적으로 동일한 것이라는 사실을 함축한다. 혈연 이타주의이론에서 '혈연'은 단지 이타적 형질의 공유를 나타내는 징표일 뿐이고, 다-수준 선택론에서의 '유유(類類)'도 이타적 형질의 공유를 나타내는 징표이기 때문이다. 유전자의 관점에서 소버와 윌슨의 이타주의가 진화하는 구조를 나타내어 보면 그 점이 잘 드러난다. 그것은 "이타적 유전자들이 이타적 행동을 통해 다른 유기체들의 이타적 유기체의 복제를 유발하게끔 해주는 구조"이다(Gildenhuys 2003, 30). 소버와 윌슨의 이타주의를 혈연이나 호혜성과 무관한 이타주의로 간주할 수 있게 만들었던 요소는 선택이 이루어지는 구조의 내면을 들춰내자 무력화되어 버린다. 그 만큼 다른 이타주의와의 차별성도 약화된다.

이 결론에 따라, 그리고 앞절 말미에 이어서, 다-수준선택론에서 "이타적으로 행동한다"는 말은 이제 다음과 같이 개체주의적 표현으로 바뀔 수 있다. "집단군 전체에서 비이타적 개체들보다 더 많은 이타적 개체들이 이타적 행동으로부터 혜택을 입는 방식으로 집단내 타개체들을 위해 희생한다". 이 말은 "그 희생의 혜택이 최대한 많은 수의 다른 이타적 개체들에게 돌아가게 함으로써 집단군 전체를 통해 이타주의가 진화하게 행동한다"는 말과 같다. 선택되는 이타적 형질은 집단내에 이타적 형질을 가진 동료들이 일정 수 이상으로 많을 때에만 동료들을 위해 희생하는 형질이다. 그래서 앞의 말은 다시 "형질집단군 내에서 자신과 동일한(이타적) 형질(을 가진 후손들)의 번성에 가장 유리한 방식으로 행동한다"는 말이 된다. 결국 다-수준선택론에서 말하는 이타주의란 실제로는 '형질집단군내'라는 특수한 환경에서의 다윈 원리에 충실한 개체의 적응 방식에 불과하다. 즉 "A

가 이타적이다"는 말은 "형질집단군의 성원으로서 A는 다윈적으로 이기적이다"는 말과 같다.

물론 이 또한 적응도로 정의된 이타주의의 예정된 결론일 것이다. 문제는 혈연 이타주의 개념에서와 마찬가지로, 그 과정에서 '이타적'의 뜻이 '자신과 동일한 형질(을 가진 후손들)의 번성에 가장 유리한 방식으로 행동하는 성향'으로 바뀐다는 점이다. 혈연 이타주의 논의에서 봤듯이, '자신과 동일한 형질' 속에는 '그 형질의 번성에 가장 유리한 방식으로 행동하려는 형질' 즉 이타적 형질도 포함되어야 하기 때문에, 여기서도 '이타적'의 의미가 부분적으로 '이타적'에 의존하는 의미론적 순환이 발생한다.

지금까지 살펴본 세 종류의 생물학적 이타주의에 대한 논의를 한 문장으로 요약하면 다음과 같다. 생물학적 이타주의는 자신을 통한 직접적 재생산보다 동일한 형질을 가진 다른 개체들을 통한 간접적 재생산이 더 유리할 때, 그것을 실행하는 다윈적인 행동 성향 또는 적응 전략이다. 생물학에서는 이타주의를 이 이상으로 더 명료히 정의 할 수는 없어 보인다. '동일성' 속에 있는 순환성 내지 중첩성 때문이다. 이 한계는 '이타주의(이타성)'라는 용어가 다윈적 의미만 갖는 규약적 용어라면 별 문제가 되지 않을 것이다. 그러나 이타주의는 일상어 출신 용어이다. 그 의미의 핵심은 '자신의 손해를 감수하면서 남에게 도움을 준다'는 것이다. 이 핵심 의미가 생물학적 이타주의 정의에서 적절히 내포될 수 없다는 데에 생물학적 이타주의의 개념적 난점이 있다. 이 난점은 생물학, 특히 도덕을 생물학화 하겠다는 사회생물학의 이타주의 논의에서 중요한 오해의 원천으로 작용하고 있다.

