

## 정량적 확률적 인과론에 관하여

김 세 중  
(경주대학교 교양학부)

**요약** 확률적 인과론은 확률관계를 통해 인과 관계를 밝히려는 이론이다. 그런데 만약 단지 C가 E의 발생 확률을 높인다는 사실을 밝히는 것으로 그치지 않고 더하여 C가 E를 발생시키는데 얼마나 기여하는지 그 기여도도 밝힐 수 있다면 우리는 원인과 결과의 관계에 대하여 훨씬 더 많은 정보를 얻을 수 있게 될 것이다. 이 글에서 나는 베이기의 개념에 기반한 멜러나 엘스의 정량적 확률적 인과론들을 살펴본 후 그 이론들이 인과적 효과도나 인과적 연관도를 밝히는데 부적절한 이론들임을 보인다. 그 후 나는 인과적 효과도를 측정하는데 보다 적절한 공식을 제시하며 이 공식에 기반하여 인과적 연관도 또는 인과적 기여도의 공식도 제시한다.

**주요어** 확률적 인과론, 정량적 확률적 인과론, 멜러, 엘스, 인과적 효과도, 인과적 기여도, 인과적 연관도.

확률적 인과론은 확률관계를 통하여 인과 관계를 밝히려는 이론이다. 그런데 대부분의 확률적 인과론자들이 주 초점을 맞추었던 것은 C가 E의 원인임을 확률 값의 비교를 통하여 밝히는 일이었다. 확률적 인과론의 가장 기본적 원칙은 C가 주어졌을 경우 E가 발생할 확률이 C가 주어지지 않았을 경우 E가 발생할 확률보다 높으면 C는 E의 원인이라는 것이다. 즉,  $P(E/C) > P(E/-C)$  이면 C는 E의 원인이라는 것이다. 예를 들자면 흡연자일 경우 폐암에 걸릴 확률이 비흡연자일 경우 폐암에 걸릴 확률보다 높으면 흡연은 폐암의 원인이라는 것이다. 물론 이 기본적 원칙에는 수많은 예외들이 있을 수 있으며 이런 예외들을 확률

값의 비교를 통해 어떻게 설명할 것인가가 확률적 인과론자들이 주로 다루는 문제들이라 할 수 있을 것이다.

그런데 만약 단지 C가 E의 발생 확률을 높인다는 사실을 밝히는 것으로 그치지 않고 더하여 C가 E를 발생시키는 데 얼마나 기여하는지 그 기여도도 밝힐 수 있다면 우리는 원인과 결과의 관계에 대하여 훨씬 더 많은 정보를 얻을 수 있게 될 것이다. 그러므로 확률적 인과론에 있어서도 단순히 확률 값을 비교하는 비교적 확률적 인과론보다는 인과적 기여도를 밝혀주는 정량적 확률적 인과론이 보다 더 발전된 형태의 확률적 인과론이라 말할 수 있을 것이다. 이런 이유로 여러 철학자들이 인과적 연관도라든지 인과적 기여도 등의 개념을 소개하였다. 인과적 연관도나 인과적 기여도는 비슷한 개념으로 한 인과적 요인이 결과 사건의 발생과 얼마나 밀접히 관련되었나를 보여준다. 굿(Good)은 log의 개념을 사용하여 인과적 기여도를 공식화하였고 기어(Giere) 또한 나름대로의 공식을 제시하였다. 그러나 이 글에서 나는 빼기의 개념에 기반한 멜러(D.H. Mellor)와 엘스(Eellery Eells)의 정량적 확률적 인과론을 살펴보고 한다. 먼저 나는 멜러의 인과적 효과도의 이론을 살펴본 후 멜러의 공식이 원인이 결과를 불러일으키는데 얼마나 효과적인가에 대한 우리의 자연스런 직관과 합치하지 않음을 밝힐 것이다. 다음으로 엘스의 인과적 연관도의 이론을 살펴본 후 그의 이론 또한 멜러의 이론과 같은 문제를 안고 있음을 밝힐 것이다. 마지막으로 나는 보다 개량된 인과적 효과도의 공식을 제시할 것이며 그 개념에 기반하여 인과적 연관도의 공식 또한 제시할 것이다. 여기에서 한가지 지적하고 싶은 것은 철학자들이 정량적 확률적 인과론을 전개할 때에 인과적 효과도와 인과적 연관도가 서로 다른 개념이라는 사실을 간과하고 있다는 점이다. 그러나 이 글을 통해 우리는 그들이 서로 다

른 개념임을 알 수 있게 될 것이다.

### 1. 멜러의 인과적 효과도 공식

『인과의 사실들』(*The Facts of Causation*)이라는 그의 책에서 멜러는 인과적 연관성에 관한 정량적 이론을 소개한다. 이 책에서 멜러가 주로 초점을 맞춘 것은 인과적 설명의 문맥이다. 비록 인과적 설명의 문제와 확률적 인과론의 문제가 서로 다른 측면을 조명한다고 볼 수는 있으나 이들이 모두 인과 관계에 관한 문제라는 점에서 서로 맥이 닿아 있다고 하겠다.

멜러는 원인  $C$ 가 다른 모든 관련된 상황  $S$ 와 결합하여 결과  $E$ 를 유발할 기회도(chance)를  $ch_C(E)$ 라는 그 자신만의 용어로 표현하였다(멜러, 25쪽). 그렇다면  $ch_C(E)$ 는 상황  $S$  속에서  $C$ 가 없을 경우  $E$ 가 발생할 기회도를 나타낸다고 하겠다. 멜러의 기회도의 개념이 우리에게 보다 익숙한 조건부 확률인  $P(E/C)$ 와 다르기 때문에 멜러의 이론을 다루는 데 있어서 우리는 그냥 그의 용어를 사용하도록 하겠다.

