

## 김성수 교수의 활성화 경로 이론에 대한 변호와 그에 대한 반론\*

김 세 화

**【국문요약】** 김성수 교수는 최근 논문 “반사실적 인과론과 인과 구조식 접근법: 활성화 경로 이론의 재검토”에서 활성화 경로 이론에 대한 변호를 펼친다. 활성화 경로 이론은 반사실적 인과론의 핵심적 직관을 유지하면서 ‘실제로 벌어진 정황’에 주목함으로써 반사실적 인과론에 제기된 반례들을 극복하고자 제시되었다. 그러나 활성화 경로 이론에 대해서도 반례가 제기된 바 있다. 김성수 교수는 이 반례를 반박함으로써 활성화 경로 이론에 대한 변호를 펼치는데, 본 논문에서 필자는 김성수 교수의 반박이 성공하지 못했음을 밝힌다.

**【주요어】** 인과, 반사실적 인과론, 활성화 경로, 반례, 김성수

---

접수일자: 2014.10.21 심사 및 수정완료일: 2014.11.25 게재확정일: 2014.12.19

\* 이 논문의 심사를 맡아 유익한 논평을 해주신 익명의 심사위원님들께 감사 드린다.

1

김성수 교수는 최근 논문 “반사실적 인과론과 인과 구조식 접근법: 활성화 경로 이론의 재검토”에서 활성화 경로 이론에 대한 변호를 펼친다.<sup>1)</sup> 필자는 본 논문에서 김성수 교수의 변호가 성공적이지 않음을 보이고자 한다.

2

활성화 경로 이론은 반사실적 인과론에 제기된 반례들을 극복하기 위해 제시된 이론인데<sup>2)</sup>, 먼저 반사실적 인과론이란 무엇인지 살펴보자.<sup>3)</sup> 반사실적 인과론은 결과가 원인에 반사실적으로 의존하는 것으로 분석하는 이론으로서, 예를 들어,

(돌팔매) 철수가 돌을 던지고 그 돌에 맞아서 유리창이 깨졌다.

와 같은 상황에서

(의존1) 만약 철수가 돌을 던지지 않았더라면 유리창은 깨지지 않았을 것이다.

라는 반사실적 조건문이 참이기 때문에 이 경우 철수가 돌을 던진 사건이 유리가 깨진 사건의 원인이라고 주장한다. 반사실적 인과론에 대해서는 잘 알려져 있는 것처럼 지금까지 (전 선점), (후 선점),

---

1) 김성수 (2012).

2) 예를 들어 Hitchcock (2001a, 2001b)을 볼 것.

3) Lewis (1973, 1986, 2000)를 볼 것.

(전이성)과 같은 반례가 제기되었다. 이들은 다음과 같다.

(전 선점) 철수와 영희 둘 다 자신이 든 돌을 유리창에 던지려고 한다. 철수가 먼저 돌을 던지고 이를 본 영희는 돌을 던지지 않기로 한다. 철수가 던진 돌이 유리창에 맞았고 유리창은 깨졌다.<sup>4)</sup>

(후 선점) 철수와 영희 둘 다 유리창을 향해 돌을 던진다. 철수가 던진 돌이 영희가 던진 돌보다 먼저 유리창을 깨고 영희가 던진 돌은 유리창이 깨지고 없는 빈 공간을 날아간다.<sup>5)</sup>

(전이성) 암살자가 보스의 커피에 독을 넣었다. 이를 본 보디가드는 커피에 해독제를 넣었고 그 해독제가 커피 속의 독을 중화시켰다. (해독제는 그것만 마셨을 경우 무해하다.) 보스는 이 커피를 마셨고 살아남았다.<sup>6)</sup>

반사실적 인과론은 위의 세 경우에 대해 우리의 직관에 반하는 결론을 낸다.

### 3

활성 경로(active route) 이론은 반사실적 인과론의 핵심적 직관을 유지하면서 ‘실제로 벌어진 정황’에 주목함으로써 위의 반례들을 극복하고자 제시되었다. 활성 경로 이론은 다음과 같다.

---

4) 김성수 (2012), pp. 262-263.

5) 김성수 (2012), p. 264.

6) 김성수 (2012), p. 264.

(활성 경로 이론) 발생한 사건  $c$ 와  $e$ 가 서로 다른 사건이고  $c$ 와  $e$ 가 변수  $X$ 와  $Z$ 의 값(alteration)이라고 하면, 적절한 인과 모델  $\langle V, E \rangle$ 에서  $X$ 로부터  $Z$ 로의 활성(인과) 경로가 있는 경우 그리고 단지 그 경우  $c$ 는  $e$ 의 원인이다.<sup>7)</sup>

이 때 활성 경로는 다음과 같이 정의된다.

