

도움 안 되는 반사실적 논변

최 훈 (삼척대)

주제분
류

형이상학, 심리철학

주
요
어

반사실적 논변, 정신 인과, 부수현상론, 인과적 힘, 익숙함

요
약
문

데이비드슨의 무법칙적 일원론은 정신 부수현상론을 함축한다는 비판을 받는다. 그 비판을 벗어날 수 있는 한 가지 방법으로서 데이비드슨의 엄격한 법칙을 반사실적 조건문으로 바꾸어 인과 관계를 설명함으로써 부수현상론을 피하려는 시도가 제시된다. 이 논문에서는 일단 어떤 속성이 인과적으로 힘이 있음을 반사실적 조건문을 통해 정의하려는 시도 자체에 매력이 있음을 인정한다. 그래서 그 정의를 계속 세련되게 만들어 그 중 가장 낮다고 생각되는 정의에 도착하겠다. 그러나 정신 속성은 그 정의를 만족하지 못하고 따라서 반사실적 논변으로도 정신 속성에 인과적 힘이 있음이 보여질 수 없다고 주장하겠다. 그 과정에서 익숙함은 인과적 효력을 따질 때 중요한 고려 기준이 아님을 지적할 것이다.

1. 머리말

데이비드슨(Donald Davidson)은 현재의 정신 인과 논의를 촉발했다고 할 수 있는 몇 편의 논문들(Davidson 1980a, 1980b, 1980c)에서 아래와 같은 세 원리들로부터 나오는 의견상의 모순을 어떻게 해소할 수 있을까 하는 고민 끝에 이른바 무법칙적 일원론을 내놓는다.

- (1) [인과적 상호작용의 원리] 모든 정신적 사건은 물리적 사건과 인과적으로 상호작용한다.
- (2) [인과 관계의 법칙론적 원리] 두 사건 사이에 인과 관계가 있으면 두 사건을 포섭하는 엄격한 법칙이 틀림없이 있다.
- (3) [정신적인 것들의 무법칙성] 엄격한 심리물리적 법칙은 없다.

(1)과 (2)가 결합하여 적어도 어떤 정신적 사건들은 법칙에 근거해서 예측되고 설명될 수 있다는 것을 함축한다. 하지만 (3)의 정신적인 것들의 무법칙성 원리는 이를 부정한다. 이게 데이비드슨의 세 원리들로부터 나오는 의견상의 모순이다. 그런데 그는 이 세 원리가 모두 참으로 받아들여질 수 있는 원리들이기 때문에 이 의견상의 모순을 해소할 수 있는 적절한 이론이 있을 것이라고 생각하고, 바로 자신의 무법칙적 일원론이 실제로 그 일을 해 낸다고 주장한다.

그런데 데이비드슨의 무법칙적 일원론에 대해서 그것은 부수현상론을 함축한다는 비판이 제기된다. 그 주장은 대체로 다음과 같은 논변으로 정리할 수 있다.¹⁾

- (4) 오직 법칙적 속성만이 인과적이다. ((2)에 의해서)
- (5) 정신 속성은 법칙적 속성이 아니다. ((3)에 의해서)
- (6) 따라서 어떤 정신 속성도 인과적 속성이 아니다. 곧 정신 속성은 부수현상적이다.

데이비드슨의 두 번째 원리인 인과 관계의 법칙론적 원리([2])는 두 사건 사이에 인과 관계가 있으면 두 사건을 포섭하는 엄격한 법칙이 틀림없이 있다고 말한다. 따라서 속성도 어떤 법칙에 포섭되어야만 인과적 속성이라고 말할 수 있을 것이다. 데이비드슨의 설명에서 기술을 속성으로 바꾸어 생각한다면, 정신 속성과 물리 속성 사이의 관련은 무법칙적이라고 했다([3]). 그러므로 어떤 정신 속성도 인과적 효력이 없고 따라서 무법칙적 일원론은 부수현상론을 함축한다는 것이다.²⁾