6. 윤리학적 함축

루스는 도덕성을 포괄적응도(자신의 유전자 복제 총량)를 높이기 위해 유전자가 고안해 낸 환상이라고 했다(Ruse 1995). 유전자 결정론과 유전자 선택론적 입장에서 보면 이 말은 물론 맞을 것이다. 그러나 유전자 결정론

은 하나의 가설에 불과하며, 또한 삶과 행위의 직접적 단위는 유전자라기보다는 개체이다. 유전자의 지배적 결정권을 인정한다고 해도 각 개체가 타개체들과 어떤 관계에 있느냐에 따라 상이한 방식의 생존 전략이 성립할 수 있다.

생물학적 이타주의 개념이 의미론적 순환에서 벗어나지 못하는 것은 그것이 일상어를 차용했으면서도 "개체를 넘어서 적응도를 극대화 함"이라는 의미만을 가질 뿐, 일상적 이타주의 개념의 핵심 내포를 갖고 있지 않기 때문이다. 그래서 일상적인 이기·이타의 구분과 생물학적 이기·이타의 구분간에는 일대일 대응관계가 성립하지 않으며, 생물학적 이타주의의 개념에는 '배려'나 '희생'과 같은 용어의 도덕적 의미도 유의미하게 드러날 수도 없다. 생물학적 이타주의 논의가 윤리적 이타주의 논의에 제공할 수 있는 것은 생물학 차원의 존재론적 토대일 뿐이기 때문이다. 존재론적 토대는 그 상부에 존재론적 근거를 제공하지만, 상부는 인식론적, 특히 의미론적 자유를 누릴 수 있고, 따라서 상부는 토대에 대해 환원불가적 의미를 갖는다(Hartmann 1997, ch. 11). 생물학이 이타적주의가 생물의 이타행이 이러저러한 유전자적 근거에서 유래한다는 것은 알려줄 수 있지만, 어떤 유형의 행동 방식이 그러한 행동 성향에 속하는지, 왜 그런지 등에 관해서 알려줄 수 없다는 말이다.

혈연 이타주의는 타인을 혈연적 근연도에 비례해서 배려하라는 상식도덕성(common sense morality)에 생물학적 토대를 제공할 수 있다. 자식에 대한 부모의 헌신적 사랑에서 그 전형을 볼 수 있는 이타성이다. 상식도덕성은 비혈연적 대상은 적대시해도 된다는 함축을 갖고, 그것은 비혈연적 관계에 있는 대상들간의 적대성을 정당화 하는 원치 않은 결과를 낳는다. 이것은 이른바 상식도덕성의 자기파괴적(self defeating) 특성에 생물학적 근거를 제시해 준다(정상모 1998 참조). 혈연 이타주의 이론의 우호적 기여는 여기까지이다. 무엇이 그리고 어떻게 하는 것이 자식에 대한 부모의 사랑인지 또는 혈족에 대한 사랑인지에 관해서 생물학이 해줄 수 있는 말은 별로 없다.

호혜 이타주의는 사회계약적 도덕성의 생물학적 토대를 제공한다. 예컨

대 롤스의 정의론에 따르면 인간 사회의 기초를 이루는 기본 조건인 "원초적 입장"(original position)"은 다음 세 가지 요소로 되어 있다. 첫째, 사회의 성원들은 가능한 한 최대의 "주요 사회적 재화"(primary social goods)"가 자신에게 돌아오기를 원하는 "합리적이고 자기-본위적인"(self-interested) 사람들이다. 둘째, 그들은 미래에 이익이 보장되는 상호협력이 더 유리하다는 점을 알고 있다. 셋째, 그들 각자는 그 사회에서 자신들이 최종적으로 차지할 미래의 지위에 관해 "무지의 장막"에 가려 있다(Rawls 1971). 이 조건들은 거의 정확하게 진화게임에서 TFT전략이 성립하기 위한 조건들에 대응한다. 그러한 조건에서 그들이 이행하기로 상호합의한 내용("자유 우선성의 원칙", "차등의 원칙", "공정한 기회 균등의 원칙")도 호의에는 호의로 대한다는 TFT전략에 대응한다. 그러나 구체적으로 무엇을 어떻게 하는 것이 호혜적으로 도덕적인지에 관해 호혜 이타주의 이론이 알려주는 바는 별로 없다. 예컨대 거지에게 적선을 하는 것과 하지 않는 것중 어느 쪽이 도덕적으로 옳은가 하는 물음에 그 이론은 답을 주지 못한다.

