

정치학과 실험*

안도경 | 서울대학교 정치외교학부

실험이란 자연적인 상황에서는 수행하기 어려운 대상이나 과정을 측정하기 위해서 또는 자연적으로 발생하는 자료로는 하기 어려운 인과적인 추론을 목적으로 하여 자연적 과정에 개입하는 것이다. 정치학은 오랫동안 경험적 연구와 규범적인 연구를 수행해 왔지만, 과학적 연구의 대표적 방법인 실험은 불가능하거나 부적절한 것으로 간주되어왔다. 기술의 발전, 정책적인 기회, 그리고 연구자들의 창의성을 바탕으로 하여 이러한 현실은 변화하고 있으며 이제 실험은 정치학의 핵심적인 연구방법 중의 하나로 정착하고 있다. 본 논문은 정치학 실험의 세 가지 유형인 실험실 실험, 현장 실험, 그리고 설문 실험을 소개하고 각 유형의 대표적인 연구들을 검토한다.

주제어: 정치학 방법론, 실험방법, 정치학과 실험

* 이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013S1A3A2055391).

실험이란 관찰 또는 인과적 추론을 목적으로 자연적 과정에 개입하는 것이다. 개입이 필요한 이유는 자연적으로 관찰되지 않는 현상을 관찰하거나, 자연적인 과정을 단축하거나 또는 자연적인 과정에 대한 관찰 자료로는 불가능한 인과적인 추론을 위해서이다. 정치학은 오랫동안 경험과학적 연구를 규범적인 연구 못지않은 주요 구성부분으로 포함해 왔지만 과학적 연구방법의 표준이라고 할 수 있는 실험 방법은 여전히 적합하지 않거나 또는 불가능하다고 간주되어 왔다. 이러한 현실은 최근에 변화를 겪고 있으며 이제 실험은 정치학의 핵심적인 연구방법의 하나로 정착하고 있다.

실험방법을 사용한 정치학적 연구로서 최근에 큰 반향을 일으킨 예로 중국의 인터넷 검열에 대한 현장실험을 보자. 인터넷을 통한 정치 정보의 유통이 표현의 자유를 확대하고 민주주의의 발전에 기여할 것이라는 희망은 과연 중국에도 적용될 것인가? ‘국경 없는 기자회’는 중국을 ‘네티즌들의 최대의 감옥’으로 묘사하고 있다.¹⁾ 반면 중국의 네티즌들은 꾸준히 검열의 장벽을 넘어 인터넷을 통한 정부 견제와 민주주의의 확대에 진전을 이루고 있다는 보고도 있다 (Human Rights Watch 2012). 중국정부는 어떤 내용의 인터넷 메시지들은 어떠한 방식으로 검열하고 있는가? 공산당과 지도자에 대한 비판은 허용되는가? 이러한 의문들을 자연적으로 관찰 가능한 자료에 근거하여 답하는 데는 한계가 있다. 관찰이 가능한 인터넷의 내용들은 이미 검열을 통과한 것이므로 검열에서 걸러진 메시지들의 내용을 추론할 수 없다. 인터넷 사용자들에 대한 탄압 보고가 유용한 정보이기는 하지만 어떤 내용이 얼마만큼 어떠한 방법으로 통제되는지에 대한 정확한 상을 그리지는 못한다.

King et al.(2014)은 현장실험을 통해서 중국의 인터넷 검열에

1) <https://en.rsf.org/china-china-12-03-2012,42077.html>

대한 매우 구체적인 상을 그리는데 성공하였다. 연구진은 중국의 여러 소셜미디어 사이트에 계정을 만들고, 그 사이트들에 상이한 내용의 메시지를 게재하여 어떠한 메시지들이 어떻게 검열되는지를 측정하였다. 연구진은 또한 중국의 인터넷 회사와의 계약을 통해 소셜미디어 사이트를 직접 만들어 중국 정부의 인터넷 검열을 참여관찰적인 방법으로 연구하여 위의 현장실험을 보완하였다. 이러한 연구의 결과 King et al.은 중국인터넷 상에서 중국정부, 지도자, 정책에 대한 비판은 거의 무제한적으로 허용되고 있지만 현실의 집단행동으로 이어질 가능성이 있는 메시지들은 체계적으로 검열되고 있다고 보고한다. 이와 같은 발견은 부정적인 일화들을 중심으로 한 국제인권기구의 보고와는 어느 정도 다른 현실을 보여준다. 그러나 여전히 중국의 인터넷이 아랍의 봄과 같은 민주화의 직접적인 매개체가 되기는 어려운 현실을 보여주기도 한다. 이 연구는 앞으로 중국에서의 표현의 자유를 논함에 있어서 바탕이 될 현실진단의 더 정확한 상을 제시하였다는 의미가 있다.

또 다른 실험연구의 예로 저개발국의 개발정책과 관련한 현장실험을 하나 살펴보자. 저개발국이 여성 참여의 제한으로 인한 개발정책적 한계를 경험하고 있는 것으로 보고되고 있는데 과연 여성의 참여는 개발정책과 그 성과에서 어떠한 차이를 낳을 것인가? 탈레반정권의 붕괴이후에도 아프가니스탄은 여전히 세계에서 가장 여성차별적인 사회들 중의 하나로 지속되고 있다. 이 나라에서 수행되는 마을 공동체 차원의 개발 프로그램들에서 여성의 참여를 확대하는 것은 여성의 경제적, 정치적 위상을 높이는데 기여할 수 있을 것인가?

자연적 관찰 자료에 입각하여 위의 문제를 다룬다고 가정해보자. 가상적으로 10개의 마을 공동체를 여성이 남성과 동등하게 마을의

공동사업 운영에 참여하는 5개의 곳과 그렇지 않은 5개의 곳으로 나누어 각 그룹에서 평균적으로 여성에 대한 인식, 여성의 교육 정도와 사회경제적 위상 등을 설문으로 측정하였다고 해보자. 각 지표마다 두 그룹의 평균치를 비교하여 여성참여의 효과를 측정할 수 있을 것인가? 이러한 연구는 원인과 결과의 관계가 분명하지 않은 내생성의 문제로 인하여 인과적 추론에 한계가 있다. 즉, 여성의 지위가 이미 높기 때문에 여성이 마을 사업에 동등하게 참여하는 것인지 아니면 동등한 참여의 결과로 여성의 지위가 높아진 것인지를 알 수 없는 것이다. 분명한 인과적 추론을 위해서는 여성의 참여가 보장되지 않고 있는 마을에서 여성이 참여하였을 경우, 그리고 여성의 참여가 보장되는 곳에서 여성이 참여하지 않았을 경우와 같은 ‘반사실적 사례들(counterfactuals)’에 대한 정보가 필요하다. 두 그룹이 여성 참여 이외의 조건에서 동일하다고 판단할 수 있을 때만이 여성 참여의 인과적 효과에 대한 측정이 가능한 것이다. 달리 말해 양 집단에서 동일하지 않은 것은 단지 여성의 참여정도만이 아니라 그에 영향을 미치는 여러 요인들일 수 있다. 즉, 공동체가 얼마나 경제적으로 발전해있는가, 여성의 역할에 대한 인식이 어떠한가, 외부 사회와의 교류정도가 활발한가, 등의 요인이 여성의 참여정도에 영향을 주고 동시에 직접, 간접으로 여성의 사회경제적 지위에 영향을 준 것일 수 있다. 이러한 경우, 여성의 참여가 여성의 사회경제적 지위에 영향을 주었는지, 주었다면 그 정도는 얼마 만큼인지 등에 대한 정확한 인과적인 판단을 내릴 수 없다. 여성에 대한 편견이 적고, 외부사회와 잘 연결되어 있으며 경제적으로 어느 정도 발전해 있는 공동체라면 여성의 마을 운영에 대한 참여 없이도 여성의 사회경제적 지위가 높을 수도 있는 것이다.