멜러는 인과 관계와 관련된 세 가지 개념들을 논하였다; 원인은 결과의 증거이고, 결과를 설명하며, 결과를 불러일으키는 유용한 수단이다. 이들은 모두 원인과 결과 사이의 연결에 관한 개념이다. 이들 개념들과 관련하여 우리는 세 가지 정량적 개념들을 생각할 수 있다: 그들은 증거적 지지도(degree of evidential support), 인과적 설명도(degree of causal explanation), 그리고 인과적 효과도(degree of causal effectiveness)이다. 이들 개념들에 관하여 멜러는 다음과 같이 주장한다:

“이 세 가지 개념들은 한가지 공통적 측정치를 갖는데 그것은 그들 모두가 C가 얼마나 효과적으로 E를 유발하는가에 대한 자연스런 측정치라는 것이다.”(멜러, 93쪽)

이는 증거적 지지도나 인과적 설명도 그리고 인과적 효과도는 모두 인과적 효과도에 의해서 표현된다는 말이다. 멜러는 또한 C로 하여금 E를 설명하게 하고 C로 하여금 E에의 유용한 수단이 되게 하는 것은 C가 E의 가능성을 높이는 데에, 즉  $ch_c(E)$ 가  $ch_{-c}(E)$ 를 초과하는 데에 있다고 보았다. 다음 인용문이 이를 분명히 보여준다:

“C가 E를 결정하는데 가까워질수록 C와 E 사이의 이들 연결이 더 강해진다: C가 E를 위해 제공하는 증거적 지지도가 강할수록 C는 E를 더 잘 설명하며 C는 E에의 더 유용한 수단이 된다.”(멜러, 93쪽)

이러한 비교적 개념으로부터 멜러는 과감하게 정량적 개념으로 비약한다. 그는 증거적 지지도와 인과적 설명도, 그리고 인과적 효과도가 모두  $ch_c(E)$ 와  $ch_{-c}(E)$  사이의 차이에 의해서, 즉  $ch_c(E) - ch_{-c}(E)$ 에 의해서 측정된다고 주장하였다. 예를들어, 인과적 설명도와 관련하여 그는 다음과 같이 논증한다:

“인과적 설명도를 측정하는 것은 주어진 상황 속에서 C가  $ch(E)$ 를 높임으로써  $-E$ 의 가능성을 얼마나 줄이느냐이다-이것이 왜  $ch_c(E) - ch_{-c}(E)$ 가 C가 E에의 얼마나 강한 증거가 되는가에 대한 측정치인 것처럼 C가 얼마나 E를 잘 설명하는가에 대한 자연스런 측정치이다.”(멜러, 96쪽)

$ch_c(E)-ch_{-c}(E)$ 에 대한 대안으로 그는  $ch_c(E)/ch_{-c}(E)$ 를 고려해보지만 곧 이를 부적합한 것으로 폐기해버리는데 그 이유는  $ch_c(E)/ch_{-c}(E)$ 는 주어진 상황 속에서 C가  $-E$ 에 반하여 E를 확실히 발생시키는데 얼마나 가까이 있는지에 대한 측정치가 되지 못하기 때문이다. 그리고 땀나는 E를 설명하는 C의 질은 C가  $ch(E)$ 를 얼마나 1에 가까이 옮기는지 즉  $ch(-E)$ 를 얼마나 0에 가까이 옮기는지에 의해서 가장 잘 측정되지  $ch_c(E)$ 가  $ch_{-c}(E)$ 보다 몇 배나 큰지에 의해서 측정되는 것은 아니라고 믿기 때문이다.

이 시점에서 우리는 적어도 두 가지 의문을 제기할 수 있다. 첫째로, 위의 세 가지 개념들이 모두  $ch_c(E)-ch_{-c}(E)$ 에 의해서 측정될 수 있는가? 둘째로, 이 공식이 우리가 관심을 갖는 인과적 연관도나 인과적 기여도를 나타낼 수 있는가?

먼저 위의 세 가지 개념들이 모두  $ch_c(E)-ch_{-c}(E)$ 에 의해서 측정될 수 있는가의 의문에 대한 나의 답은 부정적인데 그 이유는 비록 이 공식이 직관적으로 다른 개념에는 맞아 떨어질지는 몰라도 최소한 인과적 효과도를 측정하는데는 적합하지 않다고 생각하기 때문이다. 인과적 설명도와 인과적 효과도를 비교하기 위해서 다음의 경우를 살펴보자:

치사율이 0.1인 질병  $d_1$ 이 있다. 이 질병을 치료하기 위하여 한 제약회사에서 신약  $x$ 를 개발하였는데 그 약은  $d_1$ 의 치사율을 0.01로 낮춘다. 그런데 과학자들은 신약  $x$ 가 또 다른 훨씬 더 치명적인, 치사율이 0.91에 이르는 질병  $d_2$ 에도 효과가 있는 것을 발견하였다. 이 신약은  $d_1$ 의 경우와 마찬가지로  $d_2$ 의 경우에도 치사율을 0.01로 낮추었다.

이 경우에  $ch_c(E) - ch_{-c}(E)$ 는 C가 얼마나 잘 E를 설명하는지에 대한 자연스런 측정치라고 생각할 수도 있다. 환자가 x를 복용한 후  $d_2$ 로부터 회복되었다고 하더라도 x를 복용한 사실이 환자의 회복을 별로 잘 설명할 것 같지 않은데 그 이유는 설혹 그 환자가 x를 복용하지 않았더라도 회복했을 가능성이 높기 때문이다. 즉 질병  $d_1$ 의 경우 약 x를 복용하면 회복률이 0.9에서 0.99로 높아지므로  $ch_c(E) - ch_{-c}(E) = 0.99 - 0.9 = 0.09$ 로서 낮은 수치이기 때문이다. 그러나 만약 그가 x를 복용한 후  $d_2$ 로부터 회복되었다면, x를 복용한 사실은 환자의 회복에 대해 훨씬 더 잘 설명할 것 같은데 그 이유는 그 환자가 x를 복용하지 않았더라면 회복하지 못했을 가능성이 높기 때문이다. 즉 질병  $d_2$ 의 경우 약 x를 복용하면 회복률이 0.09에서 0.99로 높아지므로  $ch_c(E) - ch_{-c}(E) = 0.99 - 0.09 = 0.9$ 로서 높은 수치이기 때문이다.

