

베이즈망에 토대한 인과 모형 이론의 문제로서 심슨 역설:

충실성 조건에 대한 이영의 선생님의 비판과
이에 대한 응답[†]

김 준 성[‡]

이영의 선생님(2015)은, 심슨 역설이 대표적인 인과(구조) 모형 이론의 SGS(Spirites, Glymour, Scheins 1993) 이론의 충실성 조건에 반례가 된다는 문제를 논의한다. 이 반례는 베이즈망에 토대한 베이즈 인과 이론에도 문제가 된다. 이영의 선생님의 주장은 카트라이트(Cartwright 2001[2003]) 등 SGS 이론에 비판적인 입장에 일부 함께 하고 있다. 필자는, 켈리(Kelly 2007; 2008; 2010; 2012)가 발전시킨 오캄의 단순성(경제성) 원리에 토대하여, 인과 추론의 방법론적 관점에서 SGS의 충실성 조건이 의미를 가질 수 있다는 것을 보여준다. 특별히 예측(발견)과 설명(정당화, 해명)의 구분을 통해 SGS 이론은 예측의 효율성에 초점을 두고 있으며, 충실성 조건도 그 맥락에서 유의미하다는 것을 보여준다. 이 논의는 SGS 이론과 이 이론에 대한 비판적 입장 모두를 함께 양립하게 할 수 있을 것이다.

【주요어】 단순성, 마코프 조건, 베이즈망, 베이지안, 심슨 역설, 오캄,
이영의, 인과 구조 모형, 충실성 조건, 켈리, SGS

[†] 이 연구의 계기를 마련해 주신 이영의 선생님과 북 심포지엄으로 이끌어 주신 최훈 선생님께 감사합니다. 또한, 이 글의 오류를 확인해 주시고 개선 될 수 있는 중요한 조언을 해 주신 심사위원 선생님들께 감사합니다.

[‡] 명지대학교 철학과, jkim30@mju.ac.kr.

1. 들어가는 말

오늘날 베이즈 이론(Bayesian theory)의 눈부신 성과를 볼 때 이 이론의 적절성을 논의하는 것은 무의미하다. 오히려 베이즈 이론이 그렇게 다양한 많은 영역에서 왜 성공적이었는지 그리고 그 성공을 또 다른 영역으로 어떻게 확장할지를 논의하는 게 유의미할 것이다. 이론이나 모형의 적절성은, 세계의 질서를 우리의 인식이 얼마나 잘 통찰하는지에 달려있다. 다시 말해 세계의 질서를 이해하는 데에 적절한 논리를 세우는 데에 있다. 그 적절한 논리를 세우는 것을 합리성이라 부를 수도 있다. 논리는 이론을 객관화하는 데에 절대적이다. 따라서 그 합리성은 이론이 한 개인에게만 설득력을 갖는 게 아니라 모두를 설득시키는 것이다. 또한 이론을 통해 세계에 대한 이해뿐 아니라 세계에 대한 예측과 활용도 가능하게 해야 한다. 베이즈 이론은 바로 그 합리성이 성공한 것을 잘 보여주는 경우로 볼 수 있다. 그러나 이영의 선생님은, 최근에 출간한 저서 『베이즈주의: 합리성으로부터 객관성으로의 여정』(2015)에서 베이즈 이론이 전제하는 주관적 합리성만으로는 과학성을 확보하기 어렵고 베이즈 이론의 성공 여부는 객관성을 확보하는 정도에 달려있다고 주장한다. 이 책에서 이영의 선생님은 베이즈 이론의 합리성에 관한 이론적 틀과 베이즈 이론의 문제들을 논의하면서 이 이론의 객관성을 어떻게 확보하여 줄 수 있는지를 보여준다.

베이즈 이론의 수월성을 보여주는 요소 가운데 베이즈망(Bayes net)은 특별하다. 베이즈망은 세계의 질서를 가시적으로 재현하는 데에 있어서 탁월하다. 베이즈망은 세계의 인과 구조를 이해할 수 있게 하고, 상관(correlation)이란 규칙성에서 인과의 필연적 구조를 추론할 수 있게 한다. 인과에 대한 통계적 이론들과 철학적 이론들의 성과는 인과(구조) 모형(causal model, or modeling) 이론으로 수렴된다고 볼 수 있다. 인과 모형 이론에서 베이즈망의 역할은 절대적이다. 인과 모형에 대한 여러 이론 가운데 카네기멜론 대학 철학과의 SG(Scheines, Glymour, Spirtes 1993; 2000)의 연구가 UCLA 컴퓨터공학과의 퍼얼(Pearl 2000)의 연구와 함께 가장 많은 주목을 받아 온 것은 분명하다.¹⁾

1) 인과 르네상스라 부를만한 인과에 대한 논의가 철학뿐 아니라 컴퓨터 과학, 통계학, 경제학 등에서 특별히 1990년대를 시작으로 현재까지 진행되고 있다고 보아도 과장은 아니다. 이 논의의 유형은 세 가지로 구분될 수 있다. 인과를 확률로 해명하는 인과에 대한 확률 이론, 반사실적 의존 관계로 인과를 해명하는 인과에 대한 반사실적 이론, 구조 방정식 모형에 토대한 인과 구조 모형 이론이다. 인과에 대한 확률 이론은, 라이헨바흐(H. Reichenbach), 굿(I. J. Good), 수퍼(P. Suppe)에서 창안되어 카트라이트(N. Cartwright), 스컴즈(B. Skyrms), 새먼(W. Salmon), 소버(E. Sober), 엘스(E. Eells), 히치콕(C. Hitchcock) 등을 통해 그 이론의 체계가 갖추어졌다. 인과에 대한 반사실적 이론은, 험(D. Hume)의 아이디어에서 출발하여 루이스(D. Lewis)와 루이스의 이론을 발전시킨 후학들에 의해 진화되어 왔다. 이들 두 이론 모두, 철학의 전통에 토대한다는 공통점을 갖고 있다. 또한 경제학, 법학 등 사회과학, 컴퓨터 과학에서의 인과 논의에 여전히 주목을 받고 있다. 인과 구조 모형 이론은, 퍼얼(J. Pearl), SGS, 할펀(J. Halpern), 후버(K. Hoover) 등이 중심에 있으며 컴퓨터 과학, 심리학, 통계학, 철학의 논의가 함께 녹아 있는 학제 간 융합 이론의 특성을 갖고 있다. 이들은 확률에 대한 인과 이론에 비판적인 반면에, 반사실적 인과 이론을 적극적으로 수용하여 인과 모형 이론을 계속 발전시키고 있다. 인과 모형 이론은 앞 서 두 이론과 비교하여 전반적으로 네 가지 점에서 차별성을 갖는다. 첫째, 인과에 대한 해명이나 정당화보다는 인과 추론(발견, 예측, 설명)에 대한 연구에 더 큰 비중을 두고 있다는 것이다. 둘째, 인과성을 전제하고 있는 반면에, 인과 개념들에 대한 언급이 최소화되기를 지향한다는 것이다. (세 이론 모두, 인과를 다른 개념으로 환원하는 것이 아니라는 점에서 일치한다.) 셋째, 구체적인 형식화에서 보다 확연히 구분된다. 인과 모형 이론은 구조 방정식의 형식을 가지며, 오차항(error terms)이 보여주듯이 결정론적인 해석을 전제한다. 인과 상관자를 변수들과 이들의 값으로 고려한다. 인과 모형은 이를 변수 간의 함수 관계를 가리킨다. 이를 변수의 관계를 그림(DAG)으로 가시적으로 구조화한다. 넷째, 인과 모형을 응용하고 활용하는 데에 적극적이어서, 실용적으로 유용하다. 인과 모형 이론의 연구자들 간에도 형식과 내용에서 차이가 있다. 예를 들어 퍼얼과 SGS 모두, 컴퓨터 과학의 연구 프로그램에서 인과 기계로서 인공지능 개발에 중점을 두고 있다. 퍼얼이 반사실적 인과 이론 등 인과에 대한 개념적, 철학적 논의를 수용하는 데에 적극적인 반면에, SGS는 논쟁이 될 그런 논의를

SGS 이론에 대한 높은 인지도에도 불구하고, 형식화된 이론의 특성 때문인지 그 이론의 내용과 문제를 이해하는 게 그렇게 쉽지는 않다. 그러나 이영의 선생님의 저서 4장 4절과 6장 4절을 통해서, SGS 인과 모형 이론의 핵심과 베이즈망 그리고 그 이론의 문제에 쉽게 다가갈 수 있다.