데이비드슨의 무법칙적 일원론이 부수현상론인지 아닌지에 관한 논의는 우리에게 다음과 같은 질문을 던져 주어 정신 인과 논의를 촉발시켰다. 곧 엄격한 법칙에 어울리지 않는 정신 속성이 어떻게 해서 인과적인 효력이 있겠는가? 이 질문에 대해서 정신 속성에 인과적 효력을 돌려주고 싶은, 곧 부수현상론을 피하고 싶은 비환원적 물리주의자들은 두 가지 방법을 쓸 수 있다. 첫째는 데이비드슨의 주장을 받아들여, 사실 정신 속성도 데이비드슨이 요구하는 엄격한 법칙에 포섭된다고 주장하는 길이다. 둘째는 정신 속성에 인과적 효력이 있기 위해서는 데이비드슨이 요구하는 것처럼 엄격한 법칙에 포섭될 필요가 없다고 말하는 것이다. 대부분의 비환원적 물리주의자들은 두 번째 길을 걷는다.

이 두 번째 길도 더 자세하게 세 가지로 갈래로 나눌 수 있다. 데이비드슨이 요구하는 인과 관계의 법칙적 요구를 완화시킨다는 점에서는 똑같은데 그 방법에서 조금씩 다르다. 첫째로

1) 다음과 같은 문헌들에서 그런 비판을 하고 있다. Hess(1981), Honderich(1983), Sosa(1984), Kim(1984), Skillen(1984), Johnston(1985), Follesdal(1985), Stoutland(1985), Horgan(1989), McLaughlin(1989), Kim(1993c).

2) Davidson의 무법칙적 일원론과 그것이 부수현상론을 함축하느냐는 주제에 대해서는 최훈(2004)를 참조하라.

포더(Jerry Fodor)처럼 단서조항 구절이 들어간 엄격하지 않은 법칙을 대안으로 내놓는 이들이 있다.³⁾ 둘째로 르포(Ernest LePore), 로워(Barry Loewer), 호건(Terence Horgan) 등은 반사실적 의존을 엄격한 법칙의 대안으로 내놓는다.⁴⁾ 마지막으로 데이비드슨 스스로는 생각하는 원인들에서 심신 수반을 대안으로 삼는다.⁵⁾

나는 이 논문에서 그 중 데이비드슨의 엄격한 법칙을 반사실적 조건문(counterfactual conditional)으로 대신하여 인과 관계를 설명함으로써 정신 부수현상론을 피하려는 시도를 살펴볼 것이다. 나는 어떤 속성이 인과적으로 힘이 있음을 반사실적 조건문을 통해 정의하려는 시도 자체에 매력이 있음을 인정한다. 그래서 그 정의를 계속 세련되게 만들어 그 중 가장 낫다고 생각되는 정의에 도착하겠다. 그러나 정신 속성은 그 정의를 만족하지 못하고 따라서 반사실적 논변으로도 정신 속성에 인과적 힘이 있음이 보여질 수 없다고 주장하겠다. 그 과정에서 익숙함은 인과적 효력을 따질 때 중요한 고려 기준이 아님을 지적할 것이다.

2. 반사실주의자의 인과적 효력 정의

반사실적 조건문에 의한 인과 관계를 간단하게 표현하면 다음과 같다.

(7) c가 일어나지 않으면 e가 일어나지 않는 꼭 그 경우, c는 e가 일어나는 데 인과적으로 힘이 있다.