그러나 이 경우에도  $ch_c(E) - ch_{-c}(E)$ 가 C가 E를 일으키는 데 얼마나 효과적인가를 측정하는 공식으로서는 부적절함을 우리는 알 수 있다. 이 공식에 의하면 신약 x가  $d_1$ 을 치료하는 효과도는  $0.99 - 0.9 = 0.09$ 이다. 반면에, x가  $d_2$ 를 치료하는 효과도는  $0.99 - 0.09 = 0.9$ 이다. 그러므로 x가  $d_1$ 이나  $d_2$ 를 치료하는데 모두 매우 효과적인 약임에도 불구하고 멜러의 공식은  $d_1$ 과  $d_2$ 에 대한 x의 효과도를 극단적으로 다르게 평가한다: x는 0.09라는 아주 낮은 효과도로  $d_1$ 을 치료하지만 0.9라는 높은 효과도로  $d_2$ 를 치료한다. 아마도 신약 x를 개발한 제약회사 연구원들은 x가  $d_1$ 을 치료하는 효과도가 0.09에 불과하다는 주장에 접하면 무척 어이없어 할 것이다. 물론  $d_1$ 과  $d_2$ 에 대한 x의 효과도가 다를 것은 분명

하지만 이렇게까지 차이가 나는 것은 아니라는 생각을 하게 된다. 이러한 차이는 신약  $x$ 가 질병들을 치료하는데 얼마나 효과가 있는가를 측정하는 방법으로서 멜리의 공식을 사용하는 것은 적합하지 않음을 보여준다 하겠다.

이러한 논증에 대해 멜리는 그러한 차이를 이해할 수 있는데 그 이유는 그 약이  $d_1$ 과  $d_2$ 에 대해 다른 인과적 역할을 하기 때문이라고 답할 수 있다. 그러나 다음의 경우를 또 살펴보도록 하자:

치사율이 0.9인 질병  $d_3$ 이 있다. 이 질병을 치료하기 위하여 한 제약 회사에서 신약  $y$ 를 개발하였는데  $y$ 는  $d_3$ 의 치사율을 0.81로 낮춘다.

멜리의 공식을 따르면 신약  $y$ 가  $d_3$ 을 치료하는 효과도는  $0.19 - 0.1 = 0.09$ 이다. 이 경우 문제가 되는 것이 신약  $x$ 가  $d_1$ 을 치료하는 효과도 또한 정확히 0.09라는데 있다. 같은 치료 효과도가  $x$ 와  $y$ 에 부여되었다. 그러나  $x$ 가  $d_1$ 에 상당히 효과적인 약인데 비해  $y$ 가  $d_3$ 에 별로 효과적이지 않다는 사실을 감안하면 이들에게 같은 효과도를 부여하는 것은 직관적으로 옳지 않은 것으로 보인다.

이러한 우리의 직관이 어디에서 유래하는지를 차분하게 따져보면 우리는 효과도 측정을 위한 보다 나은 공식을 찾을 수 있게 된다. 어떤 약이 문제의 질병으로부터 사람의 생명을 구하는데 얼마나 효과적인가를 살필 때 우리는 그 질병에 걸렸더라도 죽지 않을 사람들은 고려할 필요가 없는데 왜냐면 그들은 그 질병에 걸리든 걸리지 않든 상관없이 죽지 않을 것이기 때문이다. 질병  $d_1$ 의 경우 이 질병에

걸리느냐에 상관없이 90 사람은 살아남는다. 그러므로 그 밖의 10명이 그 질병에 약한 사람들이라 할 수 있다. 이들 10명중에서 9명이 약  $x$ 에 의해서 살게 된다. 10명중에서 9명을 살리므로  $x$ 는  $d_1$ 에 매우 효과적인 약이다. 이에 반해 질병  $d_3$ 의 경우 이 질병에 걸리느냐에 상관없이 10명이 살아남는다. 그러므로 그 밖의 90명이 그 질병에 약한 사람들이라 할 수 있다. 이들 90명중에서 9명이 약  $y$ 에 의해서 살게 된다. 90명중에서 9명을 살리므로  $y$ 는  $d_3$ 에 별로 효과적이지 않은 약이라 말할 수 있다. 그러나 멜러의 공식은 이들에게 똑 같은 정도의 효과도인 0.09를 부여한다. 이러한 효과도의 비교를 통해 우리는 멜러의 공식이 효과도를 측정하는데 적합치 않음을 알 수 있다. 그러므로 우리는 한 요인이 결과를 불러일으키는데 얼마나 효과적인가를 체계적으로 측정하기 위해 멜러의 공식과는 다른 공식을 찾아낼 필요가 있다 하겠다.

내 생각으로는  $x$ 가  $d_1$ 을 치료하는 효과도로는 9/10을 부여하고  $y$ 가  $d_3$ 을 치료하는 효과도로는 9/90을 부여하는 것이 적합해 보인다. 이것이 멜러의 공식에 의한 값보다 더 적절히 약의 효과도를 나타내는 것으로 보인다. 이러한 생각에 기반하여 나중에 원인이 결과를 유발하는 효과도를 측정하는 일반적인 공식을 제시할 것이다.