심슨 역설(Simpson's paradox)과 관련하여 SGS 이론의 베이즈망과 확률 분포의 관계에서 발생하는 문제를 이영의 선생님²⁾은 잘 설명해 주고 있다.³⁾ 이영의 선생님은, 심슨 역설과 SGS 이론의 문제에 대한 논의가 베이즈 이론에 객관성을 확보하는 것과 어떤 관계가 있는지를 명시적으로 언급하지는 않는다. 이영의 선생님은, 동질적인 부분 집단에 상대하여 인과 구조를 해명하는 게 심슨 역설의 해결 방안임을 확인한다. 이 논의가 베이즈 이론의 객관성 문제와 어떤 관련이 있는지는 분명하지 않다. 그러나 충분히 미루어 짐작할 수는 있다. 집단에 상대하여 객관적인 인과 구조를 해명할 수 있는 것이고, 이런 해명이 베이즈 이론에 객관성을 확보해 준다고 볼 수 있다. 그러나 필자는, SGS 이론에 대한 이영의 선생님의 비판과 관련하여 우리는 SGS 이론의 목적이나 목표를 주목할 필요가 있다고 본다. 인과적 예측, 인과적 추론 등 발견의 맥락에서 SGS 이론은 중요한 역할과 의미를 가지며, 이 사실은 베이즈 인과 추론의 객관성도 확보하여 준다고 본다.

SGS 이론의 핵심은, 변수들 간의 확률적 독립성(probabilistic, or statistical independence)으로 베이즈망을 통해 변수들 간의 인과 구조를 이해하고 추론(발견)할 수 있게 하는 데에 있다.⁴⁾ 이를 가능하게 하는 세 가지 조건이 있는데 그 중에서 마코프 조건(Markov Condition)과 충실성 조건(Faithfulness Condition)이 중요하다. 또 다른

최소화하면서 인과 구조를 추론(발견)하는 방법론적 효율성에 더 초점을 두고 있다. 따라서 인과에 대한 개념들의 최소화와 그 방법론적 효율성을 전제하고 SGS의 이론과 그 이론의 토대가 되는 핵심 조건들을 살펴야 한다.

2) 이영의 (2015), pp. 182-88.

3) SGS (1993), pp. 64-69.

4) SGS (1993), pp. 25-40; 이영의 (2015), pp. 286-93.

하나는 최소성(Minimality) 조건이다.⁵⁾

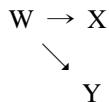
노드 V 에 대한 확률 분포 P 가, 다음 방식으로 그래프(Directed Acyclic Graph; 이후 DAG, 다시 말해 베이즈망)와 연결된다면, 확률 분포 P 는 마코프 조건을 충족한다. 그 연결 방식은 다음과 같다.

노드 V 의 모든 변수 X 와, X 의 후손을 포함하지 않는 모든 변수 Y 에 대해,

X 와 Y 가 확률적으로 독립적인 DAG이면,

$$\Pr(X \mid \text{조상}(X) \ \& \ Y) = \Pr(X \mid \text{조상}(X)).$$

다시 말해, 변수 X 에 대해 X 의 조상 변수인 $\text{조상}(X)$ 의 값이 주어질 때, (X 의 후손이 아닌) Y 의 값은 X 의 확률에 더 이상의 차이(변화)를 만들지 않는다. 이 조건이 인과 모형 내에서 의미하는 바는 다음과 같다: $\text{조상}(X)$ 가 X 의 원인이 되고 X 가 Y 의 원인이 아닌 구조, 또는 DAG이면, $\text{조상}(X)$ 가 주어질 때 X 의 확률은, $\text{조상}(X)$ 와 (X 의 후손이 아닌) Y 가 주어질 때 X 의 확률은 같다는 것이다. (마코프 조건을 충족하는 DAG과 확률 분포를 포함하는 인과 모형을, 인과 베이즈망이라 부른다.) 아래 <그림 1>이 마코프 조건을 충족하는 것을 쉽게 볼 수 있다. <그림 1>에서 W 는 X 의 조상이고, Y 는 X 의 후손이 아닌 변수이다.



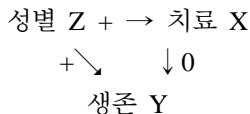
<그림 1>

충실성 조건은 마코프 조건의 역관계이다. 따라서,

5) 최소성 조건에 따르면, 마코프 조건을 충족하는 DAG이 있다고 할 때, 이 것을 구성하는 어떤 하부 DAG도 마코프 조건을 만족하지 않는다.

$\Pr(X \mid \text{조상}(X) \& Y) = \Pr(X \mid \text{조상}(X))$ 이면, X와 Y가 확률적으로 독립적인 DAG이다.

인과의 베이즈망에 여러 문제가 제기되어 왔다. 특별히 변수들 간의 확률적 독립성으로는 인과 구조로서 베이즈망을 충실히 드러내지 못한다는 문제, 다시 말해 충실성 조건에 대한 반례가 쟁점이 되었다. 충실성 조건의 반례로 제시된 것이 심슨의 역설(Simpson's paradox)이다. 이 역설이 왜 반례인지는 다음과 같다. 이영의 선생님의 저서와 SGS의 책에서 볼 수 있듯이,⁶⁾ X는 치료, Y는 생존, Z는 성별이 된다. 치료 X만이 주어졌을 때, 치료 X는 생존 Y에 대해 중립적이거나 부정적인 원인 효과를 가졌다. 그러나 성별 Z와 치료 X가 함께 주어졌을 때, 치료 X는 생존 Y에 긍정적인 원인 효과를 가진 것으로 나타났다. 정리하면, 확률적 관계는 다음과 같다. X와 Y에 대한 어떤 조건화도 없을 때 X와 Y는 확률적으로 독립 관계이다. 따라서 $\Pr(X \mid Y) = \Pr(X)$. 그러나 Z를 조건화하면 따라서 $\Pr(X \mid Z \& Y) \neq \Pr(X \mid Z)$, 더 정확히 말하면, $\Pr(X \mid Z \& Y) > \Pr(X \mid Z)$ 이다. 그러나 X, Y, Z에 대해 <그림 2>과 같은 DAG이 확인되었고, 이것은 충실화 조건을 만족시키지 못 한다. (<그림 2>에서 ‘→’는 원인의 영향과 인과 방향을, ‘+’는 원인의 역할 가운데 긍정적, 또는 기여를 가리키고, ‘-’는 원인의 역할 가운데 부정적, 또는 막는 것을 가리킨다. ‘0’은 중립적 원인 영향을 가리킬 것이다.)



<그림 2>

왜냐하면 Z는 X의 조상이지만, $\Pr(X \mid Y) = \Pr(X)$ 라는 관계에 따라

⁶⁾ SGS (1993), pp. 64-69.

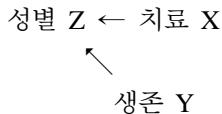
서 Y가 X의 후손은 아닌, 다시 말해 X가 Y의 조상은 아닌 것이기 때문이다. 그와 같은 확률 관계를 따르면 <그림 3>이 되어야 한다. 아래 그림에서 ‘↑’은 무관함을 가리킨다.



<그림 3>

그러나 실제로 주어진 DAG은 <그림 2>이다. 결국 확률적 독립성이 성립하여도, 상대하는 실제 DAG은 그 독립성을 보여주지 못한다. 따라서 충실성 조건에 대한 반례가 된다.

SGS는 이 반례에 대한 응답으로 다음과 같은 DAG <그림 4>를 제시하고 그것으로 충실성 조건이 여전히 성립한다는 것을 보여준다. 그러나 이영의 선생님도 지적하듯이,⁷⁾ 그 수정된 DAG을 납득하기는 쉽지 않다.



<그림 4>

<그림 4>의 DAG이 충실성 조건은 구제한다. $\Pr(X | Z \& Y) = \Pr(X | Z)$ 를 회복하기 때문이다. 그러나 우리의 직관에는 충돌한다. 치료와 생존이 성별에 영향을 주는 원인이 된다고 생각하는 사람을 없을 것 같다.

심슨 역설에 따른 충실성 조건에 대한 문제는 SGS 스스로 진단하였

⁷⁾ 이영의 (2015), pp. 305-12.

고⁸⁾ SGS와 같은 연구프로그램을 지향하는 퍼얼(Pearl 2000)도 그 문제에 주목하였다. SGS와 퍼얼은 충실성 조건을 보존하는 입장에서 그 문제에 대응한다. 반면에 이영의 선생님과 인과 구조 모형에 비판적인 카트라이트(Cartwright 2001[2003])는 충실성 조건에 비판적이다. 카트라이트는 충실성 조건뿐 아니라 마코프 조건도 비판하지만, 이영의 선생님(2005; 2015)은 마코프 조건을 비결정론적 체계 내에서 제한적으로 수용한다.