(활성 경로) 인과 모델  $\langle V, E \rangle$ 에서 경로  $\langle X, Y_1, \dots, Y_n, Z \rangle$ 는 다음의 경우 그리고 단지 다음의 경우 활성 경로이다.  $Y \in E$ 인 모든 변수  $Y$ 에 대해, 그  $Y$ 가  $X$ 와  $Z$  사이에 있지만 경로  $\langle X, Y_1, \dots, Y_n, Z \rangle$ 에 속해 있지 않을 경우 그러한  $Y$ 의 값을  $E$ 에서  $Y$ 가 실제로 갖는 값을 갖도록 만들어버리는 구조식을  $E'$ 라고 하면, 이렇게 새롭게 만들어낸  $E'$ 에서  $Z$ 는  $X$ 에 반사실적으로 의존한다. (만약 주어진 경로에 속하지 않으면서  $X$ 와  $Z$  사이에 있는 그런 변수가 없다면  $E'$ 는  $E$ 와 동일하다.)<sup>8)</sup>

구체적으로 이 이론이 (전 선점), (후 선점), 그리고 (전이성) 반례를 어떻게 극복하는지를 통해 이 이론이 원인과 결과를 어떻게 분석하는지 살펴보자.

(전 선점)의 예에서 각각의 사건을 변수로 놓으면 다음과 같다:

C: 철수가 돌을 던짐.

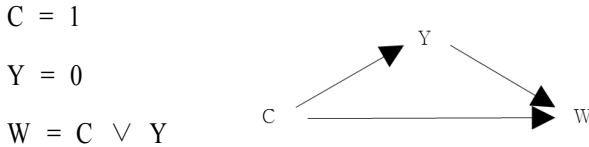
Y: 영화가 돌을 던짐.

W: 유리창이 깨짐.

<sup>7)</sup> Hitchcock (2001a), p. 287. 김성수 (2012), p. 270에서 재인용.

<sup>8)</sup> Hitchcock (2001a), p. 286. 김성수 (2012), pp. 270-271에서 재인용.

이 때 사건이 일어난 경우 그 변수의 값을 1이라고 하고, 사건이 일어나지 않은 경우 그 변수 값을 0이라고 하고, 실제로 일어난 대로 변수에 값을 부여하면 우리는 다음과 같은 구조식 E'를 가진다.



이 구조식을 풀면  $W = 1$ 이다. 이제 ‘실제로 벌어진 정황’인 영희가 돌을 던지지 않은 상황을 고정시킨 상태에서 ‘철수가 돌을 던지지 않았더라면,’ 유리창이 깨지는 사건이 일어나는지 아닌지 살펴보자. 즉,  $Y = 0$ 으로 고정시킨 상태에서,  $C$ 의 값을 0으로 한 경우  $W$ 의 값이 1이 되는지 0이 되는지 살펴보자.  $Y = 0$ 으로 고정시킨 상태에서  $C = 1$ 에서  $C = 0$ 으로 그 값을 변화시키면  $W = 0$ 으로 변하게 된다. 이것은 실제로 벌어진 정황인 영희가 돌을 던지지 않은 사건을 고정시킨 상태에서 철수가 돌을 던지지 않았더라면 유리창이 깨지지 않았을 것이라는 것을 의미하며, 따라서

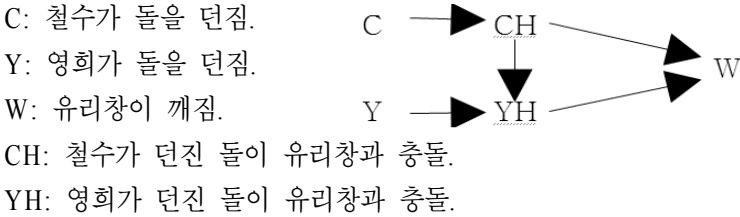
(의존1) 만약 철수가 돌을 던지지 않았더라면 유리창은 깨지지 않았을 것이다.

가 참이라는 것을 의미한다. 활성화 경로 이론에 따르면 E'에서  $W$ 는  $C$ 에 반사실적으로 의존하므로  $C - W$  경로는 활성화 경로이다. 즉,  $C$ 로부터  $W$ 로 이어지는 활성화 경로가 존재하므로 이 경우 철수가 돌을 던진 사건은 유리창이 깨진 사건의 원인이 된다.