멜러의 생각과 달리 인과적 설명도와 원인이 결과를 유발하는 효과도는 서로 다른 것으로 보인다. 그리고 내 생각에는 멜러의 공식은 적어도 인과적 효과도를 측정하기에는 부적합한 공식이다. 그렇다면 이들 중 어느 하나가 원인이 결과의 발생에 기여하는 기여도로 이해될 수 있을까? 나는 인과적 설명도나 인과적 효과도는 모두 원인이 결과의 발



생에 기여하는 기여도, 또는 원인이 결과와 인과적으로 관련된 인과적 연관도와는 다른 개념이라 생각한다. 다만 원인이 결과를 유발하는 효과도는 인과적 연관도와 밀접한 관련이 있는데 나중에 원인이 결과를 유발하는 효과도를 측정하는 공식을 제시한 후 이를 이용하여 인과적 효과도와는 다른 값인 인과적 연관도를 측정하는 공식을 제시할 것이다. 그러기에 앞서 먼저 엘스의 정량적 확률적 인과론을 살펴보도록 하자.

## 2. 엘스의 정량적 확률적 인과론

엘스의 정량적 확률적 인과론은 원인이 주어졌을 경우 결과가 발생할 확률 값에서 원인이 주어지지 않았을 경우 결과가 발생할 확률 값을 빼는 것을 공식의 기본으로 한다는 점에서 멜러의 것과 거의 유사한 것이라 할 수 있다. 엘스는 정량적 확률적 인과론을 그의 개별 사건 확률적 인과론의 기반 위에 세운다. 유형  $X$ 를 예화하는 개별 사건  $x$ 와 유형  $Y$ 를 예화하는 개별 사건  $y$  사이의 인과적 연관성을 측정하는데 사용하는 배후 문맥  $K_a$ 의 특성을 기술한 후 엘스는 개별 사건간의 인과적 중요성의 정도를(degree of causal significance) 설명한다.

엘스의 개별 사건 확률적 인과론은 한 사건이 다른 사건과 갖는 네 가지 인과적 중요성의 요소들을 설명함으로 시작한다; 즉 “때문에”(because), “불구하고”(despite), “독립성(independence)” 그리고 “자율성(autonomy)”의 네 가지 요소들에 대한 설명으로부터 그의 개별 사건 확률적 인과론은 시작한다. 먼저, “만약  $x$ 가 일어난 직전이나  $x$ 가 일어나

기 직후의 Y의 확률이 모두 같다면, y는 x와 독립적으로 Y이다.” 그리고 “만약 (i) x가 일어난 시간에 Y의 확률이 변화하고, (ii) x가 일어난 직후에 Y의 확률이 높고, (iii) 이 확률이 x가 일어나기 직전보다 높으나, (iv) x가 일어난 이후와 y가 일어나기 이전의 어떤 시점에서 Y의 확률이 낮게 떨어지면, y는 x와 자율적으로 Y이다.” 그리고 “만약 x 때문에 또는 x에도 불구하고 y가 Y라면, x는 개별 사건의 차원에서 y가 Y인 것과 인과적으로 연관이 있다; 그리고 만약 y가 Y인 것이 인과적으로 x에 독립적이거나 x에 자율적이라면, x는 개별 사건의 차원에서 y가 Y인 것과 인과적으로 무관하다.”(엘스, 297쪽) 엘스는 이 네 가지 요소들을, 즉 because, despite, independence, 그리고 autonomous를, 각각 그 첫 문자를 따서 B, D, I, 그리고 A로 표현하였다.

엘스의 정량적 확률적 인과론은 이 네 가지 경우에 각각 다른 값을 부여하는데 이 값들은 다음의 세 가지 간단한 확률 값을 사용하여 주어진다.

- (1) P는 사건 x가 일어난 시간인  $t_x$  바로 직전에 Y의 확률이다. 이 확률은 시간 t가 과거로부터  $t_x$ 에 접근하는 한계 값이다.
- (2) P\*는 사건 x가 일어난 시간인  $t_x$  바로 직후의 Y의 확률이다. 이 확률은 시간 t가 미래로부터  $t_x$ 에 접근하는 한계 값이다.
- (3) M은 시간 간격( $t_x, t_y$ ) 사이에서  $P_H(Y)$ 가 갖는 최저 값이다.

요약하자면,  $P^-$ 는  $t_x$  이전에  $Pr_t(Y)$ 가 갖는 마지막 값이고  $P^+$ 는  $t_x$  이후에  $Pr_t(Y)$ 가 갖는 첫 번째 값이고  $M$ 은  $t_x$ 와  $t_y$  사이에서  $Pr_t(Y)$ 가 갖는 최저 값이다. 엘스는 이런 세 가지 값들을 이용하여 위 네 가지 경우의 확률 값을 다음과 같이 부여한다:

$$B = M - P^-;$$

$$D = P^- - P^+;$$

$$A = P^+ - M; \text{ 그리고}$$

$$I = 1 - |P^- - P^+|. (\text{엘스, 355쪽})$$

엘스는 인과적 연관도  $B$ 가  $P^+ - P^-$  대신에  $M - P^-$ 로 정의되는 이유를 다음과 같이 설명한다:

“만약 시간  $t_x$ 에 사건  $x$ 가  $X$ 를 예화함으로써  $y$ 가  $Y$ 를 예화할 확률을 높이는데 성공했으나 나중의 사건이— $x$ 가  $X$ 를 예화한 사건의 있을 법하지 않은 결과들— $y$ 가  $Y$ 를 예화할 확률을 낮춘다면, 그렇다면 그 나중 사건들은  $x$ 가  $X$ 를 예화함으로써  $y$ 가  $Y$ 를 예화하는데 기여했던 긍정적 인과적 중요성을 부분적으로 또는 모두 빼앗아 버릴 수 있다.”(358쪽)

즉 사건  $x$ 가 어떤 시간에  $y$ 의 확률을 높이기는 하지만 나중에 다른 사건으로 인하여  $y$ 의 확률이 낮아져버리고 그럼에도 불구하고 또 다른 사건으로 인하여 결국  $y$ 가 일어나게 된다면  $x$ 를  $y$ 의 발생 원인으로 보는 것은 적절하지 않을 것이기 때문이다. 예를 들어 어떤 암살자가 사람을 죽이기 위해 독 묻힌 화살로 그 사람을 겨냥하고 정확히 쏘았으나 갑자기 센 바람이 불어와 화살이

나가는 방향이 틀어져 버렸다면 설혹 나중에 다른 이유로 그 사람이 죽었다고 하더라도 그 암살자가 화살을 쏜 사건을 그의 죽음의 원인으로 보는 것은 적절하지 않을 것이다.