이미 알려진 것처럼, SGS는 마코프 조건, 충실성 조건에 토대한 베이즈망의 방법론적 유용성을 주장한다. 방법론적 유용성이란 무엇인가? SGS의 주장과 관련하여, 필자는 정당화의 맥락과 발견의 맥락을 구분할 수 있다는 데에 주목할 필요가 있다고 본다. 잘 알려져 있듯이, 이론의 성공은 이론의 목적이 무엇인지에 달려있다. SGS의 인과 모형은 인과의 구조를 효과적으로 발견하거나 추론하는 데에 목적을 두고 있다. 그렇다면 그런 발견이나 추론의 맥락에서 SGS 이론이 수월성을 갖는지 검토할 필요가 있다. 다른 한편으로, 정당화와 발견의 맥락의 구분에 상응하여, 필자는 설명과 예측의 구분에 주목할 필요가 있다고 본다. SGS의 인과 모형의 목적은 설명의 풍부함보다는 예측의 효율성에 초점을 두고 있다고 볼 수 있다. 필자의 이런 생각은 SGS의 또 다른 동료 연구자 켈리(Kelly 2007; 2008; 2010; 2012)의 주장에 토대하고 있다. 켈리는 이론(가설)의 단순성(simplicity)과 단순성의 유용성을 통하여 충실성 조건을 방어할 수 있다고 보는 것 같다. <그림 2>는 <그림 4>에 비해 이론이 증거들에 과잉, 또는 과적합(overfitting)한 경우이다. 다시 말해 하나의 이론(가설)이 보다 더 많은 증거를 만족하는 경우가 된다. 필자는 켈리의 주장이 다음을 의도한 것으로 본다. 심슨의 역설은 예외적으로 특수한 경우일 뿐이며, 일반적으로 많은 경우에 충실성 조건은 이론(가설)의 그런 과잉 적합을 피할 수 있게 하고, 그 점에서 충실성 조건은 방법론적으로 여전히 유용할 수 있다는 것이다. 히치콕(Hitchcock 2010)도 방법론의 관점에서 충실성 조건의 유용성을 언급한다. 이론의 단순성과 이론의 과잉 적합 간에 일종의 맞교환이

8) SGS (1993), pp. 64-69.

있다는 것은 잘 알려져 있다. 이론이 더 적은 수의 증거에 적합하게 되면, 다시 말해 더 단순하다면, 예측력은 높아진다. 그러나 그 댓가로 설명력이나 정보력은 떨어지게 된다. 그와 반대로, 이론이 더 많은 수의 증거에 적합하게 되면, 설명력이나 정보력을 높아진다. 그러나 예측력을 떨어지게 된다.

이영의 선생님은 충실성 조건을 ‘인과에 대한’ 베이즈 이론과 베이즈망에 대한 한 가지 한계로 본 것 같다. 그러나 충실성 조건에 대한 문제 및 비판과 관련하여, 필자는 방법론의 관점에서 충실성 조건이 어떻게 유의미한지를 보겠다. 단순성과 정보력이란 이론 평가의 기준에 상응하는 예측과 설명(해명)은 서로 다른 논리적 구조와 목적이다. SGS 이론, 그리고 이 이론의 충실성 조건은 예측이란 목적에 중점을 두고 있는 반면에, 인과에 대한 기준의 대다수 이론들은 설명(해명)이란 목적에 중점을 두고 있다고 보겠다. 이를 통해 SGS의 이론과 충실성 조건에 대한 비판이 모두 양립할 수 있음을 보겠다. 이 논의를 위해 무엇보다 먼저 심슨 역설에 대한 정확한 이해가 필요하다. 다음 2장에서 심슨 역설의 세 가지 유형을 검토하여 그 역설의 특성에 주목하겠다. 이것으로부터 필자가 논의를 풀어가겠다. 다음으로 마코프 조건, 충실성 조건과 관련하여 SGS 이론이 지향하는 목표와 이런 이론의 배경에 대한 이해가 필요하다. 물론 그런 목표나 배경은 이영의 선생님의 책에 잘 서술되어 있지만, 인과 구조 모형 이론에 관한 최근 연구 동향에 토대하여 3장에서 필자의 해석으로 다시 정리하여 보겠다. 4장에서는 켈리가 발전시킨 오캄의 단순성, 경제성 원리에 토대한 공통 결과(common effect) 인과 구조가, 충실성 조건을 인과 추론이란 방법론으로 어떻게 유의미하고 중요한 역할을 하게 하는지 보겠다. 5장에서는 예측과 설명이 이론의 서로 다른 목적이며, SGS 이론은 대조 실험 등 조정(manipulation, or control)이나 개입(intervention)이 불가능한 비-실험적 상황에서 예측의 효율성에 초점을 두고 있는 방법론적 이론이란 것을 보겠다. 방법론적 관점에서 충실성 조건을 고려해야 하며, 이 관점에서 SGS 이론의 충실성 조건과 이에 대한 비판적 입장이 함께 양립할 수 있다는 것을 보겠다.

2. 심슨 역설의 구조와 의미

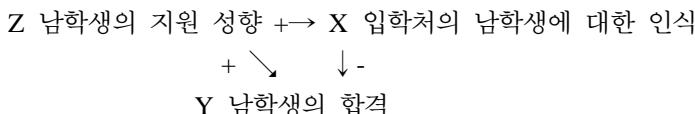
충실성 조건에 대해 심슨 역설이 왜 문제인지를 정확히 이해하기 위해 심슨 역설의 구조와 그것의 의미나 중요성을 다시 확인해 볼 필요가 있다. 심슨 역설은 (이영의 선생님도 언급한) SGS가 제시한 사례 이외에도 몇 가지 잘 알려진 대표적인 사례가 있다. 이들 사례로 세 가지가 있는데, 이들에 대한 분석으로 심슨 역설의 인과 구조를 더 정확히 파악할 수 있을 것이다.

첫 번째 사례는 캘리포니아 대학(UC Berkeley)에서 발생한 사건으로 입학생에 대한 성 차별 논란에 관한 것이다.⁹⁾ 몇 년 간 이 대학에 합격한 신입생들의 성별을 조사하였더니, 여자 신입생이 남자 신입생의 수보다 더 적었다. 이 결과에 대해 학생회는 대학에서 입학을 결정할 때 남녀 성차별이 있었다고 주장하였다. 대학은 이 사안을 조사하고자 진상 위원회를 꾸렸고 입학생 선발에서 성차별이 있었는지 학과 별로 남녀 입학생에 대해 조사를 시작하였다. 진상 위원회의 조사 결과는 다음과 같았다. 학과 별로 입학생을 보았을 때 남자보다 여자의 수가 상당히 적은 학과들이 많았다. 따라서 전체 학생 수만을 보면 남자보다 여자 입학생이 적은 것이 맞다. 그러나 진상 위원회는 또 다른 중요한 사실을 발견하였다. 학과 별로 지원한 학생 수에 상대하여 얼마나 많은 학생이 합격하였는지를 보여주는 학과별 합격률을 조사하였고, 합격률에서는 남녀 성별에 따른 차등이 거의 없는 것으로 나왔다. 합격한 학생 수와 합격률의 이런 차이는 남학생들이 지원 성향에서 비롯되었다. 남학생들은 입학 지원을 할 때 경쟁률이 상대적으로 낮은 학과에 지원하는 성향이 높았고, 상대적으로 경쟁률이 높은 학과에 지원하는 성향은 낮았다. 따라서 남학생들이 많이 지원한 경쟁률이 낮은 학과에서 남학생들의 합격률은 높았다. 이런 탓에 학과 별로 남녀 합격자 수의 차이가 심하였고, 전체 합격자 신입생 수만 보았을 때 남학생 합격자 수가 여학생보다 높게 나타났다. 학과 별로 입학생에 대한

9) Cartwright (1979); Ells (1991), pp. 62-80.

남녀 비율을 볼 때 차별은 없었다. 그러나 학과 전체를 통한 입학생의 남녀 비율은 남자가 여자보다 높게 나왔다. 학생들의 학과별 지원 성향을 고려하지 않고 남녀 입학생 비율을 보았을 때 남자 합격 비율이 여자 비율보다 높으므로 성차별이 있는 것으로 보였다. 그러나 학생들의 학과별 지원 성향을 고려하고 학과 별 남녀 입학생 비율을 보았을 때 남녀 합격자의 비율은 거의 비슷하였다.