(후 선점)의 예에서도 우리는 다음과 같은 구조식 E'를 가진다.<sup>9)</sup>

$$\begin{aligned}
 E': & C = 1 \\
 & Y = 1 \\
 & CH = C \\
 & YH = 0 \\
 & W = CH \vee YH
 \end{aligned}$$

각각의 변수는 다음의 사건을 나타낸다.



실제로 일어난 대로 변수 값을 부여하면 위에서 보듯이  $C = 1, Y = 1, YH = 0$ 이다. 이제 ‘실제로 벌어진 정황’인  $YH$ 를 0으로 고정할 때,  $C$ 의 값을 0으로 바꾸면  $CH = 0$ 으로 되고,  $W = 0$ 으로 변화된다. 이것은 실제로 벌어진 정황인 영희가 던진 돌이 유리창과 충돌하지 않은 사건을 고정시킨 상태에서 철수가 돌을 던지지 않았더라면 유리창이 깨지지 않았을 것이라는 것을 의미하며, 따라서

(의존1) 만약 철수가 돌을 던지지 않았더라면 유리창은 깨지지 않았을 것이다.

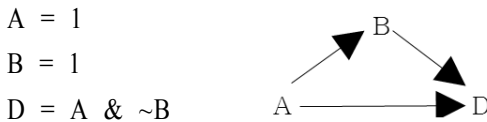
<sup>9)</sup> 김성수 (2012), p. 272.

가 참이라는 것을 의미한다. 활성 경로 이론에 따르면 E'에서 W는 C에 반사실적으로 의존한다. 이 말은 C로부터 W로 이어지는 활성 경로가 존재한다는 것을 의미하므로, 이 경우 철수가 돌을 던진 사건은 유리창이 깨진 사건의 원인이 된다. 반면 C를 실제로 일어난 값인 1로 고정하고 Y의 값을 0으로 바꿨을 때에는  $YH = 0$ ,  $CH = 1$ , 그리고  $W = 1$ 이 되므로

(의존2) 만약 영희가 돌을 던지지 않았더라면 유리창은 깨지지 않았을 것이다.

는 거짓이 되고, 이 말은 Y로부터 W로 이어지는 활성 경로가 존재하지 않는다는 것을 의미하므로 영희가 돌을 던진 사건은 유리창이 깨진 사건의 원인이 아니다.

(전이성)의 경우 암살자가 커피에 독약을 넣는 사건을 변수 A라고, 보디가드가 커피에 해독제를 넣는 사건을 변수 B라고, 보스가 커피를 마시고 죽는 사건을 변수 D라고 했을 때 우리는 다음과 같은 구조식 E'를 가진다.<sup>10)</sup>



이 구조식을 풀면  $D = 0$ 이다. 이제 ‘실제로 벌어진 정황’인 B를 1로 고정시킨 상태에서 A의 값을 0으로 변화시키면 D의 값은 변화하지 않고 여전히 0이다. 이는 D가 A에 반사실적으로 의존하지 않는다는 것을 나타내며 A에서 D에 이르는 활성 경로가 없다는 것

<sup>10)</sup> 김성수 (2012), p. 274.

을 의미한다. 따라서

(의존3) 만약 암살자가 보스의 커피에 독을 넣지 않았더라면 보스는 커피를 마시고 죽었을 것이다.

는 거짓이며, 암살자가 보스의 커피에 독을 넣은 사건은 보스가 커피를 마시고 죽은 사건의 원인이 아니라는 것을 의미한다.

이렇듯 활성화 경로 이론은 (전 선점), (후 선점), (전이성)의 경우에 모두 우리의 직관에 부합하는 답을 내놓는다.

#### 4

하지만 활성화 경로 이론에도 다음과 같은 반례가 존재한다.

(치명적 해독제) 암살자가 보스의 커피에 독을 넣었다. 이를 본 보디가드는 커피에 해독제를 넣었다. 만약 보디가드가 해독제를 넣지 않았더라면 보스는 커피를 마시고 죽었을 것이다. 반면 해독제는 그것만 마셨을 경우 치명적이며 커피에 독이 있지 않았더라면 해독제는 보스를 죽였을 것이다. 독약과 해독제는 그 중 하나만 마시면 치명적이지만 함께 마시면 서로를 중화한다. 보스는 커피를 마셨고 살아남았다.<sup>11)</sup>

암살자가 커피에 독약을 넣는 사건을 변수 A라고, 보디가드가 커피에 해독제를 넣는 사건을 변수 B라고, 보스가 커피를 마시고 죽는 사건을 변수 D라고 했을 때 우리는 다음과 같은 구조식 E'를 가진다.