엘스의 개별 사건 확률적 인과론을 살피는 것이 이 글의 주목적이 아니기 때문에 여기서는 최저값  $M$ 의 값을 고려하는 것이 적절한가의 문제를 살피지 않을 것이고 더하여 엘스가 말하는 인과적 독립성이나 인과적 자율성의 개념들이 무엇이며 왜 그들을 고려해야 하는지의 문제들은 살피지 않겠다. 단지 그가 제시한 인과적 연관도인  $B$ 의 공식이 직관적으로 그럴듯한 값을 내는지만 간단히 살펴보도록 하겠다.

엘스의 정량적 확률적 인과론에서  $B$  값을 측정하는 공식은 원인이 주어졌을 경우 결과가 발생할 확률 값에서 원인이 주어지지 않았을 경우 결과가 발생할 확률 값을 빼는 것을 기본으로 한다는 점에서 멜러의 것 보다 별로 나을 것이 없다고 말할 수 있다. 엘스가 그의 논문에서 다루었던 라이헨바하 폭포에서의 사건의 예를 조금 변형한 다음의 경우를 살펴보자:

설록 홈즈가 라이헨바하 폭포 아래쪽 계곡 길을 걸어가고 있다. 그런데 계곡 위쪽 절벽 위에는 그의 적인 모리아티가 그를 죽일 의도로 커다란 바위 덩어리를 하나 막 굴러 떨어뜨리려 하고 있다. 바로 그때 홈즈의 동료인 왓슨이 절벽 위에 도착하여 바위를 굴러 떨어뜨리려는 모리아티를 발견하였는데 그는 홈즈를 구하기 위해 그 순간 자신이 할 수 있는 일이라고는 오직 그 바위를 밀어서 떨어지는 방향을 틀어지게 함으로써 모리아티가 신중하게 겨냥하여 바위를 밀어 떨어뜨린 경우보다 홈즈가 맞아 죽을 확률을 떨어뜨리는 수밖에 없다고 생각한다. 그래서 왓슨과 모리아티가 서

로 다른 목적으로 바위를 향해 뛰어가고 있다. 그들이 바위에 먼저 도착할 확률은 서로 같다고 하자. 또 만약 아주 정확한 기술을 가진 모리아티가 그 바위를 밀어뜨린다면 홈즈가 바위에 맞아 죽을 확률은 0.999이고 만약 왓슨이 그 바위를 밀어뜨린다면 홈즈가 죽을 확률은 0.401이라고 하자. 불행히도 모리아티가 먼저 도착하여 바위를 밀어뜨렸고 그 결과로 홈즈가 바위에 맞아 죽었다고 하자.

이 경우에  $P^- = (0.999+0.401)/2 = 0.7$ 이고  $M = 0.999$ 이다. 그러므로  $B = 0.999-0.7 = 0.299$ 이다.  $M$ 이 가질 수 있는 최대치는 1이다. 그러므로 모리아티가 바위를 밀어뜨린 것이 홈즈의 죽음에 대하여 갖는 인과적 중요성의 최대치는 아무리 모리아티가 결정적으로 홈즈의 죽음에 기여했을 지라도 0.3을 넘을 수 없다. 다른 말로 하자면 모리아티가 홈즈를 죽인 방법이 아무리 효과적이었을지라도, 그 인과적 효과도는 0.3을 넘을 수 없다. 그러므로 엘스의 공식은 멜러의 공식처럼 인과적 효과도를 올바르게 반영할 수 없어 보인다.

더욱이 엘스의 공식은 멜러의 공식이 갖지 않은 또 다른 문제점을 안고 있는데 그것은 엘스의 공식이 어떤 원인에 도 결과를 불러일으킬 인과적 중요성으로 0을 부여할 가능성을 안고 있다는 것이다. 이 문제를 다음의 간단한 예를 통해 살펴보자:

시간  $t_1$ 에 모리아티는 그 바위를 홈즈를 향해 밀어뜨렸는데 그때 홈즈가 죽을 확률은 0.9가 되었다. 잠시 후인  $t_2$ 에 홈즈는 모리아티가 계획한대로 바위에 맞아 죽었다.

비록 모리아티가 홈즈를 향해 바위를 굴린 시간은  $t_1$ 이지

만  $t_1$ 에 무한대로 가까운 앞 시간에 홈즈가 죽을 확률은  $t_1$ 과 거의 차이가 없을 것이다. 왜냐면 모리아티는 바위를 거의 굴리기 시작하고 있을 것이기 때문이다. 그렇다면  $P^- = 0.9$ 에 가까울 것이고  $M - P^- = 0$ 에 가까울 것이다. 이 문제는 엘스의 공식이 개별적 확률적 인과론에 기초하고 있음으로 해서 발생하는 또 다른 문제라 할 수 있다.

### 3. 나의 정량적 확률적 인과론

우리는 멜러와 엘스의 정량적 이론이 인과적 효과도를 적절히 포착하지 못하는 것을 보았다. 빼기의 개념에 기반한 그들의 이론들은 우리로 하여금 한 요인이 결과를 불러일으키는데 있어서 얼마나 효과적인지를 측정할 수 있게 하는 직관적으로 적절한 공식을 제시하지 못했다. 이제 나는 이러한 일을 할 수 있는 공식을 제시하고자 한다.