경쟁률이 낮은 학과로 지원하는 남학생의 성향을 Z라 하자. X는 대학 입학처가 지원자를 남학생으로 판단하는 것이다. Y는 남학생의 합격(입학)을 가리킨다. X와 Y에 긍정적 상관이 있는 것처럼 보이지만, 사실상 X는 Y에 대해 부정적 원인 효과를 갖는 것으로 나왔다. 다른 한편으로, Z를 고려할 때, Z에서 X를 통해 Y로 향하는 경로가 있게 된다. 이 경로에서 Z는 X에 긍정적 원인 효과를 갖고, X는 Y에 부정적 원인 효과를 갖는다. (다시 말해, 입학처는 남학생에 대한 편향을 갖고 있지 않았다.) 다른 한편으로, Z에서 Y로 직접 향하는 경로가 있다. Z는 Y에 긍정적인 원인 효과를 갖는다. 두 경로 모두를 함께 고려 할 때, Z에서 X를 통해 Y로 향하는 부정적인 인과 경로의 원인 효과와 Z에서 Y로 향하는 긍정적인 경로의 원인 효과는 서로 상쇄된다. 따라서 Z에서 X를 통해 Y로 향한 경로의 원인 효과는 중립화되거나, Z의 Y에 대한 직접적인 (긍정) 원인 효과가 X를 통한 Y에 대한 간접적인 (부정) 원인 효과를 압도하게 된다. 이 때문에 입학생에 대한 남녀 차별이 있는 것처럼 보인다.



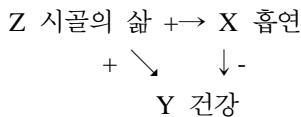
<그림 5>¹⁰⁾

두 번째 사례는 흡연과 폐암의 관계이다.¹¹⁾ 일반적으로 흡연은 폐암

¹⁰⁾ X과 Y의 관계가 상관에 불과할 수도 있다.

에 대해 긍정적인 원인 효과를 가진다. 그런데 어떤 나라에서 흡연은 폐암에 부정적인 원인 효과를 가진다는 놀라운 조사 결과가 나왔다. 보사부에서 이 결과를 확인하기 위해 조사를 하였고 특이 사항을 발견하였다. 도시 거주민과 시골 거주민이 흡연 습관에 차이가 있었다. 도시 거주민은 흡연 뿐 아니라 도시의 대기 오염이 폐암을 유발한다는 것에 크게 걱정한다. 따라서 폐암에 대한 위협이 배가되는 것을 염려하여 흡연 비율이 시골 거주민에 비해 떨어진다. 반면에 시골 거주민은 대기 오염이란 위험 요인에 대해 걱정하지 않으므로 도시 거주민보다 흡연 비율이 상대적으로 높다. 흥미롭게도 시골 거주자들에게서 흡연은 폐암을 억제하는 것으로 나타났다. 또한 이 나라에서는 도시 거주민보다 시골 거주민의 비율이 더 높다. 시골 거주라는 요인을 배제하였을 때 흡연은 폐암에 부정적인 원인 요인으로 확인되었다. 그러나 시골 거주라는 요인이 주어질 때, 이 요인은 다음과 같은 인과 구조를 형성하게 하였다. 시골 거주를 Z, 흡연을 X, 건강을 Y라 하자. 두 가지 경로가 있다. 첫 번째 경로는 Z로부터 X를 통해 Y로 향하는 경로다. Z에서 X로 향하는 경로에서 Z는 X에 긍정적인 원인 효과를 갖고, X는 Y에 대해 부정적인 원인 효과를 갖는다. 반면에, Z는 Y에 긍정적인 원인 효과를, 다시 말해 Z는 Y를 강화한다. Z에서 Y로 직접 향한 인과 경로에서 Z의 Y에 대한 긍정적 원인 효과는 Z에서 X를 통해 Y로 향한 경로에서 X의 Y에 대한 부정적인 인과 효과를 상쇄하게 된다. 따라서 Z에서 X를 통해 Y로 향한 경로에서 X의 Y에 대한 원인 효과는 중립적이 되거나, 또는 Z의 Y에 대한 직접적인 (긍정) 원인 효과가 X를 통한 Y에 대한 간접적인 (부정) 원인 효과를 압도하게 된다. 따라서 흡연이 건강을 강화하고 폐암을 억제하는 것으로 나타나게 된다.

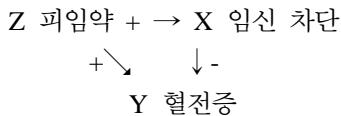
¹¹⁾ Skyrms (1980); Eells (1991), pp. 62-80.



<그림 6>

세 번째 사례는 잘 알려진 특정 피임약과 혈전증의 관계이다.¹²⁾ 특정 피임약은 혈전증에 긍정적인 원인 효과를 가진다. 그러나 특정 피임약이 혈전증에 오히려 부정적인 원인 효과를 갖는, 다시 말해 혈전증을 막는다는 조사 결과가 나왔다. 이 사실을 확인하기 위해 보사부는 진상 조사를 하였고 특히 사항을 발견하였다. 임신도 혈전증에 긍정적인 원인 효과를 가지는데, 그 피임약이 임신율을 떨어뜨려서 혈전증의 발병률을 낮추었다는 것이다. 두 가지 경로를 생각할 수 있다. 피임약을 복용하는 게 혈전증의 발병률을 높이는 경로이다. 다른 한편으로, 피임약을 복용하여 임신율이 떨어져서 (다시 말해, 임신 차단률이 높아져서) 혈전증의 발병률이 떨어지는 경로이다. 두 번째 경로를 고려하지 않았을 때, 처음 조사에서는 피임약 복용은 혈전증의 발병률을 높이는 것으로 나타났다. 그러나 진상 조사에서 두 번째 경로를 함께 고려하였을 때, 피임약 복용으로 인한 임신율 하락이 혈전증을 낮추는 것으로 나왔다. 두 번째 경로가 첫 번째 경로의 인과 효과를 상쇄하거나 압도하는 것으로 나타났다. 피임약 복용을 Z, 혈전증 발생을 Y라 하자. 임신 차단을 X라 하자. Z에서 Y로 향하는 직접적인 경로에서 Z는 Y에 대해 긍정적 원인 효과를 가진다. 반면에, Z에서 Y로 향하는 간접적인 경로에서 Z는 X에 대해 긍정적 원인 효과를 갖고, 다시 X는 Y에 대해 부정적 원인 효과를 가진다. 두 번째 경로가 첫 번째 경로를 상쇄하게 될 때, Z의 Y에 대한 원인 효과는 중립적으로 되거나, 두 번째 경로가 첫 번째 경로를 압도하여 Z는 Y에 대해 부정적인 원인 효과를 가진 것으로 나타나게 된다.

¹²⁾ Hesslow (1976); Eells (1991), pp. 173-97.



<그림 7>

심슨 역설의 이들 세 가지 사례 각각은 인과의 이행성 등 인과의 여러 문제와 관련하여 많은 논의를 필요로 하는 문제다. 여기에서는 이들 사례가 심슨 역설에 핵심이 될 요소를 함께 하고 있다는 데에 주목할 필요가 있다. 결과 요인에 도달하는 두 가지 경로가 있고, 하나의 인과 경로에서 원인 효과가 다른 인과 경로에서 원인 효과에 의해 상쇄, 또는 차단하고 있는 구조이다. 이처럼 상쇄하는 역할을 하는 인과 경로를 ‘뒷문(back-door)’ 경로로 부르기도 한다.¹³⁾ 이영의 선생님도 밝혔듯이, 심슨 역설에 대한 가장 표준적인 해결 방법은 전체 집단과 이 집단의 부분 집단을 구분하여, 이들 집단 각각에 상대하여 한 요인이 또 다른 요인에 대해 갖는 원인 효과를 해명하는 것이다.¹⁴⁾ 또 다른 최근에 제시된 방식은, 직접적인 인과 경로의 원인 효과를 상쇄하는 역할을 하는 뒷문(간접) 경로의 매개 요인을 차단하여 직접적인 인과 경로의 원인 효과를 보여주는 것이다.¹⁵⁾ 이들 두 방법의 관계와 차이에 대해서는 다른 곳에서 자세히 논의할 필요가 있다.¹⁶⁾ 이 글에서는 이영의 선생님이 비판한 SGS의 해결 방안에 대해 초점을 두고 논의하여 보겠다.

¹³⁾ Pearl (2000), pp. 79-81.

¹⁴⁾ Eells (1991).

¹⁵⁾ Hitchcock (2001).

¹⁶⁾ 김준성 (2015), (2016).