---

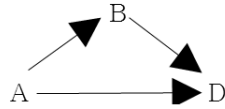
<sup>11)</sup> Menzies (2004), p. 825. 김성수 (2012), pp. 274-275에서 재인용.



$$A = 1$$

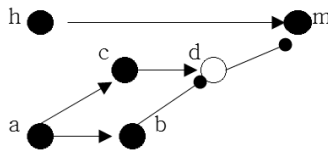
$$B = 1$$

$$D = (A \ \& \ \sim B) \ \vee \ (\sim A \ \& \ B)$$



실제 상황에서  $A = 1$ ,  $B = 1$ , 그리고  $D = 0$ 이다. ‘실제로 벌어진 정황’인  $B$ 를 0으로 고정할 때,  $A$ 를 값은 0으로 바꾸면  $D$ 의 값은 1이 되므로, 보스가 커피를 마시고 살아남는 사건은 암살자가 커피에 독약을 넣는 사건에 반사실적으로 의존하며, 따라서  $A$ - $D$  경로는 활성 경로이다. 이는 암살자가 커피에 독약을 넣는 사건이 보스가 커피를 마시고 살아남는 사건의 원인이라는 뜻인데, 이는 우리의 직관에 반하는 것이다.

두 번째 반례는 네드 홀에 의해 제기되었으며 그는 구체적인 정황의 기술이 아니라 다음과 같은 뉴런 다이어그램을 제시하였다.<sup>12)</sup>



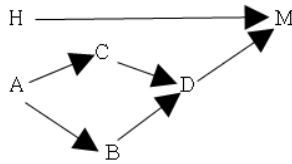
각각의 알파벳 소문자는 사건을 나타내며 속이 채워진 원은 사건이 발생했음을, 그리고 속이 빈 원은 사건이 발생하지 않았음을 나타낸다. 화살표는 인과의 방향을 나타내는데, 화살표가 시작하는 사건이 화살표 끝에 있는 사건을 촉진하는 관계를, 그리고 끝에 화살표가 아니라 속이 검게 채워진 원이 있는 선은 같은 방향으로 사건

<sup>12)</sup> Hall (2007).

을 억제하는 관계를 나타낸다. 위 다이어그램에서 사건 h는 사건 m을 촉진하며, 사건 a는 사건 c와 b를 촉진하고, 사건 c는 d를 촉진한다. 사건 b는 사건 d를 억제한다. 사건 d는 발생하지는 않았지만, 발생했다면 사건 m을 억제했을 것이다.

위 다이어그램에서 사건 a는 두 가지 일을 동시에 행한다. 즉 m을 억제하는 사건 d를 촉진하는 c를 촉진하고 동시에 m을 억제하는 사건 d를 억제하는 사건 b를 촉진한다. 홀에 의하면 바로 이러한 이유 때문에 a가 m의 원인이 아니라고 주장한다. 왜냐하면 “a는 m에 대한 위협을 만들어내는 동시에 제거”<sup>13)</sup>하기 때문이다. 그런데 이를 활성 경로 이론에 적용하면 우리는 다음과 같은 구조식 E'를 가진다.

$$\begin{aligned}
 H &= 1 \\
 A &= 1 \\
 B &= A \\
 C &= 1 \\
 D &= C \ \& \ \sim B \\
 M &= H \ \& \ \sim D
 \end{aligned}$$



구조식을 풀면 D의 실제 값은 0, 그리고 M의 실제 값은 1이다. 이제 C를 실제 값인 1로 고정하고 A = 0으로 변화시키면 B = 0, D = 1, M = 0으로 변화된다. 따라서 A-B-D-M은 활성 경로가 되며, 활성 경로 이론에 의하면 A는 M의 원인이 된다. 이는 우리의 직관에 반하는 것이다.

<sup>13)</sup> Hall (2007), p. 120.

## 5

이제 김성수 교수가 위의 반례를 어떻게 반박함으로써 활성 이론을 변호하는지 살펴보자. 먼저 그는 홀의 반례의 경우  $D = A \& (C \& \sim B)$ 가 아니라  $D = C \& \sim B$ 라는 것을 강조한다. 그렇다면 위의 다이어그램을 구체적인 예시로 만들어 보면 다음과 같아져야 한다고 그는 주장한다.