이 목적을 위해 먼저 멜러의 공식을 다룰 때 언급했던 약  $x$ 의 경우를 다시 살펴도록 하자. 그때 나는 어떤 약이 문제의 질병으로부터 사람의 생명을 구하는데 얼마나 효과적인가를 고찰할 때 우리는 그 질병에 걸렸더라도 죽지 않을 사람들은 고려할 필요가 없다고 말했는데 그 이유는 그들은 그 질병에 걸리든 걸리지 않든 상관없이 죽지 않을 것이기 때문이며 그 약의 효과도를 측정하는데 있어서 이처럼 그 약의 효과와 관련하여 상관없는 사람들은 고려할 필요가 없기 때문이다. 질병  $d_1$ 의 경우 이 질병에 걸리느냐에 상관없이 90 사람은 살아남는다. 그러므로 그 밖의 10명이 그 질병에 약한 사람들이라 할 수 있다. 이들 10명중에서 9명이 약  $x$ 에 의해서 살게 된다. 10명중에서 9명을 살

리므로  $x$ 는  $d_1$ 에 매우 효과적인 약이다. 이에 반해 질병  $d_3$ 의 경우 이 질병에 걸리느냐에 상관없이 10명은 살아남는다. 그러므로 그 밖의 90명이 그 질병에 약한 사람들이라 할 수 있다. 이들 90명중에서 9명이 약  $y$ 에 의해서 살게 된다. 90명중에서 9명을 살리므로  $y$ 는  $d_3$ 에 별로 효과적이지 않은 약이라 말할 수 있다. 그러나 펠러의 공식은 이들에게 똑 같은 정도의 효과도인 0.09를 부여한다.

내 생각으로는  $x$ 가  $d_1$ 을 치료하는 효과도로는 9/10을 부여하고  $y$ 가  $d_3$ 을 치료하는 효과도로는 9/90을 부여하는 것이 적합해 보인다. 이것이 펠러의 공식에 의한 값보다 더 적절히 약의 효과도를 나타내는 것으로 보인다. 여기에서 확률 값 9/10이 나타내는 것은 무엇인가? 그것은 그 질병에 약한 사람들 중 얼마나 많은 사람들이 그 약을 복용하여 치료되는가의 비율을 나타낸다.

이제 이런 생각을 조금 더 형식적으로 표현해보도록 하자.  $C_x$ 가 약  $x$ 를 복용하는 인과 요인이라 하고,  $E$ 는 병  $d_1$ 에서 회복하는 것이라 하고,  $C_n$ 은  $C_x$ 의 인과 요인 공간상의 중립점이라고 하자. 이 경우  $C_n$ 은 약  $x$ 를 복용하지 않는 것을 나타낸다<sup>1)</sup>. 그렇다면, 위의 전제들로부터  $P(E/C_x) = 0.99$ 이고  $P(E/C_n) = 0.90$ 임을 알 수 있다. 그 병에 약한 사람들의 비율은  $1 - P(E/C_n) = 0.10$ 이고 그 약을 복용하여 치료되는 사람들의 비율은  $P(E/C_x) - P(E/C_n) = 0.99 - 0.90 =$

1) 여기에서 인과 요인 공간상의 중립점  $C_n$ 을 소개하는 이유는 인과 요인이 주어지지 않은 경우의 확률값을 결정하는 문제, 즉  $P(E/-C)$  값을 결정하는 까다로운 문제가 제기되기 때문이다. 이 문제에 대한 답으로 나는  $C$ 라는 인과 요인이 속하는 공간상의 중립점에서의 확률값이  $P(E/-C)$ 의 값이 되어야 한다고 믿는다. 나는 이 문제를 선점적 인과요인의 문제와 결부하여 다른 논문에서 다루게 될 것이다.

0.09이다. 그렇다면 그 약의 인과적 효과도는  $(P(E/C_x) - P(E/C_n)) / (1 - P(E/C_n)) = 9/10$ 이다.

이러한 직관적 생각은 보다 일반적인 방식으로 표현할 수 있다. 다음을 가정하자:

$x = P(E/C_x)$  :  $x$ 는 특정 인과 요인  $C_x$ 가 주어졌을 경우  $E$ 가 일어날 확률이다.

$n = P(E/C_n)$  :  $n$ 은  $C_x$ 의 인과 요인 공간상의 중립점인  $C_n$ 이 주어졌을 경우  $E$ 가 일어날 확률이다.

$y = CE(C_x \rightarrow E)$  :  $y$ 는  $C_x$ 가  $E$ 를 불러일으킬 인과적 효과도이다.

그렇다면  $x$ 와  $n$ 과  $y$  사이에 성립하는 다음 세 가지의 직관적 전제에 기반하여 우리는 인과적 효과도 함수를 구성할 수 있다:

- (i) 인과 요인  $C_x$ 가  $C_n$ 일 때, 그때 인과적 효과도는 0이다. 즉 만약  $x = n$ 이면  $y = CE(C_x \rightarrow E) = 0$ 이다.
- (ii) 인과 요인  $C_x$ 가  $E$ 를 불러일으키기에 충분한 요인일 때, 그때 인과적 효과도는 1이다.  
즉 만약  $x = 1$ 이면  $y = CE(C_x \rightarrow E) = 1$ 이다.
- (iii) 인과 요인  $C_x$ 가  $E$ 의 발생을 방지하기에 충분한 요인일 때, 그때 인과적 효과도는 -1이다.  
즉 만약  $x = 0$ 이면  $y = CE(C_x \rightarrow E) = -1$ 이다.