3. 심슨 역설에서 마코프 조건과 충실성 조건의 의미

마코프 조건, 최소성 조건, 충실성 조건이 SGS 인과 모형 이론 등 베이지안 인과 이론에서 갖는 의미가 무엇인지 보자. 흄이 주장하였고 많은 사람들이 동조하듯이, 인과성 개념은 하나의 미궁이다. 이런 형이 상학적 개념을 언급하거나 이론의 토대로 삼지 않으려는 것은 논리실증주의자 등 경험주의 전통에 동조하는 모든 사람들에게 크게 다르지 않다. 글리모어(Glymour 1980) 스스로 논리경험주의의 계승자라고 선언하였듯이, SGS는 인과성 개념을 전제하지 않고 경험적, 또는 실증적 전통 내에서 인과에 대한 구조를 추론하려는 것이다. 이런 맥락에서 마코프 조건은 중요하다. 마코프 조건은 인과에 대한 어떤 언급도 없이 인과성이 갖는 특성을 보여준다. 그러나 여러 반례가 (Hausman & Woodward 1999) 이 조건에 제기되었고, 마코프 조건의 역 관계인 충실성 조건에 대해서도 다르지 않다. 반례나 비판들은, 이를 조건을 벗어나는 인과 구조의 사례들에 토대한다. 마코프 조건에서 핵심 개념은 확률적 독립성이다. 이 확률적 돋립성은, 결과 변수의 직접적인 원인으로서 변수만을 주목하게 하는 제약 조건이다. 이런 직접적인 원인 변수를 차례로 발견하고 이를 통해 차례로 단계적으로 인과의 전체적인 구조, 또는 망을 구성하여 세계의 인과 구조의 진상(眞相)에 접근하는 것이다. 이런 방식이 아니라 서로 얹혀 있는 인과의 망을 전제하고 또 다른 원인을 찾는 단계로 나아간다면, 인과의 이행성(transitivity) 등 여러 문제가 될 인과 개념들이 개입하게 되어, 시작 단계부터 한계에 부딪히게 될 것이다. 또한 인과의 항상성(invariance)를 보여주는 데에도 마코프 조건의 직접적인 최소 원인 변수만을 고려하는 것은 효과적이다. 일종의 세테리스 패리부스(ceteris paribus) 조건의 역할을 한다고 볼 수 있다.¹⁷⁾ 또한 이론적인 가설이 그것의 효과를 보여주고자 여러 제약 조건을 두는 것과 유사하다. SGS에게는 더 큰 이유가 있다. 앞서 언급하였듯이 방법론적 효율성이다. 가능한 한 단순

¹⁷⁾ Pearl (2000), pp. 48-49.

화 인과 구조에서 또 다른 인과 구조로 확장하는 과정에서, 그 시작부터 복잡하다면 그 복잡성을 해결하기 위한 구조의 확장으로 가는 것은 쉽지 않게 된다. 가설의 단순성과 관련하여, 정보력이나 설명력의 크기와 예측의 효율성의 맞교환은 이를 잘 보여준다.

심슨 역설이 충실성 조건을 위반하는 것은 다음과 같이 이해할 수 있다. 충실성 조건에 따른 해당 변수들 사이에서, 결과의 원인이 되는 모든 독립, 또는 외생(exogeneous) 변수들은 상호 간에 확률적 독립성이 마코프 조건에 따라 요구되어진다. 다시 말해, 마코프 조건에 따르면 결과에 인과적으로 관련된 원인 변수들이 있을 때 이들 변수의 결과가 아닌 변수들은 당연히 그 원인 변수들과 서로 인과적으로 관련되지 않기를 요구하는 것이다. 그러나 인과적으로 관련되는 경우가 발생하는 데, 원인 변수들의 부모 변수가 또 다른 경로를 통해 그 결과 변수에 인과적으로 연결되는 경우이다. 다시 말해 원인 변수의 부모 변수가 우회 경로를 통해 결과 변수에 인과적으로 연결되는 것이다. 앞서 보았듯이, 퍼얼(Pearl 2000)은 이것을 뒷문(back-door)으로 부른다. 심슨 역설에 대한 해결 방식은 모든 이론에 공통적으로 같은 방식으로 일반화되어 있다. 결과 변수에 우회적으로 인과로 연결되고 해당 원인 변수와 또한 인과적으로 연결된 그 독립 변수를 배경 맥락으로 고정하여, 해당 원인 변수와 결과 변수의 관계를 보는 것이다. 또한 다음과 같이 말할 수 있다. 그렇게 두 경로를 통해 결과 변수에 연결된 그 독립 변수를 실현하는 집단에 상대하여, 해당 원인 변수와 결과 변수의 관계를 보는 것이다. SGS는 심슨의 역설이 충실성 조건을 충족하지 못 하는 것에 대해, 그 충족을 실현하기 위해 해당 독립(외생) 변수와 결과 변수가 이들의 원인으로 연결된 독립 변수를 오히려 이들의 결과로서의 독립 변수로 연결하는 것이다.

필자의 생각에는 SGS가 (분명하게 보여주지는 않지만) 보다 중요한 문제를 고려하는 것으로 본다. SGS는 인과 이행성에 관한 인과 경로를 피하고 싶은 것이며, 이를 통해 인과 이행성의 문제를 다루고 싶지 않은 것이다. 그렇다면 두 가지 해석이 가능하다. 첫째, SGS는 인과 이행성을 인정하지 않는다고 볼 수 있다. 둘째, SGS는 인과 이행성이

문제가 많은 개념이므로, 이를 전제하지 않고 인과 구조에 대한 추론을 진행하고 싶다. SGS의 주장에 대한 필자의 또 다른 생각은 심슨 역설의 구조를 인과 추론을 위해 필요한 인과 구조의 기본 모듈로 수용하기를 원하지 않는 것으로 보인다. 심슨 역설의 인과 구조를 인과 구조의 기본 모듈로 수용한다면, 그 보다 더 복합적으로 상호 연결된 인과 구조도 수용해야 하는 것이고, 따라서 모든 인과 구조들 간의 경계가 불분명해지는 모호성 문제가 발생한다. 결국, 이 세계의 모든 것이 상호 연결되어 있다는 주장에 이르고, 이는 무엇도 주장하는 바와 다르지 않게 되며, 새로운 인과 구조에 대한 탐색(추론, 발견)을 시작 단계에서 막는 게 될 수도 있다. 이 사안에 대해 켈리(Kelly 2007; 2008; 2010; 2012)는 오캄(Ockham)의 면도날, 다시 말해 오캄의 단순성, 경제성 원리로 SGS의 이론이 전제하는 충실성 조건을 정당화 한다. 다음 장에서 그 정당화에 대해 논의하겠다.

4. 인과 구조에서 오캄의 단순성(경제성) 원리의 역할

베이즈망을 구성할 DAG의 기본이 되는 인과 구조를 다시 보자. X, Y, W는 변수다. $X \rightarrow Y$ 는 X는 Y의 직접 원인이 된다. 이행적인 연쇄 인과 고리 $W \rightarrow Y \rightarrow X$ 에서, W는 Y에 관한 정보를 주는 것을 통해서만 X에 관한 정보를 줄 것이다. 따라서 X에 관하여 우리가 아는 바는, Y가 W에 관하여 주는 정보가 전부다. 다시 말해, Y로 조건화 할 때, X는 W로부터 독립적이다. $Pr(X | W \& Y) = Pr(X | Y)$. 역 연쇄 인과 고리 $X \leftarrow Y \leftarrow W$ 에도 같은 논리가 적용된다. 공통 원인의 상황 $W \leftarrow Y \rightarrow X$ 에도 마찬가지다. 이들 세 상황 모두는 같은 조건화 관계가 된다. Y로 조건화할 때, W는 X로부터 독립적이다.

이번에는 공통 결과의 상황 $W \rightarrow Y \leftarrow X$ 를 보자. 이 상황에서 W만으로는 X에 관한 어떤 정보도 주지 않는다. 두 변수 W와 X는 인과적으로 독립적이기 때문이다. 그러나 Y로 조건화할 때, 변수 X는 W에 관한 어떤 정보를 준다. 다시 말해 Y로 조건화할 때, 변수 X는 W

에 독립적이지 않게 된다. $\text{Pr}(X | W \& Y) \neq \text{Pr}(X | Y)$. (또한 $(\text{Pr}(X | W \& Y) = \text{Pr}(X | W))$ 이므로, 충실성 조건은 만족된다.) W와 X는 특정한 방식으로 Y에서 관찰된 값을 생성하는 데에 공조하기 때문이다. 예를 들어 어떤 사람이 알칼리 액과 산성 액을 모두 마셨을 때 (두 성분의 결합으로 중성 액이 되어) 그는 살아남는다. 물론 둘 중 하나만 마셨다면 죽을 것이다. 따라서, 알칼리 액을 마시는 것과 생존의 주어질 때 산성 액을 마시는 것에 대한 우리의 평가와, 산성 액을 마시는 것이 주어질 때 생존하는 것에 대한 우리의 평가는 다를 것이다.