Beath et al.(2013)은 세계은행의 지원으로 아프가니스탄 정부

가 수행하고 있는 공동체 단위 개발프로그램에 무작위대조군시험(Randomized Controlled Trial, RCT)을 결합시켜 여성의 참여가 여성지위의 향상을 가져오는지 연구하였다. 구체적으로 500개 마을을 상호 비교 가능한 250개의 쌍으로 나누고 각 쌍에서 한 마을에 대해서는 여성의 참여를 조건으로 하는 마을 지원프로그램을 실시한 것이다. 프로그램 실행 2년 뒤에 13,000여명을 대상으로 설문을 통해 프로그램의 효과를 측정된 결과 연구진은 여성 참여가 여성의 사회, 경제 정치적 지위의 향상을 가져오기는 하였으나 가족 내에서의 여성의 발언권 확대나 여성에 대한 인식의 개선으로 이어지지는 않았음을 보고하고 있다.

위에서 소개한 두 연구는 현장실험 방법을 사용하여 정치학적으로 중요한 문제들에 대하여 기존의 관찰연구의 한계를 극복한 예이다. 물론 정치학적 실험연구에서 현장실험이 기본이라거나 또는 가장 보편적인 것은 아니다. 그 외에도 실험실 실험, 설문 실험 등의 방법이 있다.

정치학 실험의 목적은 정치현상의 정확한 측정, 변수들 간의 인과적 관계에 대한 추론, 인과메커니즘의 규명 등에 있어서 자연적 관찰데이터가 가지는 한계를 극복하는데 있다. 흔히 실험의 목적은 인과적인 추론에 있다고 여겨지지만 연구대상의 상태에 대한 정확한 측정을 위해서도 실험이 필요한 경우가 많다. 예를 들어 물리학자가 빛의 속도를 측정하기 위해 실험을 하는 것은 인과적인 이론을 정립하거나 가설을 검증하기 이전에 대상에 대한 측정 그 자체를 목적으로 하는 것이다. 앞서 소개한 King 등의 중국 인터넷 검열에 대한 실험은 메시지의 종류에 따른 검열의 유무를 확인하는 인과적 이론 검증의 요소를 포함하고 있기는 하지만 그 이전에 중국에서의 인터넷 검열은 어느 정도이며 어떠한 방식으로 행해지는가

라는 질문에 답하였다는 것에 더 큰 의미가 있다고 볼 수 있다. 물론, 자연적으로 주어지는 데이터의 한계를 극복하고 데이터의 생성 과정을 관찰함으로써 인과적인 추론을 행하고, 더 나아가서 인과적 기제를 재구성해 내는 것 또는 실험의 매우 중요한, 어쩌면 가장 중요한 역할이기도 하다. 실험이 필요한 것은 정치학적으로 중요한 여러 질문들이 관찰 자료를 바탕으로 대답하기 어렵거나 또는 관찰 자료를 바탕으로 주어진 대답을 좀 더 정교한 방식으로 검증할 필요가 있기 때문이다.

정치학 실험의 종류는 크게 랩(실험실) 실험, 현장 실험, 설문 실험 등으로 나눌 수 있다. 이 세 가지 종류의 실험 모두 정치학에서 상당한 역사를 가지고 있다. 특히 최근 수십 년 동안 실험방법이 정치학에서 확대된 것은 실험경제학의 영향을 받은 실험실 실험의 확대, 그리고 그를 뒷받침한 컴퓨터 및 전산 네트워킹 기술의 발전에 기인한다고 할 수 있다. 그리고 최근에는 현장실험의 방법의 사용이 급격하게 증대하고 있는데 이 역시 인터넷의 발달과 개발협력의 확대와 같은 기술적 조건과 연구 환경의 변화를 연구자들이 적극적으로 활용한 결과이다.

1. 랩(실험실) 실험

랩 실험 혹은 실험실 실험(laboratory experiment)은 피험자들을 특정한 장소에 모아서 연구자가 인위적으로 통제하는 조건하에서 피험자의 반응 또는 상호작용을 측정하는 것이다. 실험실 실험은

그 수행 방법에 있어서 실험의 이상에 가장 부응하는 실험의 형태라고 할 수 있다. 실험의 목적이 측정과 인과적 추론을 위해 자연적인 과정에 개입하는데 있다고 할 때, 실험실 실험은 그 여러 가지 문제점에도 불구하고 피험자의 특성, 피험자의 상호작용의 환경, 피험자가 접하는 정보의 내용 등에 대한 가장 높은 수준의 통제가 가능하다는 장점이 있다.

예를 들어 ‘죄수의 딜레마 게임이 반복되면 맞대응(Tit-for-Tat) 전략을 통한 협동이 가능하다’라고 하는 악셀로드의 주장(Axelrod 1984)을 실험을 통해서 검증한다고 해보자. 과연 사람들은 맞대응 전략을 사용할 것인가? 그를 통해서 협동이 유지될 것인가? 악셀로드의 이론, 그리고 그 이후의 반복게임에서의 협동의 가능성에 대한 이론들은 죄수의 딜레마 게임이 종결점이 불명확한 상태에서 반복된다는 것을 가정하고 있다(Fudenberg and Maskin 1986). 그렇다면 죄수의 딜레마 게임이 유한반복 될 경우, 예를 들어 20회를 반복하기로 하고 그러한 사실을 피험자들이 알고 있을 경우 협동은 발생하지 않을 것인가?

위와 같은 연구 질문들을 답하기 위한 실험실 실험을 어떻게 수행할 것인가? 죄수의 딜레마라고 하는 매우 간단한 게임에 대한 실험을 행하기 위해서도 만만치 않은 준비가 필요함을 독자들은 깨달을 수 있을 것이다. 먼저 누구를 피험자로 할 것인가? 일단 실험을 대학교 내에서 수행하기로 하고 대학생들을 충원하기로 하자. 그렇다면 어떠한 방식을 충원할 것인가? 대학교 포털사이트에 실험참가 공고문을 올릴 수도 있고 또는 특정한 수업에 참여하는 학생들에게 실험 참여를 권유하거나 또는 심지어 ‘강제’하는 방법도 생각할 수 있다. 실험참여자들에게 보상을 제공할 것인가? 제공한다면 어떠한 보상을 제공할 것인가? 예를 들어 기본 참가비 오천 원에 게임의

결과에 따른 추가적인 보상을 제공할 수 있다. 실험정치학에서는 이 제는 많이 사용되지 않는 방법이기도 하지만 학점 산출에서 추가적인 점수를 줄 수도 있다. 죄수의 딜레마 게임 자체는 어떻게 구성할 것인가? 한 번에 두 명만 실험실로 불러들여서는 충분한 관찰 자료를 모으는 시간이 너무 많이 걸릴 수 있기 때문에 열 명 또는 스무 명을 한 세션에 충원하는 것이 좋을 것이다. 그렇다면 죄수의 딜레마는 2인 게임이므로 어떻게 짝짓기를 할 것인가를 결정하여야 한다. 그리고 보상표도 만들어야 할 것이다. 예를 들어 두 명이 다 협동을 하면 70점, 둘 다 배신을 선택하면 30점, 한명이 협동을 하고 한명은 배반을 한다면 협동자는 0점 배반자는 100점을 주고, 이 점수는 실험 후에 합계하여 현금으로 전환시켜 지급할 수 있다.

위의 예는 죄수의 딜레마라고 하는 매우 간단한 실험을 수행하는 하나의 방식일 뿐이다. 언급되지 않은 요인들 중에 특히 중요한 것은 실험실 조건에서 피험자간의 상호작용의 환경이다. 현재 가장 보편적으로 사용되는 방식은 컴퓨터화된 상호작용으로서 마치 매우 간단한 컴퓨터게임을 연상하면 된다. 한 실험실에 앉아있는 피험자들이 컴퓨터프로그램을 통해서 무작위로 짝이 지워지고 익명의 상대방과의 게임에서 컴퓨터 모니터를 통해 자신의 선택을 입력하고 게임의 결과에 대한 피드백을 받는 것이다. 컴퓨터 사용의 보편화 이전에는 ‘종이와 연필(paper and pencil)’ 실험을 흔히 행하였는데 예를 들면 실험참여자가 개인별 결정작성지(decision sheet)에 자신의 선택을 기입하고 실험자가 그 결과를 모아서 짝지어진 상대방의 선택과 결부시켜 보상을 정하는 방식이다. ‘종이와 연필’을 통한 실험이 학술적인 가치가 떨어진다고 할 수는 없다. 다만 실험상황이 복잡해질수록 시간이 많이 걸리고 통제가 어려워질 수 있다는 문제점이 있다. 이에 비해 컴퓨터화된 실험은 실험실에서 이루어지거나

또는 인터넷을 이용하여 온라인으로 진행된다면 종이와 연필 실험에 비해서 훨씬 더 복잡한 형태의 실험을 효과적으로 구현할 수 있다.