(독이 이미 들어 있는 커피에 넣은 촉매제) 암살자가 커피에 독을 넣는다. 이를 본 조력자는 이미 독이 들어 있는 커피에 촉매제를 넣고, 이(암살자가 커피에 독을 넣는 것)를 본 보디가드는 커피에 해독제를 넣는다. 커피에 넣은 독이 효력을 발휘하기 위해서는 촉매제가 필요하지만 해독제는 그렇지 않다. 해독제와 촉매제는 그것만 마시든 함께 마시든 무해하다. 만약 보디가드가 해독제를 넣지 않았더라면 보스는 커피를 마시고 죽었을 것이다. 실제로 보스는 커피를 마셨고 살아남았다. 덕분에 보스는 해설자의 도움으로 오전 미팅을 성공적으로 수행하였다.

김성수 교수는, 위의 구조식에서 C는 단지 조력자가 커피에 촉매제를 넣는 것이 아니라 C가 이미 독이 들어 있는 커피에 촉매제를 넣는 것이 되어야 하며, 이 때 이제 C를 실제 값인 1로 고정하고  $A = 0$ 으로 변화시키는 경우는 “이미 독이 들어 있는 커피에 촉매제를 넣었다는 사실을 고정하는 동시에 암살자가 커피에 독을 넣지 않는 상황은, 실제 상황과는 아주 동떨어져있는, 생각하기 어려운 상황”이라고 주장한다.<sup>14)</sup> 김성수 교수의 이 주장은 사건 A가 0이 되는 경우 C가 1로 고정되는 경우를 우리가 상상 혹은 고려해 보

<sup>14)</sup> 김성수 (2012), p. 283.

는 것은 불가능하거나 매우 어려운 일로서 쉽게 B, D, 그리고 M의 값을 구하기 어렵다는 것이다. 이는 김성수 교수가 홀의 다이어그램이 활성 경로 이론에 대한 적절한 반례가 되기 힘들다고 반박하는 것으로 해석된다.

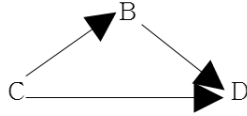
(치명적 해독제) 반례에 대해서 김성수 교수는 위의 내용을 바탕으로 다음을 고려해 보라고 한다(필자는 이 경우를 (독이 이미 들어있는 커피와 치명적 해독제)라고 부르겠다).<sup>15)</sup>

- B: 보디가드가 커피에 치명적 해독제를 넣음.
- C: 커피에 이미 독이 들어있음.
- D: 보스가 죽음.

$$C = 1$$

$$B = C$$

$$D = (C \ \& \ \sim B) \ \vee \ (\sim C \ \& \ B)$$



D의 실제 값은 0이다. 실제로 일어난 정황인 B = 1로 고정했을 때 C를 0으로 변화시키면 D = 1로 변화되므로 C-D 경로는 활성 경로이며 따라서 C는 D의 원인이다. 즉, 커피에 이미 독이 들어있는 사건은 보스가 살아남은 사건의 원인이다. 그리고 이것은 우리의 직관과도 부합한다. 김성수 교수는 이를 바탕으로 다음과 같이 주장한다.

그런데 커피에 독이 들어있다는 것이 보스가 살아남은 것의 원인이라는 주장에 문제가 없다면 암살자가 커피에 독을 넣은 것이 보스가 살아남은 것의 원인이라는 주장 역시 언뜻 생각하는 것보다 더 설득력이 있어 보인다. 만약 커피에 독이 있는 것이 보스가 살아남은 것의 원인이라는 직관과 암살자가 커피에 독을 넣은

---

<sup>15)</sup> 김성수 (2012), p. 286.

것은 보스가 살아남은 것의 원인이 아니라는 직관 모두를 갖고 있다면, 이 두 직관을 동시에 유지하는 것이 정당화될 수 있는지 생각해볼 필요가 있다.<sup>16)</sup>

이는 (치명적 해독제)의 경우에 암살자가 커피에 독을 넣은 것이 보스가 살아남은 것의 원인이라고 주장하는 데에 아무런 문제가 없을 수 있다는 주장, 혹은 적어도 커피에 독이 있는 것이 보스가 살아남은 것의 원인이라고 주장하면서 암살자가 커피에 독을 넣은 것은 보스가 살아남은 것의 원인이 아니라고 주장하는 것에는 중대한 문제가 있다고 반박하는 것으로 해석된다.