윤리학에서 순수한 의미의 이타주의는 헌신이나 희생과 같이 타인에 대해 대가없는 배려를 요구하는 성자(聖子)적 도덕성에서 그 전형을 볼 수 있는 희소한 이타주의이다.[18] 이런 이타주의는 소버와 윌슨의 형질집단선택론에서 부분적으로 (형질집단내에서만 성립한다는 의미로) 그 생물학적 토대를 찾을 수 있다. 그 이론은 서로 피를 나눈 적도 없고 계약적 관계도 아닌 전혀 모르는 사람들을 위해 대가 없이 희생하는 이타성의 존재, 그 희생으로 사회 전체가 번성하는 사태, 그리고 그러한 번성 중에 이타적 형질을 가진 사람들도 함께 번성하는 사태 등을 잘 설명해 준다. 인간 사회가 상벌을 통해 이타적 행동을 장려하고 이기적 행동을 억제하는 것은 유유상종 조건에 대응한다. 이타주의가 매우 까다로운 조건하에서만 성립한다는 점을 통해 혈연이타성이나 호혜 이타성이상의 이타성이 인간 사회에

18) "이타주의자는 이면의 욕구(other-directed desire)를 가져서는 안 된다. '타인을 위한다는'이 욕구를 오직 비도구적 방식으로 가져야 한다. 타인의 선이 목적이어야 할 뿐, 어떤 이기적 만족을 위한 수단이어서는 안 된다"(Sober and Wilson 1998, 467).

서 매우 희귀하다는 점을 반영해 주기도 한다. 또한 일견 순수하게 희생적인 이타주의 실현으로 보이는 행동도 결과적으로는 어떤 형태로든 합당한 보상을 받는다는 예측도 제공해 준다. 그러나 무엇을 어떻게 하는 것이 그런 이타주의에 적합한 행동 유형인지에 관해서 이 이론이 제공해 주는 것은 물론 없다.

보편적 도덕성의 기본은 보편적 인간애이다. 그것은 본능적으로 자신에게 향하는 생물의 이기적 성향을 거스르라는 명령의 형식으로 존재한다. 예컨대 동서고금의 윤리설이 공통적으로 인정하는 보편적 도덕 원리인 황금률[19]은 나의 이기적 욕구를 포기하고 타인에 대한 인간애를 발휘할 것, 즉 남을 적어도 나만큼은 배려하라고 요구한다. "너 자신의 인격과 다른 모든 사람의 인격 가운데 있는 인간성을 언제나 동시에 목적으로서 사용하고 결코 수단으로서 사용하지 않도록 행위하라"는 칸트 윤리설과 "최대다수의 최대의 행복" 그리고 "한 사람에게는 한 몫"을 원리로 내세우는 공리주의는 황금률의 보편적 도덕성을 잘 반영하고 있다(Velasquez & Rostankowski 1984). 이런 보편적 도덕성의 생물학적 기초는 어디서 나올 수 있는가? 혈연에 대한 배타적 배려는 당연히 보편적 도덕성의 지위로서 부족하다. 호혜적 이타주의도 그 동기가 자기 이익의 극대화라는 점에서 보편적 도덕성에는 미흡하다. 그러나 다른 사람들 또한 나와 마찬가지로 자기 이익의 극대화를 추구한다는 사실을 인정하고 상대의 행동에 따라 조건부로 호의를 베풀기 때문에 제한적으로는 보편적 인간애가 반영되고 있다. '집단내'라는 단서를 단다면, 집단내 다른 성원을 위해 희생하는 형질집단선택론의 이타주의는 보편적 도덕성을 도리어 얼마쯤 지나친 듯하다. 따라서 기존의 생물학적 이타주의 이론에 그것에 딱 맞는 것은 없어 보인다. 호혜 이타주의 이론과 형질집단선택론의 중간쯤에서 찾을 수 있을지 모르겠다. 그것이 어떤 이론일 것인가는 다른 연구의 몫으로 남겨둔다.

---

19) 서양에서는 예수의 산상수훈에 나오는"그러므로 무엇이든지 남에게 대접을 받고자 하는 대로 너희도 남을 대접하라"(마태복음)에서 유래한다. 동양에도 동일한 의미의 공자 말 "己所不欲勿施於人"(논어, 안연편)이 보편적 도덕률인 황금률을 형성해왔다.