그 이외에도 실험의 프레이밍을 어떻게 할 것인가의 문제가 있다. 추상적인 의사결정 실험이라고 하고 선택지를 A와 B로 제시할 것인가? 아니면 협동과 배반이라는 용어를 명시적으로 사용할 것인가? 이러한 문제에는 정답이 있는 것이 아니라 연구자의 목적과 필요에 따라서 적절하게 선택할 수 있지만 다만 현대의 실험 사회과학에서는 가치중립적인 용어를 사용하는 것이 보편적이고 특별한 이유가 있을 경우에는 현실적인 프레임을 한다.

실험실 실험은 피험자가 누구인가, 실험실이 어디에 설치되는가, 실험참가자들 사이의 상호작용이 어떠한 방식으로 이루어지는가와 같은 주요한 속성과 그 이외에도 실험의 프레이밍을 어떠한 방식으로 제공하는가, 실험참가자에 대한 보상이 어떻게 이루어지는가, 실험에 기만(deception)이 사용되는가와 같은 속성들에 의해서도 분류될 수 있다. 그 이외에도 실험실 실험의 유형에 대한 중요한 구분은 실험과 이론과의 관계, 특히 게임이론과의 관계이다. 실험실 실험 중에서도 이후에 살펴볼 설문실험에 근접하는 형태의 실험, 즉 집단 내 개인들의 상호작용의 요소가 적은 형태의 실험이 있는가 하면 다른 한편으로는 게임이론적인 모형에 입각하여 상호작용 환경이 설계되고 게임이론적인 균형분석을 통한 예측을 도출한 전제위에서 행해지는 실험도 있다. 후자에 가까운 실험 중에서도 상호작용의 상황의 복잡성, 또는 게임균형분석을 통한 예측의 한계 등으로 인하여 ‘이론검증’보다는 ‘탐색’에 가까운 형태의 실험도 있다.

정치학에서 실험실 실험을 통해서 다루어지는 주제는 매우 다양하며 하나의 주제에 대해서도 여기서 포괄적인 문헌검토를 하는 것

은 불가능하다. 이 논문에서는 두 가지 주제에 대한 실험실 실험 연구 문헌을 간단히 소개하기로 하겠다. 먼저 정치학에서 실험실 실험이 가장 활발하게 수행된 주제들 중의 하나는 집합행동과 협동의 문제이다. 집합행동의 문제는 정치학이 다루는 여러 가지 문제의 핵심에 위치한다(Ostrom 1998). 이 문제는 죄수의 딜레마, 공공재 공급게임, 공유자원 게임, 신뢰게임 등의 형태로 정치학자와 경제학자, 사회학자 심리학자 등의 가장 활발한 공동연구의 주제이며, 물리학, 생물학, 컴퓨터공학자들의 실험 및 비실험 연구와 연결되어 사회과학과 자연과학을 포괄하는 가장 폭넓은 영역에서 학문적 논의의 장이 마련되어 있는 주제이기도 하다.²⁾ 이 주제에 대한 실험연구에는 초기부터 정치학자들이 활발하게 참여하여 중요한 기여를 하였다(예를 들면 Orbell and Dawes 1991; 1993; Ostrom, Walker and Gardner 1992 등). 초기 연구가 선호의 이질성과 자발적 협동의 가능성을 보여주는데 초점을 맞추었다면 최근에는 다양한 제도적 요인들이 집합행동에 미치는 영향, 그리고 제도 그 자체의 선택 또는 진화와 관련된 문제로 초점이 이동하고 있다(Dal Bo 2013).

게임이론적인 모형을 활용한 실험실 실험에 적합한 또 하나의 정치학적 주제는 투표와 선거이다. 이 주제는 수많은 게임모델이 존재하여 실험의 디자인과 이론적인 분석이 가능하고 기존의 이론적 경험적 문헌과의 연결도 용이하다는 장점이 있다. 투표에 대한 초기의 연구는 투표방식에 따른 균형분석을 실험으로 검증하거나 게임이론적 예측이 분명하지 않은 경우의 결과를 탐색하는데 초점을 두었다(McKelvey and Ordeshook 1983; 1990). 이후 투표 참여의 집합행동 문제에 대한 실험이 활발하게 행해졌으며(Grosser and

2) 이 주제에 대한 실험실 실험에 대한 문헌검토는 Ostrom(1998)과 Chaudhuri(2011)를 참조할 것.

Schram 2006; 2010) 최근에는 투표 및 투표참여에 있어서의 사회학적 심리학적 요인들을 게임이론적 합리적 선택적인 모델과 결합하는 실험연구도 활발하게 진행되고 있다(Ahn, Huckfeldt and Ryan 2013; Ahn et al. 2014). 그 외에도 투표 결과에 대한 숙의(deliberation)의 영향(Dickson, Hafer, and Landa 2008)과 정치정보와 부동산 투표자의 전략적 선택에 대한 실험연구(Battaglini, Morton, and Palfrey 2010)등이 있다.

위의 주제들에 대한 실험실 실험은 주로 게임이론의 모형에 입각한 실험이며, 게임이론 자체의 균형분석에 입각한 경험적인 검증은 아니라 할지라도 최소한 게임이론적인 용어와 관점에서 실험의 환경을 구성하고 있다. 이러한 실험을 나누는 한 가지 방식은 그 실험적 상호작용 환경에 대한 이론적인 분석이 얼마나 가능한가이다. 이론적 분석은 강한 의미의 예측이라고 할 수 있고, 연구자가 실험의 결과가 게임이론적 균형 분석과 다를 것으로 이미 예측하고 있는 경우도 실험결과를 이해하는 기준(benchmark)의 역할을 할 수 있다. 게임이론적 균형 분석이 하나의 예측으로 귀결되는 것이 아니라 다수의 균형을 낳은 경우에는 그 중 어느 균형이 현실적으로 실현되는지를 실험으로 검증할 수 있다. 마지막으로 실험 환경이 게임이론적인 용어로 구성되기는 하였으나, 균형분석이 불가능한 경우도 있는데 이러한 경우 실험은 이론검증보다는 탐색적인(exploratory) 의미를 지닌다(Smith 1982).

정치학 실험실 실험이 모두 게임이론적 모형에 입각한 것은 아니다. 집단 의사결정이 아닌 개인적인 의사결정과 반응에 대한 실험 연구도 특히 정치심리학 분야와 정치커뮤니케이션 분야에서 오랫동안 수행되어 왔다. 예를 들어 Ansolabehere and Iyengar(1995)의 네거티브 캠페인과 투표참여에 대한 실험은 피험자에게 제공되는

정치 메시지들을 체계적으로 변화시켜 캠페인 전략의 투표참여에 대한 영향을 연구한다. 여기서 결과는 실제 행동이 아니라 실험 후 설문에서 투표의향을 물어보는 방식으로 측정되었다. 이 연구는 자연적 환경에서 유권자들이 노출되는 메시지의 특성과 투표참여 결정에 대한 효과를 관찰하기 어려운 문제를 실험실 실험의 방식으로 극복하고자 한 것이다. 물론 특정 선거의 맥락에서 설문을 통하여 유권자에게 자신이 접한 캠페인 메시지의 특성과 투표참여 여부를 묻고 그 상관관계를 연구하는 것이 가능하기는 하다. 그러나 자기보고의 문제, 투표참여에 대한 알려진 과대보고의 문제, 또한 메시지에 대한 선택적 노출 등 인과적 추론을 어렵게 하는 여러 가지 요인이 있다. 이 문제를 실험실 실험으로 다루는 것 또한 여러 가지 한계가 있기는 하지만 설문조사나 현장실험 등 다른 방식의 연구와 결합하여 캠페인 방식과 투표참여의 관계를 규명하는데 기여할 수 있는 것이다. 이러한 유형의 실험실 실험은 정치적 성향에 따른 정보에 대한 선택적 노출(Iyengar and Hahn 2009), 유권자의 후보자에 대한 평가 방식(Lodge, McGraw, and Stroh 1989; Lodge and Steenbergen 1995) 등 매우 다양한 정치학적 질문들을 다룬다.