이 세 관계가 주어지면 우리는 함수  $y$ 를  $x$ 와  $n$ 으로 표현할 수 있다.  $y$ 를 일차함수인  $ax+b$ 라고 가정하자. 인과적으로 기여하는 경우를 생각하면, 관계 (i)과 (ii)를 이용하여  $a$



와  $b$ 를  $n$ 으로 표현하기에 충분하다. 관계 (i)에 따르면  $y = an+b = 0$ 이며 관계 (ii)에 따르면  $y = a+b = 1$ 이다. 그렇다면  $a = -1/(n-1)$ 이 되며  $b = n/(n-1)$ 이 된다. 그러므로  $y = -x/(n-1)+n/(n-1) = (x-n)/(1-n)$ 이 된다. 반면에 인과적으로 방해하는 경우를 생각하면 관계 (i)와 (iii)를 이용하여  $a$ 와  $b$ 를  $n$ 으로 표현하기에 충분하다. 관계 (i)에 따르면  $y = an+b = 0$ 이며 관계 (iii)에 따르면  $y = b = -1$ 이다. 그렇다면  $a = 1/n$ 이 된다. 그러므로  $y = x/n-1$ 이 된다. 이렇게 우리는 인과적으로 기여하는 경우와 인과적으로 방해하는 경우에 대한 두 개의 공식을 얻게 된다. 이를 정리하면 우리는 인과적 효과도의 함수인  $CE(C_x \rightarrow E)$ 를 다음과 같이 정의할 수 있다:

$$y = \frac{x - n}{1 - n}, P(E/C_x) > n \text{인 경우(인과적으로 기여하는 경우),}$$

$$y = \frac{x - n}{n}, P(E/C_x) < n \text{인 경우(인과적으로 방해하는 경우)}$$

내 생각에 이 공식들은 인과적 효과도를 적절히 포착하는 것 같다. 신약  $x$ 와 질병들  $d_1$ 과  $d_2$ 의 경우를 살펴보자. 멜러의 공식은,  $Ch_c(E)-Ch_{-c}(E)$ ,  $x$ 가 질병  $d_1$ 을 치료하는 효과도는  $0.99-0.9 = 0.09$ 이고 질병  $d_2$ 를 치료하는 효과도는  $0.99-0.09 = 0.9$ 로 판정한다. 그러나  $x$ 가 이 두 질병에 모두 효과적인 약이라는 사실을 감안하면 이 판정은 받아들이기 어려워 보인다. 이에 반해 나의 공식을 따르면 질병  $d_1$ 의

경우  $P(E/C_x) = 0.99$ 이고  $P(E/C_n) = 0.9$ 이므로  $y = CE(C_x \rightarrow E) = 9/10$ 이다. 또한 질병  $d_2$ 의 경우  $P(E/C_x) = 0.99$ 이고  $P(E/C_n) = 0.09$ 이므로  $y = CE(C_x \rightarrow E) = 90/91$ 이다. 이 두 경우 다  $x$ 가 병을 치료하는 효과도는 매우 높게 나타난다. 이러한 결과는  $x$ 가 이 두 질병들에 모두 높은 효과가 있다는 사실과 부합된다.

이제 모리아티가 홈즈를 죽인 경우를 생각해보자. 그 경우  $P^- = (0.999+0.401)/2 = 0.7$ 이고  $M = 0.999$ 이다. 그러므로 엘스의 공식에 따르면  $B = 0.999-0.7 = 0.299$ 이다.  $M$ 이 가질 수 있는 최대치는 1이므로 모리아티가 바위를 밀어 홈즈를 죽인 것이 아무리 효과적이었다고 할지라도 그 인과적 효과도는 0.3을 넘지 못한다. 이에 반해서 나의 공식에 따르면 보다 우리는 적절한 수치를 얻을 수 있음을 알 수 있다. 이 경우에 모리아티가 바위를 밀지 않았을 경우 홈즈가 죽을 확률을 어떻게 계산할 것인가의 문제가 제기된다. 어떤 이들은 그 경우 왓슨이 바위를 밀었을 것이므로 확률은 0.401이라고 주장할 수 있을 것이다. 이 경우 모리아티가 바위를 민 행위가 홈즈의 죽음을 야기시키는데 얼마나 효과적인가를 측정하자면,  $y = (999/1000 - 401/1000) / (1-401 / 1000) = 598/599$ 이다. 이는 모리아티가 바위를 민 행위가 매우 효과적이었다는 우리의 직관과 부합된다고 하겠다. 또 어떤 이들은 아무도 바위를 밀지 않은 경우가 바위를 민 모리아티의 행위와 관련하여 중립 상태라고 주장할 수도 있을 것이다. 이 경우 홈즈가 죽을 확률은 무척 낮을 것이다. 예로 아무도 바위를 밀지 않았을 경우 폼즈가 죽을 확률은 0.001이라고 가정하자. 그러면  $P(E/C_x) = 999/1000$ 이고  $P(E/C_n)=1/1000$ 이므로  $y=(999/1000-1/1000)/$

$(1-1/1000) = 998/999$ 이다. 다시 이 결과는 모리아티가 홈즈를 죽인 것이 매우 효과적이었다는 우리의 직관과 부합된다고 하겠다.

이제까지는 인과적 효과도만을 다루었다. 이제는 왜 인과적 효과도가 인과적 연관도일 수 없는가에 대해 살펴볼까 한다. 이를 위해 철수가 두 개의 중요한 인과적 요인을 안고 있는 운석도라는 섬으로 여행을 떠난 경우를 살펴보도록 하자:

철수는 운석도로 연구 여행을 떠났다. 운석도라는 이름에 겁을 먹은 철수는 아주 비싼 운석 보호기를 사서 착용하고 여행을 떠났다. 그러나 철수의 여행기간 동안 운석도에 운석이 떨어져 철수를 맞춰 철수를 죽일 확률은  $10^{-20}$ 에 불과했다. 비록 운석에 맞았을 경우 살아남을 확률은 0.01에 불과했지만, 그 운석 보호기는 또한 아주 잘 만들어져 있어서 사람이 운석에 맞아도 0.99의 확률로 보호기 착용자를 보호해준다. 그런데 그 섬에는 아주 치명적 풍토병이 번지고 있었는데 외지인인 철수가 그 섬에서 그 병에 걸릴 확률은 0.9이며 그 병의 치사율은 0.3이었다. 그 풍토병에 관해 잘 아는 철수는 또한 그 병에 걸려 죽지 않게 해주는 약을 가져갔는데 그 약 또한 0.99의 확률로 복용자를 보호해 주는 것이었다. 여러 가지 준비를 철저히 한 관계로 철수는 연구 여행을 무사히 마치고 건강히 귀가할 수 있었다. 그러나 그 기간 동안에 그 섬에 운석이 떨어지지 않는 않았다.