7. 나오는 말

일반적으로 동물의 경계음은 동족을 포식자로부터 보호하려는 이타적 행위로 해석해 왔다. 그러나 반대 해석의 여지도 있다. 최초로 경계음을 내는 개체는 대체로 민첩하고 강건한 개체이고 현명한 포식자는 그런 개체보다는 느리고 병약한 개체를 노리기 때문이다. 선행(善行)과 건강에 관한 최근의 연구에 따르면, 선행을 하는 사람이 그렇지 않은 사람 또는 선행을 받는 사람보다 면역력이 더 높고 수명도 더 길다고 한다. '이기'와 '이타' 개념이 혼란스러워지는 대목이다. 이런 결과들은 동물계에 존재한다는 이타적 그리고 이기적 행동의 실체 규명이 더 필요함을 보여준다. 규명 결과, 인간을 포함한 동물계에 사실상 순수한 이타주의가 없는 것으로 판명되면, 진화 생물학의 이타주의 논의는 윤리학에 유전자 차원의 존재론적 토대를 적절히 제공하는 셈이 된다. 그렇지 않은 경우에는, 적응도 중립적인 이타주의 형질의 존재를 인정해야 할지도 모른다.

적응도 중립적 이타주의 이론으로 대표적인 것은 피터 싱어(Singer 1981)의 "진화적 파급 효과(evolutionary spin-off)" 이론과 "문화적 진화론"이다.[20] 전자는 이타주의를 직접 진화하는 형질이 아니고 다른 지적·정서적 유전 형질에 의해 파생적으로 형성되는 행동 성향으로 간주하는 것이다. 문화적 진화론(Cavali-Sforza and Feldman 1981, Boyd and Richerson 1985)은 한 걸음 더 나아가, 한 사회가 이타주의에 부여하는 높은 가치와 그러한 가치에 대한 부단한 사회적 학습이나 교육과 같은 순전히 문화적 적응성을 통해 이타주의가 선택되고 진화될 수 있다고 본다.[21]

적응도 중립적 접근도 그 나름으로 중요한 의의를 갖지만, 지금은 적응

20) 모방을 통해 한사람의 두뇌에서 다른 사람의 두뇌로 복제되는 방식으로 진화한다는 도킨스의 "밈"(meme) 이론도 문화적 진화론의 일종이라고 볼 수 있다.

21) 그 좋은 예가 19세기 서구 사회의 출산율 감소 현상이다. 출산율의 적정 수준 이하로의 감소 현상은 생물학적 적응도와 유전 이론으로 설명되지 않는다. 그것이 다윈적 비효용(생물학적 적응도 감소)에도 불구하고 선택되는 것은 그것이 갖는 문화적 적응성 때문이다. 또 그것을 전달하는 것은 유전자가 아니라 학습이나 교육이다(Sober and Wilson, 1998)

주의와 비적응주의에 관한 논의가 아직 진행 중이고, 신종합의 대미를 장식할 진화발생생물학(Evo-Devo: evolutionary developmental biology)이 양자간의 새로운 종합의 가능성도 예고하는 시점이다. 이보디보가 '유전자 대 환경'(nature vs. nurture)이라는 대립적 구조를 해소함으로써 적응주의와 비적응주의간의 간격을 획기적으로 좁혀줄 것으로 기대될 뿐 아니라, 생물학적 진화와 문화적 진화의 구분도 허물 수 있을 것이라는 기대도 주고 있다. 이보디보와 함께, 진화생리학과 진화심리학 등과 같은 학문의 신속한 진보는 생물학이 윤리학에 존재론적 토대를 넘어서 인식론적·의미론적 기초도 제공할 여지를 갖고 있음도 암시한다. 그래서 지금으로서는 적응도 중립적 이타주의 이론으로 나아가는 것이 다소 성급한 전략으로 보인다.

참고 문헌

장대익 (2005), 「이보디보 관점에서 본 유전자, 선택, 그리고 마음」, 서울대학교, 박사학위논문.

정상모 (1998), 「파핏의 상식 도덕성」, 『철학』54, 한국철학회: 219-257.

_____ (2004), 「적응도의 늪」, 『과학철학』7권 1호: 91-108, 2004.

_____ (2006), 「이타주의 논의에서 적응도 개념의 다의성」, 『대동철학』37: 93-118.

_____ (2007), 「"평균화의 오류"의 오류」, 2007 과학철학회 하계발표회.

최종덕 (2004), 「생물학적 이타주의의 가능성」, 『철학연구』, 철학연구회 64: 179-198.

Axelrod, R.(1984), *The Evolution of Cooperation*. Basic Books, New York.

Axelrod, R and W. D. Hamilton (1981), "The Evolution of Cooperation," *Science 211*: 1390-96.