실험실 실험의 문제점으로는 흔히 다음과 같은 점들이 제기되는데 되는데 공히 실험 결과의 일반화 가능성, 또는 외적타당성(external validity)과 관계된 것들이다(Levitt and List 2006). 첫째, 실험의 인위적인 환경이 외부세계의 조건과 다르다는 점이다. 인위적인 환경의 문제는 실험실에서의 상호작용 상황이 얼마나 외부세계의 상황의 핵심을 구현하고 있는가의 문제뿐만 아니라 자신의 결정이 관찰의 대상이 된다는 사실 자체가 실험실에서의 행위와 선택에 영향을 줄 수도 있다는 것을 의미한다. 둘째, 실험참가자들이 흔히 대학생들이 경우가 많고 뿐만 아니라 자기선택(self-

selection)에 의해서 실험에 참여하는 경우가 대부분이기 때문에 특정한 속성을 지닌, 예를 들어 더 경쟁적이라거나 금전적인 보상에 대한 관심이 많다거나하는 피험자들이 과대 대표될 가능성이 있다는 점이다.

외적인 타당성의 문제는 어느 정도는 대부분의 경험적인 정치학 연구가 직면하는 문제이다. 대표적으로는 사례연구 또는 적은 수의 사례에 입각한 비교연구의 한계를 들 수 있다. 심지어 많은 수의 사람들을 대상으로 한 설문연구조차도 과연 조사 대상이 아닌 인구집단에 대해 일반화 될 수 있는가에 대한 문제가 제기되기도 한다. 무작위 대표표본에 대한 설문조사는 인구집단 전체에 대한 추론의 근거가 되기는 하지만 그럼에도 불구하고 과연 다른 인구집단, 또는 다른 시기의 인구집단에 대해서도 연구의 결과가 적용되는가의 문제는 여전히 남는다. 이에 더하여 최근에는 대표표본의 추출 자체가 어려워지는 문제가 더해지기도 하였다. 물론 일반적인 사례연구, 비교연구 또는 조사연구는 그 연구의 직접적인 대상 자체가, 일반화의 가능성을 떠나서 자체적으로 관심의 대상인 경우가 많다는 점에서 외적타당성의 제한에도 불구하고 연구의 의의가 완전히 훼손되지는 않을 것이다.

실험과 비실험연구에서 외적 타당성의 문제는 정도의 문제이지 실험실 실험에만 고유한 문제는 아닌 것이며 또한 외적타당성 문제에 대한 일반적인 처방의 연장선상에서 그 해결책을 찾을 수 있을 것이다. 즉 일반화의 한계가 의심되기는 하지만 그 자체로서 흥미롭고 중요한 기여가 있을 수 있는 연구의 경우는 추가적인 실험, 다른 유형의 실험, 그리고 비실험 연구방법을 통한 검증 등을 통해서 검증하거나 보충하는 것이 타당할 것이다.

2. 현장 실험

실험실 실험은 연구자에 의한 환경통제의 가능성이 높아서 내적 타당성을 지닌 인과적 추론이 가능하다는 장점이 있지만, 앞서 말한 바와 같이 실험실 환경의 인위성, 관찰의 대상이 된다는 사실이 행동에 영향을 미칠 가능성, 피험자의 대표성 등과 같은 한계를 지니고 있다. 최근에는 이러한 실험실 실험의 단점을 극복하면서도 반사 실험 환경을 만들어 관찰과 인과적 추론을 가능케 하여 실험연구의 장점을 극대화하고자 하는 시도가 활발하게 이루어지고 있는데, 그것이 바로 현장실험(field experiment)이다. 현장실험도 여러 가지 유형이 있기 때문에 한마디로 간단하게 정의하기는 어렵다. 그러나 이상적인 현장실험이란 연구자가 연구대상으로 삼는 자연적인 환경 속에서 또한 그 속에서 자연적으로 행동하는 행위자들을 대상으로 하여 대상들이 자신이 연구를 위한 관찰의 대상이 된다는 사실을 의식하지 아니하는 상태에서 행해지는 실험이다. 즉 현장실험은 실험실의 인위적인 환경이 아니라 관찰대상의 자연적인 활동 상황, 또는 상호작용의 상황에 연구자가 개입하여 그 결과를 관찰하는 것이다. 자연적인 상황에 개입은 그 방식과 정도에 있어서 다양할 수 있고 따라서 현장실험도 그 ‘자연성’의 정도에 따라 여러 가지로 나눌 수 있다(이에 대해서는 Harrison and List 2004 참조).

본 논문의 서두에 예를 든 두 가지의 실험은 모두 현장실험에 해당하지만 그 자연성의 정도에 있어서 약간의 차이가 있다. 즉 중국의 인터넷 검열에 대한 King et al.의 현장실험은 연구자의 개입의 대상이 되는 상황에 존재하는 행위자들이 연구자의 존재 자체를 완전히 모르고 따라서 자신들의 일상적인 활동방식을 그대로 수행

하였을 것으로 판단할 수 있는 반면, 아프카니스탄의 개발프로젝트는 마을사람들이 여성의 참여를 조건부로 하는 지원프로그램의 존재로 인하여 연구자 또는 최소한 정책주체의 존재를 알며 자신들이 특정한 상호작용의 맥락에 있다는 사실을 의식하였을 것으로 판단할 수 있다. 그럼에도 불구하고 이 연구는 인위적인 실험실 환경이 아니라 피험자들의 실제적인 삶의 공간에 대한 체계적인 개입이라는 점에서 현장실험으로 분류된다. 현장실험의 의미는 실험의 조건과 현장의 조건의 간극을 줄이는 것이고, 그 줄임의 방식과 정도에 따라서 현장실험의 종류를 나눌 수 있다. 현장실험에서 사실성 또는 자연성이라고 하는 것이 중요한 가치인데, 이는 처지(treatment), 참여자, 맥락, 결과(outcome measures)라고 하는 네 가지의 기준으로 판단할 수 있다(Green et al. 2011, 4).

Harrison and List(2004: 1014-5)는 현장실험의 종류를 그 현장성의 정도에 따라서 다음과 같이 분류한다. 먼저 인위적인 현장실험(artefactual field experiment)은 실험실 실험과 유사하지만 일반적인 실험실 실험과는 다른 피험자들을 충원하고, 때로는 피험자들을 실험실로 부르는 것이 아니라 실험실을 피험자들이 활동하는 현장으로 가지고가서 수행된다. 종종 학생이 아닌 관심대상 인구전체에 대한 대표성이 높은 표본집단을 대상으로 실험을 수행하기도 한다(Harrison, Lau and Williams 2002). 프레임화된 현장실험(framed field experiment)에서는 실험참여자들에게 대한 보상, 피험자들의 선택과 행동, 그리고 피험자들에게 제공되는 정보 등에 현장성을 부여한 형태로 수행된다. 자연적인 현장실험(natural field experiment)은 현장실험의 이상형으로서 피험자들이 자연적으로 상호작용하는 환경에 연구자가 개입하고 피험자들은 그러한 개입의 존재여부, 즉 실험의 존재여부를 알지 못하는 상황에서 관찰이 수행

된다.