이 경우에 운석 보호기가 사람이 운석에 맞아 죽을 가능성으로부터 보호해줄 인과적 효과도를 계산하면  $P(E/C_x) = 99/100$ 이고  $P(E/C_n) = 1/100$ 이므로  $y = CE(C_x \rightarrow E) = 98/99$ 이다. 그러므로 이 운석 보호기가 운석으로부터 사람을 보호해주는 인과적 효과도는 98/99로서 아주 잘 만들어져 있다는 전제에 부합한다. 마

찬가지로 그 약이 치명적 풍토병으로부터 사람을 보호해줄 인과적 효과도를 계산하면  $P(E/C_x) = 99/100$ 이고  $P(E/C_n) = 70/100$ 이므로  $y = CE(C_x \rightarrow E) = 29/30$ 이다. 그러므로 이 약이 질병으로부터 사람을 보호해주는 인과적 효과도는 29/30으로서 아주 효과적인 약이라는 전제에 잘 부합한다.

그러나 멜리의 공식에 의하면 C가 운석 보호기에 의해 보호받는 것이고 E가 여행으로부터 안전하게 귀가하는 것이라면 C의 E에 대한 인과적 효과도는  $ch_C(E) - ch_{-C}(E)$ 에 의해서 포착이 되므로  $ch_C(E) - ch_{-C}(E) = (1 - 10^{-22}) - (1 - 10^{-20}) = 10^{-20} - 10^{-22} = 99 \times 10^{-22}$ 이다. 이 값은 그 운석 보호기가 엄청나게 비효율적이고 쓸모 없는 장비임을 드러낸다. 분명히 이 장비의 제조업자들은 이런 평가에 분개할 것이다. 또한 멜리의 공식은 풍토병으로부터 사람을 보호하는 그 약의 효과성에도 낮은 효과도를 부여한다:  $0.99 - 0.7 = 0.29$ . 이 수치에 대해서도 그 약을 제조한 제약업자들은 무척 분개할 것이다.

비록 내 공식이 그 운석 보호기가 운석으로부터 사람을 보호하는 인과적 효과도를 측정함에 있어 적절한 값을 제시하기는 하지만 그 값이 그 운석 보호기와 철수가 무사히 여행을 마치고 귀가하는 사건과의 인과적 연관도를 가리키는 값은 아니다. 실제로 철수가 그런 비싼 장비를 여행에 가지고 간 것은 그 섬에 거의 무시할만한 확률로, 즉  $10^{-20}$ 의 확률로, 운석이 떨어진다는 사실을 감안한다면 불합리한 선택이라고까지 말할 수 있다. 그러므로 철수가 운석 보호기를 가져간 것이 철수의 무사 귀환에 대해 갖는 인과적 연관도가 98/99일 수는 없다. 내 생각으로는 직관적으로 적합한 인과적 연관도를 얻기 위해서는 우리는 인과적 효과

도에 위험 요소인 운석이 떨어질 확률을 곱해야 하는 것 같다. 그렇다면 철수가 운석 보호기를 가져간 것이 무사 귀환에 대해 갖는 인과적 연관도는  $98/99 \times 10^{-20}$ 이 될 것이다. 이 값은 철수가 운석 보호기를 가져간 것이 무사 귀환에 거의 무관해 보인다는 우리의 관찰과 잘 맞아 떨어지는 것 같다.

일반적으로 우리는 인과적 연관도의 함수인  $CR(C_x \rightarrow E)$ 를 다음과 같이 정의할 수 있다:

$CR(C_x \rightarrow E) = CE(C_x \rightarrow E) \times P(D)$ , 여기에서  $D$ 는 관련된 위험 요소를 나타낸다.

이 공식을 갖고 우리는 철수가 약을 가져간 사실이 그의 무사 귀환과 어떤 인과적 연관도를 갖는지를 계산할 수 있다.  $CR(\text{철수가 약을 가져감} \rightarrow \text{철수의 무사 귀환}) = CE(\text{철수가 약을 가져감} \rightarrow \text{철수의 무사 귀환}) \times P(\text{철수가 그 병에 걸림}) = 29/30 \times 9/10 = 0.87$ . 이 수치는 철수가 약을 가져간 사실이 철수의 무사 귀환에 상당히 연관되어 있음을 보여준다. 이에 비해 멜러나 엘스의 공식들은 또한 적절치 않은 결론을 내리게 하므로 바람직하지 않다 하겠다.

이 논문에서 나는 원인이 결과를 일으키는데 얼마나 효과적인지, 그리고 얼마나 기여했는지를 평가하는 멜러와 엘스의 공식들을 살펴보고 그 공식들이 적절치 않음을 보였다. 또한 나는 인과적 효과도와 인과적 연관도를 측정하기에 보다 적절하다고 생각되는 나름대로의 공식들을 제시하였다. 나는 이러한 공식들을 통해 이 논문이 인과적 효과성과 인과적 연관성이라는 개념들을 명료화하는데 조금이나마 기여하였기를 바란다.

【참고문헌】

Eells, Ellery(1991), *Probabilistic Causality*, Cambridge:  
Cambridge University Press.

Mellor, H.D.(1995), *The Facts of Causation*, New York:  
Routledge.