## 6

활성 이론에 대해서 제시된 반례에 대한 김성수 교수의 반박은 다음과 같이 정리될 수 있다.

(1) (치명적 해독제) 반례에 대한 김성수 교수의 변호: (독이 이미 들어있는 커피와 치명적 해독제)의 경우에는 커피에 독이 있는 것이 보스가 살아남은 것의 원인이라고 주장하는 동시에, (치명적 해독제)의 경우에는 암살자가 커피에 독을 넣은 것은 보스가 살아남은 것의 원인이 아니라고 주장하는 것은 정당화되지 않는다.

(2) 홀의 반례에 대한 김성수 교수의 변호: 실제 일어난 정황을 정확하게 기술하고 고정했을 때 원인과 결과에 대한 판단이 불가능하거나 매우 판단하기 힘든 사례이다.

김성수 교수는 이와 같이 두 가지 반례에 모두 문제가 있음을 지적함으로써 활성 경로 이론을 지지하고 변호하고자 한다. 필자는

---

<sup>16)</sup> 김성수 (2012), p. 286.

이 각각에 대해 반론을 제기하고자 한다.

먼저 (치명적 해독제)에 대한 김성수 교수의 반박을 보자. 여기서의 반박은 (독이 이미 들어있는 커피와 치명적 해독제)의 경우에는 커피에 독이 있는 것이 보스가 살아남은 것의 원인이라고 주장하는 동시에, (치명적 해독제)의 경우에는 암살자가 커피에 독을 넣은 것은 보스가 살아남은 것의 원인이 아니라고 주장하는 것은 정당화되지 않는다는 것이다. 필자가 보기에 이것은 (독이 이미 들어있는 커피와 치명적 해독제)의 경우에는 커피에 독이 있는 것이 보스가 살아남은 것의 원인이라는 것을 일단 받아들이면, 커피에 독이 들어 있는 것의 원인은 누군가가 커피에 독을 넣은 것이므로, 당연히 누군가가 커피에 독을 넣은 것 역시 보스가 살아남은 것의 원인이 되어야 하지 않은가라는 주장으로 보인다. 그리고 이로부터 (독이 이미 들어있는 커피와 치명적 해독제)에서 누군가가 커피에 독을 넣은 것이 보스가 살아남은 것의 원인이라는 것을 받아들인다면, (치명적 해독제)에서도 암살자가 커피에 독을 넣은 것이 보스가 살아남은 것의 원인으로 받아들이는 것이 맞다는 주장을 끌어내는 것으로 보인다.

그러나 이것은 인과의 전이성을 전제할 때 가능한 논변이며, 따라서 필자가 보기에 김성수 교수의 (치명적 해독제)에 대한 반박은 인과의 전이성을 전제하고 있다. 그러나 앞서 3절에서 보았듯이 김성수 교수가 변호하는 활성 경로 이론은 인과의 전이성을 받아들이지 않으며, 따라서 김성수 교수도 이를 받아들이지 말아야 하는 것으로 보인다. (치명적 해독제)에 대한 김성수 교수의 반박은 그가 변호하고자 하는 이론이 부인하는 전제를 받아들인 상태에서 행해지기 때문에 잘못된 반박이다.

김성수 교수의 (치명적 해독제)에 대한 반박을 다른 식으로 보면, 다음과 같이 해석될 수도 있다. 위의 (독이 이미 들어있는 커피와

치명적 해독제)를 다시 보자.

B: 보디가드가 커피에 치명적 해독제를 넣음.

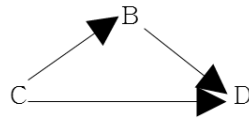
C: 커피에 이미 독이 들어있음.

D: 보스가 죽음.

$$C = 1$$

$$B = C$$

$$D = (C \ \& \ \sim B) \ \vee \ (\sim C \ \& \ B)$$



위에서 우리는 이미 커피에 독이 들어있는 것은 보스가 살아남은 것의 원인이라는 것을 보았다. 또한 C의 값을 실제 값인 1로 고정하고  $B = 0$ 으로 변화시키면  $D = 1$ 로 변화된다. 이것은 보디가드가 커피에 치명적 해독제를 넣은 것 역시 보스가 살아남은 것의 원인이 된다는 것을 보인다. 즉, (독이 이미 들어있는 커피와 치명적 해독제)의 경우 커피에 독이 들어있는 것과 보디가드가 커피에 치명적 해독제를 넣은 것 모두가 보스가 살아남은 것의 원인이 된다. 이를 통해 김성수 교수가 주장하는 것은 (치명적 해독제)에서도 마찬가지로 암살자가 보스의 커피에 독을 넣은 것과 보디가드가 커피에 해독제를 넣는 것 모두 원인으로 보는 것이 마땅하다는 것으로 해석될 수 있다.