정치학에서의 현장실험은 의외로 긴 역사를 지니고 있는데, 예를 들어 Gosnell(1926; 1927)의 투표참여 독려(get-out-the-vote) 실험은 정치학 실험의 고전으로 여겨진다. Gosnell은 12,000명의 시카고 시민을 선별하여 그 중에서 3,000명에게 선거에 대한 정보를 제공하는 우편물을 보내고 우편물을 받은 시민들과 받지 않은 시민들 사이에 유권자등록률을 비교하였다. 투표참여율을 높이는 독려 방법에 대한 현장 실험은 Gosnell 이후에도 계속 이어져 예를 들면 Eldersveld(1956)는 현장실험을 통해서 어떠한 방식의 투표참여 독려가 더 효과적인지를 연구하였다. Eldersveld의 실험은 전화나 방문을 받은 유권자들이 투표에 참여할 확률이 그러한 처치를 받지 않은 유권자들에 비해서 훨씬 높다는 것을 보이기는 하였으나 전화 통화나 방문에 응하는 유권자들이 대상자의 부분 집합에 불과한 자기선택(self-selection)의 문제로 인하여 그 효과가 과대하게 측정되었다는 비판을 받게 된다. 개인들을 대상으로 한 초기의 현장실험은 이와 같은 방법론적인 문제가 있었으며 동시에 현대적인 설문조사 기법의 발전으로 연구자들이 설문 방법을 선호하게 됨에 따라 현장실험 방법은 크게 확대되지 못하였다. Gosnell과 Eldersveld의 연구 이후에도 투표참여 독려에 대한 현장실험이 간간히 행해지기는 하였지만(Adams and Smith 1980; Miller, Bositis and Baer 1981 등) 2000년대 들어와서 Gerber와 Green 등 예일대학교 연구팀이 더욱 정교한 방법으로 대규모의 현장실험을 수행하여 현장실험 방법이 새로이 주목받게 되는 계기가 되었다.

예일대 연구팀의 기여(Gerber and Green 2000, 2001; Gerber, Green and Shachar 2003, Green and Gerber 2008 등)는 근본적으로 새로운 연구 질문이나 연구방법론에 있다기보다는 기존에 한계

에 부딪혔던 현장실험 방법을 그 설계와 규모, 결과의 분석에 이르기까지 체계적으로 혁신하여 훨씬 설득력 있는 연구결과를 제시하였다는 데 있다고 볼 수 있다. 예를 들어 Gerber and Green(2000)은 뉴헤이븐시의 약 삼만 명에 달하는 등록된 유권자들을 전화접촉 그룹과 가정방문 그룹으로 나누고 양 그룹 각각에서 일부에 대해서는 우편을 통한 선거참여 독려를 추가하였다. Gerber와 Green은 각각 방법의 효과를 측정함에 있어서 각 그룹의 일부가 연구자들의 시도에도 불구하고 결과적으로 직접적인 접촉이 성사되지 않았다는 점을 감안할 수 있는 방법론을 사용하였다. 그 결과 직접방문은 약 10% 투표율을 높이는 효과가 있으며, 우편물의 효과는 1% 미만, 그리고 전화통화의 효과는 유의하지 않음을 보였다.

정치학에서 현장실험은 그 동안 투표 참여와 정치커뮤니케이션 부분에서 가장 활발하게 진행되어 왔고 앞의 King et al.(2014)의 연구처럼 다른 주제들로 확산되고 있다. 그 중에서 특히 최근에 매우 활발하게 현장실험의 방법이 사용되고 있는 분야는 개발정치경제학(political economy of development) 분야이다. 이 분야에서 수행되는 현장실험은 흔히 무작위대조군시험(Randomized controlled trials, RCT)이라고 불린다. 무작위대조군시험 방법이 개발정치경제학적 현장실험에 많이 사용되기 시작하면서 마치 서론에서 제시한 예와 같은 공동체 단위의 현장실험을 RCT와 동일시하는 경우가 있다. 엄밀히 말하면 RCT는 현장실험에 고유한 방법이 아니라 일반적인 실험의 한 유형이며 특히 의학 임상실험에서 처치(treatment)의 효과를 정확하게 측정하기 위한 이상적인 실험 방법이라고 할 수 있다. 모든 실험이 항상 인과적인 효과를 측정하는 것을 목적으로 하는 것은 아니며 모든 실험에 반드시 통제군이 설정되어야 하는 것은 아니다. 무작위대조시험의 특징은 처치군과 통제군을 무작위의

방법으로 나누고 실험적 개입의 과정에서 두 집단 간에 실험적 처치의 여부 이외에 모든 부분에서 동일성을 유지하여, 처치의 효과를 분명하게 측정하는 것을 목표로 한다.

개발협력분야에서 무작위대조군시험 형식의 현장실험이 급속하게 확대되고 있는 것은 개발원조 프로그램의 효과측정상의 어려움에 대해서 학계와 공여국 그리고 국제기구가 오랫동안 인식을 공유하고 함께 그 해결책을 모색해 온 결과이다. 과연 개발원조 프로그램들이 수혜국 또는 수혜 공동체들에 대해 공여국과 공여기구가 의도한 긍정적인 효과를 낳고 있는가? 어떠한 방법의 개발원조가 더 효과적으로 개발협력의 목표를 달성하는가? 이러한 정책적인 질문들이 사회와 공동체의 발전 요인에 대한 사회과학의 오랜 연구 질문들과 결합되는 지점에서 개발정치경제학적 현장실험이 의미를 가지게 된다(Humphreys and Weinstein 2009). 개발정치경제학적 현장실험의 대표적인 주제와 연구들을 간략히 소개하자면 다음과 같다.

먼저 공동체 단위에서 정치제도 특히 의사결정규칙의 효과에 대한 연구들이 있는데, 이 중에서 Olken(2010)은 인도네시아의 49개 공동체들을 대상으로 한 현장실험에서 주민들의 직접참여에 의한 의사결정이 대표자 선출을 통한 의사결정 보다 주민들의 만족도와 지식획득 정도 등에서 훨씬 더 효과적임을 보였다. 이 글의 도입부에서 소개한 Beath et al.(2013)의 연구도 이러한 맥락에서 이해할 수 있다. 공동체의 자발적인 집합행동 해결 능력을 어떻게 향상시킬 수 있는가 역시 개발정치경제학의 주요한 관심사이다. Fearon et al.(2009)은 라이베리아에서의 현장실험을 통해서 공동체주도 개발 프로그램에의 참여 경험이 그 이후 공동체 구성원들의 자발적인 집합행동 능력이 향상에 기여함을 발견하였다.

정치인과 정부관리의 부패와 책임성의 부재가 개발의 주요한 걸림돌이 되고 있다는 일반적인 인식을 바탕으로 하여 이 주제에 대한 현장연구도 수행되고 있다. 예를 들어 Olken(2007)의 인도네시아에서의 현장실험은 감사(audit)의 위협과 같은 하향식 부패통제 방식이 주민들의 참여를 통한 방식에 비해서 실효성이 훨씬 떨어진다는 것을 보였으며, 마찬가지로 Bjorkman and Svensson(2009)은 우간다에서의 현장실험을 통해서 시민들에 의한 감독(oversight)이 보건서비스의 질을 제고함을 보였다.

개발정치경제학적 현장실험은 개발원조 프로그램의 효과를 측정하고 또한 정책과 사회과학이론의 결합을 통해 더욱 효과적인 정책의 설계에 기여할 것으로 기대된다. 그러나 동시에 몇 가지 극복해야 할 난제를 지니고 있기도 하다.³⁾ 무엇보다도 현장실험 방식을 통한 개발프로그램의 효과측정은 매우 긴 시간이 걸린다는 점이다. 위에서 소개한 연구들이 정책집행과 효과측정 사이에 수년 정도의 시간적인 간격을 두고 있어 일반적인 실험에 비해서 훨씬 더 긴 시기를 연구 대상으로 하고 있기는 하지만 이조차도 과연 효과를 측정하는데 충분한 시간인지 의심을 가질 수 있다. 또한 주로 공동체와 같은 집단이 분석 단위가 되기 때문에 인과적 추론을 가능케 할 정도로 충분한 수의 참여공동체를 확보하기가 용이하지 않고 비용이 많이 든다는 어려움이 있다. 인종이나 종교와 같이 개발정책적으로 중요시되는 몇 가지 변수들은 실험적 조작의 대상이 되기가 어렵다는 문제도 있으며 또한 현장실험의 수행이 가능하다는 사실 자체가 공동체의 중요한 속성을 반영한다고 볼 때, 이러한 연구의 결과들이 다른 지역이나 사례들에 얼마나 적실성 있는 정책적 함의를 지니는가에 대한 의문이 제기될 수도 있다. 더 큰 문제는 현장실험이 주로

3) 이에 대해서는 Humphreys and Weinstein(2009, 373-376) 참조.

마을단위의 공동체를 대상으로 이루어지며, 이를 국가와 같은 대단위를 대상으로 수행하기 어렵다는 점이다.