그와 같이 공통 결과 구조는 앞서 세 가지 구조와 구분된다. 공통 결과의 인과 구조는, 주어진 정보로부터 관찰가능한 조건부 (독립이 아닌) 의존 관계에 의해 결정될 수 있다. 공통 결과 구조는 심슨 역설이 의도한 바를 보여줄 수 있다. 예를 들어, Z를 공통 결과 Y의 직접적인 결과라고 하자. $Y \rightarrow Z$ 가 Y와 Z의 간접된 공통 원인 C에 의해 실제로 산출된다고 하자. 이 경우에 공통 결과의 상황을 적용하면, W는, Y와 C가 주어질 때 공통 원인 C에 관한 정보를 주게 된다. (그림 8] 참고) 마치 Y의 직접 원인인 X가 Y에 관한 정보를 주는 것처럼, 혼재된 인과 상황에서 동질적인 의존성이 나온다. 다른 한편, W는 Y만 주어진 경우에 (Y만 조건화할 때) Z로부터 독립적이다. Y가 W에 관한 정보를 산출하는 것에 의해서만 Z에게 W에 관한 정보를 생성하기 때문이다. 따라서 대조 실험을 할 수 없는, 또는 의존 변수(결과)의 변화를 보기 위한 조정(manipulation) 등이 불가능한 비-실험적 연구의 정보를 분석하여, 혼재된 인과 상황에서 동질적인 원인 효과들에 관한 인과적 결론을 내리는 게 가능하다.¹⁸⁾

$$\begin{array}{c}
 W \rightarrow Y \leftarrow X \\
 \downarrow \swarrow \\
 Z \leftarrow C
 \end{array}$$

<그림 8>

¹⁸⁾ Kelly (2010), pp. 18-20.

관심의 대상인 일부 변수를 연결하는 참된 인과 이론이 주어질 때 그리고 그 이론의 자유 모수에 대한 엄밀한 추정이 주어질 때, 우리는 자연 규칙에 따라 엄밀한 반사실적 예측을 얻을 수 있다. 변수 X 에 대한 간섭(intervening)의 결과를 예측하기 위해, 그 변수에 값 x 를 부여하자. 우선, X 로 향하는 모든 인과 회살을 지우자. 이 때, 다른 이론의 모수들에 대해서는 이전 값들로 고정하고 (다시 말해, 결과로서 의존 변수 Y 에 영향을 줄, 그리고 독립(외생) 변수 X 의 결과가 아닌 모든 변수들의 값을 배경 조건으로 고정하고), 관심의 대상인 변수 Y 의 값을 예측하자. 예를 들어, X 가 결과이고, X 에 어떤 값이 주어지면, X 와 다른 변수들 간의 모든 연결은 단절될 것이다. 따라서 다른 변수들의 값에는 변화가 없는 것으로 예측될 것이다. 반면에 X 가 원인이고, X 에 어떤 값이 주어지면, X 의 결과의 값들은 바뀔 것이다. 그와 같은 반사실적 예측에서 그 예측이 엄밀한 것이 되려면, 그 엄밀성은 변수들 간의 참인 인과 관계에 상응하는 인과 모형을 추론하는 데에 달려 있다. 그렇다면 비-실험적 정보에서 인과 구조를 발견하는 것은 오캄의 면도날(Ockham's razor), 또는 경제성이나 단순성 원리에 달려있게 된다.

인과 구조에 대한 발견은 조건부 상관들이 보여주는 패턴의 양상에서 읽어내는 것이다. 그 경우에 조건부 상관을 0으로 가정하는 경향이 있다면, 그 경향은 바로 단순성 원리를 따른 것이다. SGS에게는 마코프 조건과 충실성 조건이 그 역할을 하게 된다. 0이 되는 조건부 상관이란, 이미 보았듯이 마코프 조건과 충실성 조건이 논리적으로 합축하는 확률의 독립성을 말한다. 인과 구조를 추론하는 데에 왜 단순성 원리에 부합하는 경향을 갖는 게 필요한지는 다음과 같다. 0이 아닌 조건부 상관이란, 자유 모수들(parameters)(다시 말해 결과에 변화를 줄 또 다른 변수들)이 있다는 것이다. 우리가 예측을 하려면 이들 자유 모수를 추정해야 한다. 단순성 원리에 따르면 우리는 가능한 한 그런 자유 모수가 적은 성향을 취해야 한다. 단순성 원리에 어긋나면 어떻게 되는가? 그처럼 보다 적은 자유 모수들로 이루어진 인과 이론이 없다면, 인과에 대한 유의미한 결론을 얻지 못한다고 본다. 많은 자유 모수

들로 이루어진 가장 복잡한 인과 이론이 있다고 가정하자. 이 이론은 가장 촘촘한 인과망을 갖고 있으므로 (비-실험적) 정보에 가장 잘 적합(fit)할 수 있다. 그러나 처음부터 그런 복잡한 인과 이론으로 인과 추론을 시작한다면, 하나의 경로에서 변수 X가 또 다른 변수 Y에 대해 갖는 원인 효과는 (심슨 역설에서 볼 수 있듯이) 또 다른 경로에서 같은 변수 X가 또 다른 변수(자유 모수) Z를 통해 Y에 대해 갖는 원인 효과에 의해 상쇄될 수 있기 때문이다. 이처럼 상쇄된 결과를 전제하고 인과 예측을 하거나 또 다른 인과 구조로 추론이 진행되면 잘못된 예측이나 왜곡된 인과 구조로 추론할 가능성은 매우 높을 것이다. 결국 정보에 가장 잘 적합한 인과 이론이란 정보에 과잉 적합하게 되고, 그런 이론을 출발점으로 삼을 때 (비-실험적 상황에서) 예측의 효율성이나 추론의 정당성을 보장받기는 어렵다.

치아에 쌓인 니코틴의 양에서 폐암의 발생을 염밀하게 추정(estimate)하려는 경우를 보자. 이들 두 요인 간에 강한 통계적 상관이 발견되었다고 하자.¹⁹⁾ 그렇게 추정된 상관은 흡연의 폐암에 대한 원인 효과를 확인하는 데에 이용될 수 있다. 이를 위해 그 통계적 상관과 관련하여 가능한 한 많은 정보에 부합하는 과잉 적합 모형을 사용한다고 가정하자. 과잉 적합이므로 칫솔질 보조금이 폐암 감소에 대해 갖는 원인 효과도 확인한다고 하자. 그렇다면, 앞 서 추정된 (치아의 니코틴 양과 폐암의) 상관으로 칫솔질 보조금이 폐암 감소에 갖는 원인 효과를 염밀하게 예측하기는 어려울 것이다. 왜냐하면 그 보조금 정책이 앞서 추정된 (치아의 니코틴 양과 폐암의) 상관을 무관하게 할 정도로 표본 분포를 변화시킬 것이기 때문이다. 이 문제는 그런 표본에 기초한 인과 이론을 오류로 만들게 되어, 실제 집단에 관한 염밀한 예측도 쓸모없게 만들 수 있다. 정리하면, 표본 집단을 바꾸는 새로운 정책은 그 정책이 영향을 줄 (반사실적) 결과를 예측하는 게 무의미하게 된다. 과잉 적합 모형의 문제를 보여준다.

¹⁹⁾ Ibid.

5. 이영의 선생님의 비판에 대해 SGS의 가능한 응답: 예측과 설명의 구분에 따른 집단 상대성에 따른 문제 해결과 SGS 조건의 양립

SGS가 세 가지(마코프, 충실성, 최소성) 조건을 인과 구조를 정당화하는 절대적인 조건으로 보는지는 분명하지 않다. SGS의 주장은 보다 완곡하다고 볼 수 있다. SGS는 사회 과학 등을 조사한 결과 이들 세 조건이 충족되는 경우가 일반적이고, 따라서 주어진 정보를 간의 상관으로 인과 구조를 추론할 때 이들 조건이 최소 제약으로 주어질 필요가 있다고 보는 것 같다. 이들 세 조건이 충분하지는 않지만 필요하다는 입장이다. 심슨의 역설을 충실성 조건의 반례가 되는 특수한 경우로 제시하는 데에서도 짐작할 수 있다. SGS는 심슨 역설과 이것이 어떻게 충실성 조건에 대한 반례가 되는지 소개한 후, 페얼이 제시한 두 가지 DAG을 비교하면서 심슨 역설의 경우에 충실성 조건을 위반하지 않으려면 공통 결과 구조의 DAG을 받아들여야 한다고 주장한다.²⁰⁾ SGS는 선행 지식의 차이에 따라 DAG에 대한 결정이 달라진다고 주장한다. 심슨의 역설에 대한 배경 지식이 있다면 첫 번째 DAG을 주목해야 하지만, 그렇지 않다면 두 번째 DAG을 주목해야 한다는 것이다. 이는 SGS가 인과 구조의 해명이나 정당화보다는 발견이나 추론에 큰 비중을 두고 있음을 보여준다.