그러나 필자가 보기에 우리의 직관은 이 두 경우에서 다른 답을 내는 것으로 보인다. 즉, 우리의 직관에 의하면, (독이 이미 들어있는 커피와 치명적 해독제)의 경우 커피에 독이 들어있는 것과 보디가드가 커피에 치명적 해독제를 넣은 것 모두가 보스가 살아남은 것의 원인인데 반해, (치명적 해독제)에서는 보디가드가 커피에 해독제를 넣는 것만이 원인이다. 그리고 설사 (치명적 해독제)에서도

보디가드가 커피에 넣은 해독제는 그것만 먹으면 치명적이므로 이것 이외에 또 다른 원인이 필요하다는 것을 인정하더라도, 여기에서 필요한 또 다른 원인은 (독이 이미 들어있는 커피와 치명적 해독제)의 경우와 마찬가지로 ‘커피에 독이 들어있는 것’이지 ‘암살자가 보스의 커피에 독을 넣는 것’은 아닌 것으로 보인다.

이제 흠의 반례에 대한 김성수 교수의 반박을 보자. 이 반박에서 아쉬운 점은 김성수 교수가 택하고 있는 방법이 앞으로 나올 수 있는 비슷한 유형의 다른 반례에 모두 적용될 수 있을 만큼 일반적이지 않다는 것이다. 즉, 김성수 교수의 반박이 흠의 다이어그램 자체에 대한 반박이라기보다는 이 다이어그램을 예시하는 한 구체적인 사례에 대한 반박에 그칠 수 있다는 것이다. 그렇다면 이 다이어그램에 상응하는 다른 반례가 얼마든지 새로이 만들어질 수 있다. 이를 보이기 위해 필자는 다음의 예시를 들어 보겠다.

(명령하는 암살자) 암살자가 조력자에게 보스의 커피에 독을 넣으라고 명령한다. 이 명령을 듣고 조력자는 커피에 독을 넣는다. 한편 암살자가 조력자에게 명령하는 것을 본 보디가드는 커피에 해독제를 넣는다. 해독제는 그것만 마시는 경우 무해하다. 실제로 보스는 커피를 마셨고 살아남았다. 덕분에 보스는 해설자의 도움으로 오전 미팅을 성공적으로 수행하였다.

이를 구조식으로 나타내면 다음과 같다.

- H: 해설자가 도움.
- A: 암살자가 명령을 내림.
- C: 조력자가 독을 넣음.



B: 보디가드가 해독제를 넣음.

D: 보스가 죽음.

M: 보스가 오전 미팅을 성공적으로 수행함.

H = 1

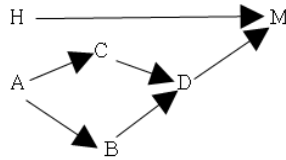
A = 1

B = A

C = A

D = C & ~B

M = H & ~D



구조식을 풀면 D의 실제 값은 0, 그리고 M의 실제 값은 1이다. 이제 C를 실제 값인 1로 고정하고 A = 0으로 변화시키면 B = 0, D = 1, M = 0으로 변화된다. 따라서 A-B-D-M은 활성화 경로가 되며, 활성화 경로 이론에 의하면 A는 M의 원인이 된다. 즉, 암살자가 명령을 내린 사건은 보스가 살아남은 사건의 원인이 된다. 그러나 이는 우리의 직관에 반하는 것이다.

이 경우 김성수 교수가 (독이 이미 들어있는 커피와 치명적 해독제)에 대해 지적한 바와 같이 조력자가 커피에 독을 넣는데 명령을 내리지 않은 상황은 실제 상황과는 아주 동떨어져있는, 생각하기 어려운 상황이라고는 할 수 없다. 왜냐하면, 암살자가 독을 넣지 않았는데도 커피에 독이 이미 들어있는 상황은 거의 모순이 되는, 따라서 거의 생각할 수 없는 상황이지만, 암살자가 명령을 내리지도 않았는데 조력자가 커피에 독을 넣는 상황은 모순이 되는 상황은 아니므로, 혼하지는 않다고 하더라도 얼마든지 일어날 수 있는 상황이기 때문이다. (명령하는 암살자)는 홀의 다이어그램에 준하는 그리고 김성수 교수의 반박이 적용되지 않는, 활성화 경로 이론에 대

한 반례이다.