위와 같은 어려움에도 불구하고, 현장실험의 방법은 지금까지 난제로 여겨져 왔던 개발정책의 효과측정, 이론과 증거에 기반을 둔 개발정책의 수립에 앞으로 계속 중요한 기여를 할 수 있을 것으로 보인다. 개발정치경제학적 현장실험 그 자체가 수십 년 전만 하더라도 상상하기 힘들었다는 점을 고려하면 앞으로 위의 난제들을 극복을 통해서 이 분야에서 현장실험이 기여할 여지는 매우 크다고 보아야 할 것이다. 한국의 개발협력 규모가 2조원을 넘어섰고,⁴⁾ 규모의 증가에 부합하는 개발협력 정책의 질적인 수준 제고가 문제시되고 있는 현실에서 최근에 국제 개발학계에서 국제원조기구들과 협력하여 광범위하게 진행하고 있는 현장실험 방식의 개발원조 집행은 충분히 주목할 만한 가치가 있다.

3. 설문실험

설문조사 자료에 대한 통계적인 분석은 정치학에서 시민들의 정책과 정당에 대한 선호를 연구하는 기본적인 방법으로 정착되어왔다. 그러나 설문에 근거한 연구는 대표성의 문제, 허위 관계(spurious correlation), 측정오류, 내생성의 문제 등 인과적 추론을 어렵게 만드는 여러 가지 한계를 지니고 있다(Gains, Kuklinski, and

4) 정부는 2014년 공적개발원조(ODA) 총규모는 약 2조 2,666억원으로, 국민총소득(GNI) 대비 비율은 0.16% 내외가 될 것으로 추정하고 있다(국무조정실 등, 2014).

Quirk 2007). 설문실험(survey experiment)은 설문연구의 틀 내에서 이러한 문제들을 해결하고자 하는 유력한 방법으로 발전되어왔다.

초기의 설문실험은 특정한 연구 질문에 답하기 위해 발전했다기 보다는 설문이라고 하는 방법자체의 문제점을 발견하고 그에 대처하는 방안을 모색하고자 하는 목적이 강했다고 볼 수 있다. 예를 들어 냉전시기 미국인들은 소련의 기자들이 미국에서 활동하는 것을 허용해야 한다고 생각하였을까? 이러한 의문에 답하기 위해 당연히 되었던 방법은 시민들에게 설문을 통해서 물어보는 것이다. 그러나 Hyman and Sheatsley(1950)는 이 질문이 어떠한 방식으로 제시되는가가 설문의 결과에 큰 영향을 미친다는 것을 발견하였다. 즉 이 질문에 앞서 미국기자들이 소련에서 취재를 하고 있다는 정보가 제공되었느냐의 여부가 미국시민들의 소련기자들에 대한 수용성 정도를 크게 변화시킨다는 것을 발견한 것이다. 문항의 순서, 자구, 제공되는 정보의 차이 등에 따라서 설문 결과가 달라진다는 것은 시민들의 견해의 부재 또는 불안정성에 대한 증거로 이해되기도 한다(Zaller 1992). 그러나 일부 정치학자들은 이러한 경향을 부정적으로만 보지 않고 오히려 더 정교한 인과적 추론을 위한 기회로서 적극적으로 활용하고자 하였다. 예를 들어, 위의 결과는 적대국의 언론에 대한 개방성은 호혜성의 원칙에 입각하여 결정되어야 한다는 당시 미국인들의 여론을 드러내는 것으로 해석될 수 있는 것이다.

일반적으로 실험이 자연적 과정에 대한 연구자의 의도적이고 체계적인 개입으로 정의될 수 있다면 설문 실험은 설문조사 과정에 대한 체계적인 개입인 것이며 특히 질문의 순서, 질문의 자구, 질문의 프레임, 질문과 결부되어 제공되는 정보 등을 체계적으로 변화시켜 그러한 변화에 따른 피설문자들의 반응의 차이가 있다면 이를

바탕으로 하여 피설문자들의 의사결정과 선호형성의 메커니즘에 대해 인과적인 추론을 하고자 하는 시도로 정의될 수 있다.

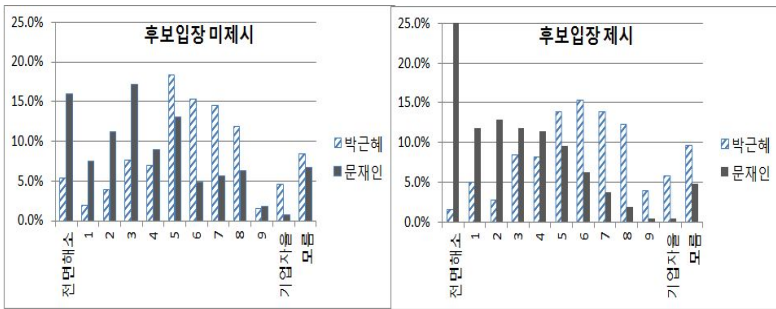
예를 들어 박원호·안도경·한규섭(2012)의 제 18대 대통령 선거에서 경제민주화 이슈에 대한 유권자들의 선호형성에 대한 연구에서는 다음과 같은 방식으로 후보자선호와 정책선호간의 인과관계를 연구하였다. 이 연구가 분석하고 있는 실험설문 문항들 중의 하나는 18대 대선을 전후하여 정치적인 이슈가 되었던 재벌그룹의 순환출자에 대한 입장을 묻고 있다. 설문참여자들을 무작위로 두 그룹으로 나누어서 한 그룹에게는 이 문제에 대한 후보자들의 입장을 밝힌 상태로, 다른 그룹에게는 후보자들의 입장을 제시하지 않고서 순환출자 규제에 대한 입장을 물었다. 구체적인 문항은 다음과 같다:

- ▶ [후보 입장 제공 프레임]: “재벌의 순환출자를 전면적으로 해소해야 한다는 입장이 제기되는 한편 기존 순환출자를 인정하고 신규만 금지해야 한다는 입장도 제기되고 있습니다. 이 문제에 대한 귀하의 입장은 어디에 가장 가깝습니까?”
- ▶ [후보 입장 제공 프레임]: “이번 19대 대선의 양대 후보는 향후 재벌의 추가적인 순환출자를 금지하겠다는 입장을 밝혔습니다. 이와 관련해 민주당 문재인 후보는 재벌의 순환출자를 전면적으로 해소해야 한다는 입장인 반면 박근혜 후보는 기존 순환출자를 인정하고 신규만 금지해야 한다는 입장입니다. 이 문제에 대한 귀하의 입장은 어디에 가장 가깝습니까?”

[그림 1]은 실험설문의 결과를 보여준다. 그림에서 나타나는 바와 같이 순환출자 규제에 대한 유권자들의 정책 선호는 후보자의 입장에 대한 정보가 제시되었는가의 여부에 따라 상당한 영향을 받는다는 것을 알 수 있다. 박근혜 후보와 문재인 후보 지지자들의 선호는 후보입장이 제시되지 않았을 때도 5.45 대 3.49로 유의한 차이를 보이지만 후보입장이 제시되었을 때는 5.70 대 2.65로 그 차

이가 더 벌어진다. 특히 문재인 후보 지지자들은 후보자입장을 제시하지 않은 설문에서는 3.49, 제시되었을 때는 2.65로 그 평균치와 분포에서 큰 차이를 보인다. 따라서 이 연구는 설문실험을 통하여 기존에 정책으로부터 후보자 선호라고 하는 고전적 민주주의 이론의(Downs 1957; Dahl 1971; Schumpeter 1942) 인과관계와는 역의 방향을 가지는, 즉 후보자에 대한 선호를 바탕으로 하여 정책에 대한 선호가 결정되는 메커니즘의 존재를 보여주고 있는 것이다.

[그림 1] 순환출자에 대한 설문실험 결과



* 출처: 박원호 등 2013, 21쪽 그림 3.