여기서 배경(사전) 지식이란 무엇인가? 무엇이든 될 수도 있지만, 심슨 역설에서는 원인 효과가 어떤 집단에 상대하는지에 따라 달라지는 것에 주목해야 한다.²¹⁾ 따라서 집단 전체에 대한 정보와 그 전체 집단의 부분(하부) 집단(sub-population)에 대한 정보가 중요하다. 또한 이들 부분 집단 가운데 어떤 집단에 상대할 것인지에 관한 정보가 중요하다. 그러나 SGS는 이들 정보가 분명하지 않거나 매우 제한적이라는 전제 하에서, 극단적인 상황에서는 그런 정보가 사실상 없다는 전제

²⁰⁾ SGS (1993), p. 68; 이영의 (2015), p. 306.

²¹⁾ Eells (1991), 김준성 (2008).

하에서 인과 구조에 대한 추론을 할 때 과연 어떤 DAG이 유용한지를 보여주려는 것 같다. 또한 중요한 점은, 대조 실험을 할 수 없는, 또는 의존 변수(결과 요인)의 변화를 보기 위한 조정(manipulation, or control) 등이 불가능한 비-실험적(non-experimental) 연구에 대한 정보에서 인과 추론을 한다는 것이다. 실험적 연구와 다르게, 비-실험적 연구에서 인과를 추론한다는 것은 매우 어렵다.

필자는 설명과 예측이란 두 가지 구분된 목적을 토대로 SGS의 주장과 SGS에 대한 비판도 함께 수용할 수 있다고 본다. 과학 이론의 성공은 어떤 목적을 전제하는지에 달려 있고, 그 목적이란 단순성, 정보력, 설명력, 예측력, 논리적 염밀성 등 다양하다. 이들 목적 모두를 함께 충족하는 게 어렵다는 논의는 이미 일반적이다. 단순성과 정보력 간에 맞교환이 있고 이들 각각에 상응하여 예측력과 설명력 간에 맞교환이 있다. 다시 말해 한 가지 목적을 실현하는 과정에서 다른 목적에 대한 희생이 불가피하다. SGS 이론과 (심슨 역설과 관련하여) 그 이론을 비판하는 입장에서 중점을 두는 목표가 다르다고 볼 수 있다. SGS 이론이 예측과 발견에 중점을 두고 있다면, 이 이론에 비판적 입장은 설명(해명)과 정당화에 중점을 둔다고 볼 수 있다. 예측과 설명이 다르다는 것은 예측과 설명의 논리적 구조가 대칭적이지 않다는, 다시 말해 같지 않다는 것에서 잘 알 수 있다.²²⁾ 예를 들어 일어나기 매우 어려운 사건이지만 매우 낮은 확률 관계로도 그 사건에 대한 예측을 말할 수 있다. 반면에, 그처럼 불확실한 사건이 발생하였을 때 왜 발생하였는지를 그렇게 낮은 확률 관계로 설명하지는 않는다.

이미 앞서 보았듯이 SGS 이론은 인과 추론(예측)의 효율성에 목표를 두고 인과 구조의 단순성을 지향한다. 따라서 심슨 역설 유형의 구조와 관련하여 충실성 조건을 만족할 공통 결과의 인과 구조를 제시한다. 경험적으로 주어진 정보만을 토대로 예측을 하는 과정에서, 심슨 역설처럼 다양한 인과 경로가 서로 상쇄하는 구조는 정보에 과잉 적합한 것이고, 따라서 그런 구조는 잘못된 예측으로 이끌 수 있다. 정보가 과잉 적합한 경우에 설명력과 해명의 정도는 떨어질 수 있다.²³⁾ 따라

22) 이영의 선생님(2015, pp. 140-42)도 설명과 예측의 비대칭성을 언급한다.

서 치료와 환자의 회복에 관한 심슨의 역설에서, 치료나 환자의 회복이 성별에 영향을 준다는 추론을 배제할 수 없게 된다. 이 추론이 설명이라면 그것은 분명히 잘못된 설명이다. 그러나 우리가 치료를 받은 환자나 회복된 환자로부터 이들의 성별을 예측하고자 한다면 그 예측이 기이한 것만은 아니다. 흥미롭게도 시골 생활, 흡연, 폐암에 관한 심슨 역설은 예측과 관련하여 인과의 일방성에 크게 제약을 받지 않는 구조를 보여준다. 흡연이 시골 생활의 원인이 되고 건강이 시골 생활의 원인이 된다. 그런데 강한 흡연 습관자가 시골 생활자가 될 것이라 예측하거나, 건강한 자가 시골 생활자가 될 것이라 예측하는 것은 자연스럽다. 다른 한편으로, 흡연 습관이 시골에서 생활하게 하는 영향을 주었고, 건강한 습관이 시골에서 생활하는 데에 영향을 주었다고 설명하는 것은, 인과의 일방성을 고려하여도 그렇게 이상한 설명은 아니다. 변수들 간의 인과 일방성에 관한 정보도 인과 구조를 통해 추론되고 수정 될 대상이다. 예측의 효율성을 극대화하는 데에 단순성, 경제성의 원리가 전제되고, 그 관점에서 충실성 조건이 유지될 수 있게 된다.

앞 장에서 보았듯이 SGS는 인과 구조의 주어진 변수(요인)들 간의 관계를 통해 또 다른 독립된 변수(요인)에 대한 정보를 얻을 수 있는 것에 주목한다. SGS에게는 그런 정보를 얻는 데에 있어서 인과의 일방성이나 방향성(asymmetry, temporal priority)을 늘 고정된 것으로 확정하는 게 우선은 아닌 것 같다. SGS에게서 인과의 일방성이 또 다른 변수에 따라 달라질 수 있는 가능성은 늘 열려 있는 문제다. 같은 정보로부터 동일한 변수들을 얻게 되고, 이들 변수 간에 인과 방향성이 서로 다른 여러 모형들이 서로 양립 가능하다. 이 경우에 어떤 모형을 최선으로 선택할지에 관한 비결정성 문제가 발생한다. 그러나 이 문제는 SGS의 연구 프로그램을 제고하게 하는 원동력이 된다.

23) 필자는 여기서 설명력과 정보력 간의 맞교환(trade-off)은 논의하지 않았다. 이에 대한 자세한 논의는 여러 문헌에서 확인할 수 있으며 특별히 주목할 만한 연구는 Forster & Sober (1994)이다. 설명력과 정보력의 관계나 이들의 맞교환 여부는 여전히 논쟁거리다. 그러나 그 맞교환을 확인하는 것은 어렵지 않다.

설명이나 해명의 관점에서 볼 때, 인과 구조가 가능한 한 많은 정보에 적합하다는 것은 중요하다. 제한된 정보가 제한된 설명을 주는 것은 당연하다. 그러나 과다한 정보가 정말로 좋은 설명을 제공할 수 있을지도 논쟁거리가 될 수 있다. 어찌되었든 인과에 대한 기준의 대다수 이론은 인과 구조를 해명하는 데에 초점을 두고 있다. 그리고 이들 이론에 반례나 문제가 될 인과 구조가 제시되고 이를 반례나 문제를 해결하는 데에 중점을 두어 왔다. 심슨 역설의 경우에 이영의 선생님도 밝혔듯이 일반적인 해결 방법은 집단에 대한 구분과 이처럼 구분된 집단에 상대하여 인과 구조를 밝히는 것이다. 예를 들어 전체 집단에서 흡연은 폐암에 긍정적인 원인 역할로 해명되지만, 시골 생활자로 이루어진 부분 집단에서 흡연은 폐암에 부정적인 원인 역할로 해명된다. 집단이 구체화될수록 부분집단에 대한 정보는 점점 더 많아지고 이들 각 정보에 상대하여 인과 구조에 대한 해명은 보다 더 정교하게 된다. 다시 말해 설명의 정도가 점점 더 높아지는 것이다. 그러나 심슨 역설처럼 다수의 인과 경로가 서로를 상쇄(간섭)하는 경우에 설명력이 높은 구조가 예측의 효율성을 떨어뜨리는 경우를 어렵지 않게 생각해 볼 수 있다. 피임약이 혈전증에 긍정적 원인 효과를 갖지만, 부정적이거나 중립적인 원인 효과를 갖는 경우를 보자. 피임약이 혈전증의 긍정적 원인이 될 수 있지만, 피임약이 임신율을 떨어뜨려 혈전증 발생을 낮추는 경로는 피임약이 혈전증에 대한 긍정적 원인을 상쇄한다. 혈전증과 관련하여, 피임약은 임신과 인과 상호작용을 하는 것이고, 피임약은 혈전증에 대해 혼재된(mixed) 원인 역할을 하게 된다. 혼재된 원인 역할로서 피임약에 대한 설명은 충분하지만, 피임약의 그 혼재된 역할이 혈전증에 대해 긍정적, 부정적, 중립적 역할 중 어떤 결과를 초래할지를 예측하기는 쉽지 않다. 정보에 많이 적합된 인과 구조에서 예측의 정확성이 떨어지는 것을 보여준다. 또한 임신율의 하락이 계속 더 강하게 유지되는 한, 혈전증 발생에 피임약 복용이 긍정적인 원인 효과를 가진다는 것에 관한 어떤 정보도 얻지 못 할 수도 있다.