Boyd, R. & Richerson, P.(1985), *Culture and The Evolution Process*. University of Chicago Press, Chicago.

Cavali-Sforza, L and Feldman, M.(1981), *Cultural Transmission and Evolution: A Quantitative Approach.* Princeton University Press, Princeton.

Darwin, C.(1859): *The Origin of Species.*『종의 기원』, 홍성표 역, 홍신문화사, 2007.

_____ (1871), *The Descent of Man and Selection in Relation to Sex.* John Murray, London.

Dawkins, R.(1976), *Selfish Gene.*『이기적 유전자』, 홍영남 역, 을유문화사, 2000.

Dugatkin L. A. and H. K. Reeve (1994), "Behavioral Ecology and Levels of Selection: Dissolving the Group Selection Controversy." *Advanced in the Study of Behavior* 23 101-133.

Gildenhuys, P.(2003), "The Evolution of Altruism: The Sober/Wilson Model", *Philosophy of Science* 70: 27-48.

Hamilton, W. D.(1964), "The Genetical Evolution of Social Behavior I and II," *Journal of Theoretical Biology* 7: 1-15.

Hartmann, N.(1997), *Neue Wege der Ontologie.* 5th ed.,『존재론의 새로운 길』, 손동현 역, 서광사. 2000.

Kerr, B and P. Godfrey-Smith(2002), "Individualist and Multi-level Perspectives on Selection in Structured Populations," *Biology and Philosophy* 17: 477-517.

Lacks, D.(1954), *The Natural Regulation of Animal Numbers.* Oxford University Press, Oxford.

Maynard Smith, J.(1964), "Group Selection and Kin Selection," *Nature 201: 1145-47.*

_____ (2002), "Commentary on Kerr and Godfrey-Smith," *Biology and Philosophy* 17: 523-527.

Rawls, J.(1971), *Theory of Justice.*『정의론』. 황경식 역, 서광사(1990).

Reeve, H. K.(2000), "Book Review of Unto Other," *Evolution and Human*

Behavior 21: 65-72.

Rosenberg, A.(1992), "Altruism: Theoretical Contexts," in E. F. Keller and E. A.

Ruse M.(1995), *Evolutionary Naturalism: Selected Essays.* Routledge, London.

Singer, P.(1981), *The Expanding Circle.* Farrar, Straus, and Giroux, New York.

Sober, E.(1993), T*he Philosophy of Biology*. Oxford University Press, Oxford.

_____ (1998), "What is Evolutionary Altruism?" in D. Hull and M. Ruse(eds.), *The Philosophy of Biology.* Oxford University Press, Oxford, 459-478.

_____ (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Harvard University Press, Cambridge.

Sterelny, K. (1996), "The Return of Group", *Philosophy of Science* 63: 526-84.

Trivers, R. L.(1971), "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46: 35-57.

_____ (1985) *Social Evolution.* Benjamin Cummings.

Velasquez and Rostankowski (1984), *Ethics: Theory and Practice.* Prentice-Hall, Englewood Cliffs, NJ.

Williams, G. C.(1966), *Adaption and Natural Selection*. Princeton University Press, Princeton.

Wilson, D. S. and L. A. Dugatkin (1992), "Altruism: Contemporary Debates", in Keller and Lloyd 1992.

Wilson, R. A.(2002), "Pluralism, Entwinement, and the Levels of Selection," *Philosophy of Science* 70: 531-552, 2002.

_____ (2004), "Test Cases, Resolvability, and Group Selection: A Critical Examination of the Myxoma Case," *Philosophy of Science* 71:

380-401, 2004.

Wynne-Edwards, V. C.(1962), *Animal Dispersion in Relation to Social Behavior*. Oliver and Boyd, Edinburgh.

Conceptual Difficulties and Ethical Implications of Evolutionary Altruism

Sang Mo Jung

Evolutionary altruism is defined as the propensity of behavior to cost one own fitness for the purpose of benefiting others' fitness. There is a paradox in this definition such that "altruism is nonadaptive as well as adaptive propensity of behavior." Three main solutions of the paradox are three theories of kin selection, reciprocal altruism, and multi-level selection. In this paper, firstly I analyse these theories in order to make clear their concepts of altruism and reveal a semantical circularity shared among them. I also explicate ethical implications of the results of the analysis. In addition to the analysis, I compare ordinary and biological meanings of altruism and examine the possibility of biological grounding of ethical altruism.

[Key Words] biological altruism, semantical circularity, individual selection theory. multi-level selection theory, morality