설문 순서 효과(question order effect)나 어구 효과(phrasing effect)에 대한 설문실험이 설문을 통한 선호측정의 내적인 타당성 또는 신뢰도(reliability)를 검증하고 제고하는데 목적을 두고 있는 반면, 위의 예와 같은 프레이밍효과에 대한 설문실험은 오히려 다른 설문 방식을 실험적 처치(treatment)로 활용하여 가설검증과 인과적 추론을 수행한다. 프레이밍은 정보내용에 대한 '해석적인 구조'에서부터 이미지와 은유, 판단의 단서(cue)에 이르기까지 정보 내용의 해석 및 그에 대한 반응에 영향을 미치는 제반의 요인들을 지칭한다(Kinder and Nelson 2005). 프레이밍에 대한 설문실험은 정보의

등가성을 전제로 하고 다른 전달방법의 효과를 검증할 수도 있고, 또 반대 극단에서는 구조화된 인식체계 자체의 효과를 검증하는 것을 목표로 할 수도 있다. 그 사이에서 더 흔히 사용되는 프레이밍은 선택의 기준이나 강조점(Gitlin 1980)에 따른 효과를 보고자한다. 이러한 설문실험 연구의 대표적인 예는 소우인중우대정책에 대한 백인들의 입장에 대한 연구(Kinder and Sanders 1990) 그리고 정부의 빈곤층 보조 정책에 대한 시민들의 선호 연구(Sniderman and Therault 1999) 등이 있다.

프레이밍에 따라 설문의 결과가 달라지는 것을 어떻게 해석할 것인가? 이에 대해서는 Zaller(1992)와 같이 대중 여론의 불안정성에 대한 Converse(1964)의 이론의 연장선에서 이상적인 유권자와 현실 유권자의 차이로 해석할 수도 있다. 그러나 프레이밍 효과를 반드시 부정적인 것으로 볼 필요는 없다. 즉, 프레이밍 효과는 복잡한 정책 이슈에 대한 선호의 형성에 있어서 시민들이 가용한 정보와 문제의 맥락을 고려하고 있음을 보여주는 것으로 긍정적인 해석도 가능한 것이다(Druckman 2001).

설문실험의 방법론적인 쟁점은 크게 효과의 지속성의 문제(duration effect), 내생성(endogeneity) 문제의 잔존, 그리고 통제 그룹의 불명확성 또는 누락 등의 문제가 있다(각 문제에 대한 자세한 내용 및 그 해결을 위한 시도들은 Gains et al. 2007을 참조할 것). 즉 설문실험은 실험이라고 하는 방법에도 불구하고 여전히 실험실 실험이나 현장실험에 비해 불명확하게 정의되는 태도나 선호에 대한 측정이 그 결과로 산출되고 따라서 그 현실의 정치 과정과 행태에 대한 함의가 직접적이지 않다는 문제점이 있는 것이다. 그럼에도 불구하고 설문실험은 실험적 요소가 없는 일반적인 설문의 방식에 비해서 훨씬 더 풍부한 자료를 제공하여 그 사용의 빈도가 꾸준히

준히 증가하고 있다.

프레이밍 효과 이외에도 다른 한 가지 중요한 설문실험의 방법은 열거식 설문실험(list experiment)이다. 열거식 설문실험은 사회적으로 민감한 문제, 또는 ‘정치적인 올바른(politically correct)’ 답이 비교적 명백한 설문문항에서 설문참여자들이 자신의 진정한 견해를 드러내기를 꺼려함으로 인하여 발생하는 측정의 신뢰성(reliability) 문제를 극복하기 위한 설문 방법이다(김승현 2010; Druckman et al. 2006). 예를 들어 설문대상자들의 부패에 대한 인식을 조사하기 위해서 다음과 같은 진술에 대한 찬반을 묻는다고 가정해 보자: “가족이나 친한 친구가 부탁하면 자신의 지위를 이용하여 어느 정도의 특혜를 주는 것은 용인될 수 있다.” 이러한 질문에 대하여 과연 응답자들은 자신의 생각을 정직하게 표현할 것인가? 부패의 부정적인 영향에 대한 공적 담론이 활성화되어 있을수록 아마 자신의 진정한 견해를 숨기는 경우가 많을 것이다. 이 문제는 예를 들어 다음과 같은 방식의 열거식 실험으로 다룰 수 있다. 일단 설문대상자들을 무작위로 두 그룹으로 나눈 다음 통제집단에 대해서는 위의 문항이 포함되지 않은 서너 개의 진술을 제시하고 그 중에서 동의하는 진술의 수를 기입하게 한다. 이 때 설문 참여자들은 각각의 진술들에 대한 찬반이 아니라 열거된 진술들 중에서 찬성하는 진술의 개수만을 기입해야 한다. 이 문항들은 규범적인 정답이 존재하지 않는 것들로 선택하는 것이 좋을 것이다. 예를 들면, ‘내각제 개헌이 필요하다,’ ‘증세를 통해서라고 개발협력 예산을 확대해야 한다’ 등과 같은 문항들을 생각해볼 수 있다. 반면 처치그룹(treatment group)에는 위의 문항들에 더하여 부패용인에 대한 문항을 포함시킨다. 만약 통제그룹에서 열거된 진술들 중에 평균적으로 2개에 찬성하고 처치그룹에서는 2.5개에 대해서 찬성하였다면,

www.kci.go.kr

양 집단의 수가 충분히 크고 응답자의 분할이 무작위로 이루어졌다는 가정 하에, 설문대상 인구의 50%는 제시된 맥락에서의 특혜가 용인될 수 있다는 견해를 가진 것으로 추론할 수 있다. 여기에 기존의 연구, 또는 병행연구를 통해서 부패의 용인에 대한 진술에 대한 직접적인 찬반을 측정한 결과가 있다면 사회적 수용성 또는 정치적 올바름에 대한 고려로 인하여 직접 설문에서 진정한 선호를 드러내지 않은 응답자의 비율 또한 추론할 수 있다. 예를 들어 직접설문에서 찬성의견 표명이 30%에 머물렀다면 이는 20%의 응답자들이 자신의 견해를 그대로 드러내지 않은 것으로 추론할 수 있는 것이다. 설문참여자의 수가 충분하다면 특히 어느 집단(성별, 세대별, 교육 수준별 등)에서 부패에 대한 용인의 정도가 더 높은지, 또한 어느 집단이 진정한 견해를 숨기는 정도가 더 높은지에 대한 추론도 가능할 것이다.⁵⁾

4. 결론을 대신하여

지금까지 이 논문은 정치학에서 사용되는 실험연구 방법을 실험실 실험, 현장 실험, 설문 실험으로 나누어 각 방법의 내용과 한계 그리고 대표적인 연구들에 대해서 간략하게 소개하였다. 소개한 바와 같이 예전에는 실험의 대상이라고 여겨지지 않았던 주제들에 대해서 인터넷의 발달 정책입안자들과의 협력, 그리고 연구자의 창의성을 바탕으로 하여 실험연구의 적용 영역이 확장되고 있으며 중요

5) 열거식 설문실험의 설계와 결과 분석과 관련한 방법론적인 쟁점에 대해서는 Corstange(2009)와 Glynn(2013) 등이 논하고 있다.

한 연구업적들이 생산되고 있다. 이러한 연구들은 다른 방법을 사용한 연구들과의 상호보완을 통해서 정치학의 질문들을 다듬고 그에 대한 더 만족할 만한 대답을 제시하는데 기여할 것으로 기대된다. 정치학은 과거에도 현재에도 미래에도 항상 그 주제와 방법론의 다양성을 통해서 발전해왔고 발전할 것이다. 지식의 축적과 연구의 필요, 그리고 기술의 발전, 사회적 조건의 변화 등에 따라서 새로운 연구 방법이 필요해지고 가능해진다. 중요한 것은 연구의 질문과 그 질문에 답하는 과정에서 얼마나 흥미있는 이야기를 전할 수 있는가 하는 것이다. 방법론에 대한 열린 자세로 임할 필요가 있는데 현재에 이는 실험 방법에 대한 편견을 버리고 그 가능성을 충분히 활용하는 것을 의미한다.