같은 맥락에서 인과의 이행적인(transitive) 구조를 보자. 서울에서 전화를 걸기 위해 전화기 버튼을 누르면, 부산에 있는 상대방의 전화

가 울리게 되고, 전화를 받게 된다. 그러나 대전에 있는 다른 사람이 전화기 버튼을 누르고, 부산에 있는 사람의 전화가 울려, 그 사람이 전화를 받는 경우에, 서울에서 전화는 차단된다. 인과의 이행성이 유지되지 않는 경우다. 인과의 이행성은 여전히 논쟁거리인데, 이행성을 보여줄 한 가지 방식은, 이행 구조에서 중간에 매개하는 요인의 원인이 될 요인들을 배경 조건으로 모두 고정하는 것이다. 대전에서 전화걸기를 초래하는 요인을 배경 조건으로 고정하는 것이다.²⁴⁾ 따라서 그 배경 조건에 따라 인과의 이행성 여부를 결정할 수 있게 된다. 마찬가지로 이행성에 대한 이런 인과 구조는 정보에 과 적합한 경우가 된다. 이행성을 설명하는 데에는 탁월하지만, 서울에서 전화가 어떤 원인 역할을 할 수 있을지는 예측하기 어렵다. 또한 대전에서 전화걸기가 계속 유지되는 한, 부산에서 전화 응답에 대해 서울에서 전화 걸기가 가질 원인 효과에 관한 어떤 정보도 얻을 수 없게 된다.

6. 나가는 말

이영의 선생님 등 여러 학자들이 심슨 역설을 SGS 이론의 충실성 조건에 대한 반례로서 주목을 하였다. SGS도 그 문제를 인식하고 그에 대해 논의하였지만, SGS 이론의 목적과 의미를 충분히 고려하지는 못 하였던 것 같다. 켈리(Kelly 2007; 2008; 2010; 2012)가 발전시킨 단순성 원리에 토대하여, 필자는 SGS의 이론을 예측의 효율성이란 방법론적 관점에서 여전히 유망한 연구프로그램이 되며, 그와 관련하여 충실성 조건이 의미를 가질 수 있다고 보았다. 모든 베이즈 이론의 핵심은 조건부 확률의 개념이다. 같은 맥락에서 베이즈 이론 자체도 그런 개념의 대상이라면, SGS의 인과 구조 이론은 그런 개념이 보다 바람직한 방향으로 진행될 수 있다고 본다. 충실성 조건도 그런 맥락에서 볼 필요가 있다. 이영의 선생님의 비판도 그런 바람직한 방향을 위

²⁴⁾ Eells & Sober (1983), Eells (1991), 김준성 (2008).

한 하나의 제안일 것이라 생각한다. 따라서 이영의 선생님도 베이즈주의와 발을 맞추고 있는 것이며, 『베이즈주의: 합리성으로부터 객관성으로의 여정』도 그 간신의 한 단계로 역할을 할 것이다.

참고문헌

- 김준성 (2008), 『확률과 인과』, 아카넷.
- _____ (2015), 「이행적 인과 경로를 통한 원인 효과에 대한 해명: 구조 방정식에 토대한 인과 모형의 원인 효과 개념에 대한 평가와 대안」, 『논리연구』 18권 1호, pp. 83-133.
- _____ (2016), 「거시 계량경제학의 인과 모형에서 딜레마와 이에 대한 해결 방안」, 『인문과학논총』 37권 1호, pp. 429-49.
- 이영의 (2005), 「인과적 마코프 조건과 비결정론적 세계」, 『논리연구』 8권 1호, pp. 47-67.
- _____ (2015), 『베이즈주의: 합리성으로부터 객관성으로의 여정』, 한국문화사.
- Cartwright, N. (1979), “Causal Laws and Effective Strategies”, *Nous* 13: pp. 419-37.
- _____ (2001), “What is Wrong with Bayes Nets?”, *The Monist*, 84(2): pp. 242-64. Reprinted in Kyburg, H. E. & Thalos, M. (eds.) (2003), *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, Open Court Publishing.
- Eells, E. (1991), *Probabilistic Causality*, Cambridge University Press.
- Eells, E. & Sober, E. (1983), “Probabilistic Causality and the Question of Transitivity”, *Philosophy of science* 50: pp. 35-57.
- Forster, M. & Sober, E. (1994). “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions”, *The British Journal for the Philosophy of Science* 45(1): pp. 1-35.
- Glymour, C. (1980), *Theory and Evidence*, Princeton University Press.
- Halpern, J. Y. & Pearl, J. (2005), “Causes and Explanations: A Structural-model Approach. Part I: Causes”, *The British Journal for the Philosophy of Science* 56(4): pp. 843-87.

- Hausman, D. M. & Woodward, J. (1999), "Independence, Invariance and the Causal Markov Condition", *The British Journal for the Philosophy of Science* 50(4): pp. 521-83.
- Hesslow, G. (1976), "Two Notes on the Probabilistic Approach to Causality", *Philosophy of Science* 43(2): pp. 290-92.
- Hitchcock, C. (2001), "A Tale of Two Effects", *The Philosophical Review* 110(3): pp. 361-96.
- Hitchcock, H. (2010), "Probabilistic Causation", in *The Stanford Encyclopedia of Philosophy*.
- Hoover, K. D. (2001), *Causality in Macroeconomics*, Cambridge University Press.
- Kelly, K. (2007), "Ockham's Razor, Empirical Complexity, and Truth-Finding Efficiency", *Theoretical Computer Science* 383: pp. 270-89.
- ____ (2008), "Ockham's Razor, Truth, and Information", in Adriaans, P. & van Benthem, J. (eds.), *Philosophy of Information*, North Holland. pp. 321-60.
- ____ (2010), *Simplicity, Truth, and Probability*, Research Showcase @ CMU (URL=<http://repository.cmu.edu>).
- Kelly, K. T. & Mayo-Wilson, C. (2012), "Causal Conclusions That Flip Repeatedly and Their Justification", arXiv preprint arXiv:1203.3488.
- Pearl, J. (2000), *Causality*, Cambridge University Press.
- Skyrms, B. (1980), *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*, Yale University Press.
- Spirites, P., Glymour, C. & Scheines, R. (1993/2000), *Causation, Prediction, and Search*, MIT press.

논문 투고일	2016. 10. 22.
심사 완료일	2016. 11. 18.
게재 확정일	2016. 11. 24.

Simpson Paradox, Bayes Nets and Causal Modeling: On Young E. Rhee's Criticism of Faithful Condition

Joonsung Kim

In his outstanding book on Bayesianism, Young E. Rhee(2015) critically examines Bayes nets in terms of Simpson's paradox that refutes the Faithfulness condition essential to SGS's theory(Spirites, Glymour, Scheins 1993). His criticism of the Faithfulness condition is in some ways aligned with the opponents of SGS's theory such as Cartwright(2001[2003]). In this paper, I show that the Faithfulness condition of SGS's theory has the methodological significance for causal inferences in non-experimental circumstances, while considering Ockham's principle of simplicity, or economy Kelly(2007; 2008; 2010; 2012) has considerably developed. I show that SGS's theory primarily concerns the efficiency of prediction, while noticing the distinction between prediction (discovery) and explanation (justification). I argue that, from the viewpoint of prediction, the Faithfulness condition is viable and is worth noticing. I expect my discussions on SGS's theory to reconcile SGS's theory and the opponents of it.

Keywords: Bayes net, Bayesian, causal structural model,
Faithfulness condition, Markov condition, simplicity,
Simpson paradox, Kelly, SGS, Ockham, Young E. Rhee