필자의 반례에 대해 C를 실제 값으로 고정할 때, 조력자가 커피에 독을 넣는 사건을 그대로 복사해와야 하며 이 경우 이미 조력자가 암살자로부터 명령을 받은 사실도 그대로 복사해와야 한다는 반론이 있을 수 있다. 그러나 조력자가 커피에 독을 넣는 사건을 그대로 복사할 때에는, 그 이전에 일어난 ‘명령받은 사실’ 자체를 복사하는 것이 아니라 조력자가 가진 ‘암살자로부터 명령을 받았다는 믿음’만 그대로 복사해오면 되는 것이다. 그 이전에 암살자의 명령이 실제로는 없었음에도 조력자가 ‘암살자로부터 명령을 받았다는 믿음’을, 즉, 거짓인 믿음을 가지는 경우는 얼마든지 있을 수 있다.

## 7

본 논문에서 필자는 활성 경로 이론에 제시된 반례를 김성수 교수가 어떻게 반박하였는지 살펴보고 그에 대한 재반박을 하였다. 필자의 재반박의 핵심은 첫째, (치명적 해독제)에 대한 김성수 교수의 반박은 그가 변호하고자 하는 이론이 부인하는 전제를 받아들인 상태에서 행해지기 때문에 활성 경로 이론에 대한 변호로서 적절하지 않다는 것, 그리고 우리의 직관은 (독이 이미 들어있는 커피와 치명적 해독제)의 경우와 (치명적 해독제)의 경우에 다른 답을 내린다는 것이다. 둘째, 홀에 대한 김성수 교수의 반박은 홀의 다이어그램을 예시하는 한 사례에 대한 반박에 그치며 따라서 얼마든지 다른 반례가 생길 수 있다는 점에서 미흡하다는 것이다. 그리고 실제로 필자는 김성수 교수의 반박을 피할 수 있는 다른 반례를 제시하였다.

필자는 활성 경로 이론이 매우 흥미로우면서도 국내에 잘 알려

저 있지 않은 이론이라고 생각하며, 따라서 본 논문에서의 문제 제시가 활성화 경로 이론에 대한 폐기가 아닌, 활성화 경로 이론 자체에 대한 보다 많은 논의를 불러일으키는 계기, 더 나아가서는 이러한 문제를 극복할 수 있는 더 나은 활성화 경로 이론, 혹은 더 나은 인과이론을 모색해보는 계기로 이어지기를 바라는 마음이다.

## 참고문헌

- 김성수 (2012), “반사실적 인과론과 인과 구조식 접근법: 활성화 경로 이론의 재검토”, 『철학』 제113집, pp. 259-291.
- Hall, Ned (2007), “Structural Equations and Causation”, *Philosophical Studies*, 132(1), pp. 109-136.
- Hitchcock, Christopher (2001a), “The Intransitivity of Causation Revealed in Equations and Graphs”, *Journal of Philosophy*, 98, pp. 273-299.
- Hitchcock, Christopher (2001b), “A Tale of Two Effects”, *Philosophical Review*, 110, pp. 361-396.
- Lewis, David (1973), “Causation”, *Journal of Philosophy*, 70, pp. 556-567.
- Lewis, David (1986), “Postscripts to “Causation””, in D. Lewis, *Philosophical Papers*, Vol. 2, New York: Oxford University Press, pp.172-213.
- Lewis, David (2000), “Causation as Influence”, *Journal of Philosophy*, 97, pp. 182-197.
- Menzies, Peter (2004), “Causal Models, Token Causation, and Processes”, *Philosophy of Science*, 71, pp. 820-832.

이화여자대학교 스크랜튼학부

Scranton Honors Program, Ewha Womans University  
seahwak@ewha.ac.kr

---

## Objections to Sungsu Kim's Defense of the Active Route Account

Seahwa Kim

---

In his paper "Structural Equations Approach to Token Causation: The Active Route Account Revisited" Professor Sungsu Kim defends the active route account. The active route account is the theory of causation which overcomes counterexamples to the counterfactual theories of causation, while maintaining the counterfactual theorist's essential intuition that an effect depends counterfactually on a cause. Unfortunately, there are counterexamples to the active route account itself. Professor Sungsu Kim attempts to defend the active route account by rebutting those counterexamples. In this paper, I argue that his defense of the active route account is not successful.

Key Words: causation, counterfactual theories of causation, active route account, counterexample, Sungsu Kim