투고일: 2014년 9월 13일

심사일: 2014년 9월 29일

게재확정일: 2014년 10월 1일

참고문헌

- 국무조정실·기획재정부·외교부. 2014. 『2014년 ODA 합동 워크샵 자료집』.
- 김승현. 2010. “가족주의와 공공성 : 전통적 가치와 현대적 가치의 측정에서 사회적 소망성 편향.” 『한국정치학회보』 44(3). 53-74.
- 박원호. 2013. “생태학적 추론(Ecological Inference) 서설: 통계학적 연금술 혹은 선거사 연구의 희망?.” 『평화연구』 21(2). 395-426.
- 박원호·안도경·한규섭. 2013. “제18대 대통령 선거의 경제민주화 쟁점에 대한 실험설문: 프레임링 효과를 중심으로.” 『한국정치연구』 22. 1-28.
- Ahn, T.K., Robert Huckfeldt, and John B. Ryan. 2010. “Communication, Influences, and Information Asymmetry among Voters.” *Political Psychology* 31(5). pp.763-787.
- Ahn, T.K., R. Huckfeldt, A. Mayer, and J.B. Ryan. 2013. “Expertise and Bias in Political Communication Networks.” *American Journal of Political Science* 57(2). pp.357-373.
- Ansolabehere, S. and S. Iyengar. 1994. *Going Negative: How Attack Ads Shrink and Polarize the Electorate*. Free Press New York.
- Adams, William C. and Dennis J. Smith. 1980. “Effects of Telephone Canvassing on Turnout and Preferences: A Field Experiment.” *Public Opinion Quarterly* 44(3). pp.389-395.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Battaglini, Marco, Rebecca Morton, and Thomas R. Palfrey. “The Swing Voter's Curse in the Laboratory.” *Review of Economic Studies* 77(1). pp.61-89.
- Beath, Andrew, Fotini Christina, and Ruben Enikolopov. 2013.

- “Empowering Women through Development Aid: Evidence from a Field Experiment in Afghanistan.” *American Political Science Review* 107(3). pp.540–557.
- Bianco, William T., Michael S. Lynch, Gary J. Miller, and Itai Sened. 2008. “The Constrained Instability of Majority Rule: Experiments on the Robustness of the Uncovered Set.” *Political Analysis* 16. pp.115–137.
- Chaudhuri, Ananish. 2011. “Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature.” *Experimental Economics* 14(1). pp.47–83.
- Converse, Paul E. 1964. “The Nature of Belief Systems in Mass Publics.” In D.E. Apter, ed. *Ideology and Discontent*, 206–61. New York: Free.
- Corstange, Daniel. 2009. “Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT.” *Political Analysis* 17. pp.45–63
- Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Dal Bo, Pedro. 2013. “Experimental Evidence on the Workings of Democratic Institutions.” Working Paper.
- Dickson, Eric S., Catherine Hafer, and Dimitri Landa. 2008. “Cognition and Strategy: A Deliberation Experiment.” *Journal of Politics* 70(4). pp.974–989.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Druckman James N. 2001. “The Implications of Framing Effects for Citizen Competence.” *Political Behavior* 23. pp.225–56.
- Eldersveld, Samuel T. 1956. “Experimental Propaganda Techniques and Voting Behaviour.” *American Political Science Review* 50(1). pp.154–165.
- Fearon, James D., Macartan Humphreys, and Jeremy M. Weinstein.

2009. “Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post–Conflict Liberia?” *American Economic Review Papers & Proceedings* 99(2). pp.287–291.
- Fudenberg, Drew and Eric Maskin. 1986. “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information.” *Econometrica* 54(3). pp.533–54.
- Gains, Brian J., James H. Kuklinski, and Paul J. Quirk. 2007. “The Logic of the Survey Experiment Reexamined.” *Political Analysis* 15. pp.1–20.
- Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis and Interpretation*. New York: Norton & Company.
- Gitlin, Todd. 1980. *The Whole World Is Watching: Mass Media in the Making & Unmaking of the New Left*. Berkeley: University of California Press.
- Glynn, Adam N. 2013. “What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment.” *Public Opinion Quarterly* 77. pp.159–72.
- Gosnell, Harold F. 1926. “An Experiment in the Stimulation of Voting.” *American Political Science Review* 20(4). pp.869–874.
- Gosnell, Harold F. 1927. *Getting–Out–The–Vote: An Experiment in the Stimulation of Voting*. Chicago: Chicago University Press.
- Green, Donald P., Allison J. Sovey, and Joel Middleton. 2011. “Political Communication: Insights from Field Experiments.” Working Paper.
- Green, Donald P., Allison J. Sovey, and Joel Middleton. 2011. “Political Communication: Insights from Field Experiments.” Working Paper.

- Grosser, Jens, and Arthur Schram. 2006. "Neighborhood Information Exchange and Voter Participation: An Experimental Study." *American Political Science Review* 100. pp.235–248.
- Grosser, Jens, and Arthur Schram. 2010. "Public Opinion Polls, Voter Turnout, and Welfare: An Experimental Study." *American Journal of Political Science* 54(3): 700-17.
- Harrison, Glenn W., Morten Igel Lau and Melonie B. Williams. 2002. "Estimating Individual Discount Rates for Denmark: A Field Experiment." *American Economic Review* 925. pp.1606–17.
- Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* XLII. pp.1009–1055.
- Human Rights Watch 2012. Country Summary: China.
- Humphreys, Macartan, and Jeremy M. Weinstein. 2009 "Field Experiments and the Political Economy of Development." *Annual Review of Political Science* 12. pp.367–378.
- Iyengar, Shanto and Kyu S. Hahn. 2009. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication* 59. pp.19-39.
- Kinder, Donald.R., and Thomas E. Nelson. 2005. "Democratic Debate and Real Opinions." In K Callaghan and F. Schnell. Eds. *Framing American Politics*, pp.103–122. Pittsburgh: University of Pittsburgh Press
- Kinder, Donald R. and Thomas R. Palfrey. 1993. "On Behalf of an Experimental Political Science." In Donald R. Kinder and Thomas R. Palfrey eds., *Experimental Foundations of Political Science* 1–39. Ann Arbor: University of Michigan Press.
- Lodge, Milton, Kathleen M. McGraw, and Patrick Stroh. 1989. "An Impression-Driven Model of Candidate Evaluation." *American Political Science Review* 83. pp.399-420.

- Lodge, Milton, Marco Steenbergen and Shawn Brau. 1995. "The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation." *American Political Science Review* 89. pp.309-326.
- McKelvey, Richard D., and Peter C. Ordeshook. 1983. "Some Experimental Results that Fail to Support the Competitive Solution." *Public Choice* 40. pp.281-91.
- _____. 1990. "A Decade of Experimental Research in Spatial Models of Elections and Committees. In *Advances in the Spatial Theory of Voting*, ed. James M. Enelow and Melvin J. Hinich, 99-144. Cambridge: Cambridge University Press.
- Miller, Roy E., David A. Bositis and Denise L. Baer. 1981. "Stimulating Voter Turnout in a Primary: Field Experiment with a Precinct Committeeman." *International Political Science Review* 2. pp.445-459.
- Olken Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115. pp.200-49.
- Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104. pp.243-267.
- Orbell, John M., and Robyn M. Dawes. 1991. "A 'Cognitive Miser' Theory of Cooperators' Advantage." *American Political Science Review* 85. pp.515-28.
- Orbell, John M., and Robyn M. Dawes. 1993. "Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game." *American Sociological Review* 58. pp.787-800.
- Ostrom, Elinor, 1998, "A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association." *American Political Science Review* 92. pp.1-22.

- Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 86. pp.404-17.
- Reporters Without Borders. 2010. China. <https://en.rsf.org/china-china-12-03-2012,42077.html>
- Schumpeter, Joseph A. 1942(2006). *Capitalism, Socialism and Democracy*. Routledge.
- Smith, Vernon L. 1982. "Microeconomic Systems as an Experimental Science." *American Economic Review* 72(5). pp.923-955.
- Steven D. Levitt, Steven D., and John A List. 2006. "What Do Laboratory Experiments Tell Us About the Real World?" Working Paper.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

Political Science and Experiments

Toh-Kyeong Ahn(Seoul National University)

An experiment is an intervention into naturally occurring situations in order to measure objects or processes that are not feasible under natural circumstances or to make causal inferences that are limited with naturally occurring data. Political science has long included both empirical and normative research, but experimentation, a gold standard of scientific method, has been considered impossible or irrelevant. Technological advances, policy opportunities and researchers' creativity have now changed the situation and experiments are now becoming one of the essential research methods in political science. This paper introduces three types of political science experiments – laboratory, field, and survey experiments – and reviews representative research in each of the three types.

Keywords: Political Method, Experimental Method, Political Experiments