자동차 열쇠를 돌리지 않았다면 자동차의 시동이 걸리지 않았을 것이다. 따라서 자동차 열쇠를 돌림은 자동차의 시동이 켜지는 데 인과적으로 효력이 있다. 방의 커튼을 젖히지 않았다면 햇빛이 들어오지 않았을 것이다. 그러므로 방의 커튼을 젖힘은 햇빛이 들어오는 데 인과적으로 효력이 있다. 반사실적 조건문은 이렇게 직관적으로 원인, 결과 관계를 설명해 주는 좋은 틀이 된다. 두 사건 또는 속성 사이에서 성립하는 법칙을 통해 원인, 결과 관계를 설명하려는 시도는 쉽지 않다. “그 다리는 용량을 초과하는 트럭이 진입했기 때문에 무너졌다”라는 예처럼 인과적으로 효력이 있는 원인, 결과 속성들이 (또는 사건들이) 엄격한 법칙으로 표현되지 않는 경우가 많기 때문이다. c가 일어나지 않는 경우 e가 일어나지 않지만 하면 되며 그 때 e가 일어나지 않는다는 것을 함축하는 법칙이 있어야 하는 것은 아닌 인과 관계를 생각해 보는 것은 손쉬운 일이다. 그러므로 (7)을 정신 인과에도 적용할 수 있다. 곧 내가 냉장고에 맥주가 있다는 믿음이 없었다면 나는 냉장고로 가지 않았을 것이다. 따라서 냉장고에 맥주가 있다는 나의 믿음과 내가 냉장고로 간 행동 사이에는 인과 관계가 있다. 그렇다면 나의 믿음이라는 정신 속성이 냉장고로 가는 행동을 불러일으키는 데 효력이 있는 원인이라고 말할 수 있을 것이다.

부수현상론을 극복하기 위해 인과 관계를 반사실적 조건문으로 읽어야 한다고 생각하는 이들을 편의상 반사실주의자라고 불러보자. 그리고 그들의 논변을 반사실적 논변이라고 해 보자. 잘 하면 부수현상론을 피할 수 있을 것 같은 정의 (7)은 몇 가지 점에서 세련되게 표현

3) Fodor(1990). 나는 최훈(2000)의 4장에서 이 첫 번째 길에 대해 논의하였다. 거기서 Fodor가 결국 첫번째 길과 두번째 길을 다 걸은 셈이라고 곧 지적했다.

4) Heil and Mele(1991), LePore and Loewer(1987, 1989), Horgan(1989)를 보라. 한편 Burge(1993)은 정신적인 것이 물리적인 것의 ‘기저에 있다’(underlie)의 뜻을 반사실적 관계로 정의내리고 있다(98쪽). 정신 인과와 관련된 반사실적 의존 분석에 대한 논의는 다음을 보라. Kim(1996): 139-144쪽, Kim(1995): 133-4쪽, Tomberlin(1995), Braun(1995)를 보라.

5) Davidson 외에 심신수반을 대안으로 내놓는 이는 초기의 Jaegwon Kim과 그리고 Stephan(1997)이 있다.

되어야 한다. 우선 (7)에서는 c와 e를 사건과 속성을 구분하지 않고 무차별적으로 썼다. 그러나 그러면 같은 사건에서도 인과적으로 관련 있는 속성과 관련 없는 속성이 구분됨을 잡아내지 못한다. 한 소프라노가 “깨져라!”라는 의미의 고음을 소리 냈고 그래서 유리컵이 깨졌다고 해 보자. 이 사건은 다음과 같이 반사실적 조건문을 따로 표현할 수 있다.

- (8) 소프라노의 고음이 “깨져라!”가 아니었다면 유리컵은 깨지지 않았을 것이다.
- (9) 소프라노의 고음의 진폭이 달랐다면 유리컵은 깨지지 않았을 것이다.

(8)은 거짓이지만 (9)는 참이다. 소프라노는 “깨져라!”가 아니라 “유리컵!”이라고 고음을 냈더라도 진폭과 음조만 그대로라면 유리컵은 여전히 깨졌을 것이다. 그러므로 반사실주의자들은 한 사건이 다른 사건을 일으켰다고 할 때 무슨 자격으로(qua), 무엇 때문에(in virtue of) 일으켰는지 꼭 물어보아야 한다. 차의 열쇠가 은색의 자격으로서 차의 시동을 걸리게 한 것은 아니고, 장인환의 총 발사가 큰 총소리의 자격으로서 스티븐슨을 저격한 것은 아니다. 차의 열쇠가 금색이더라도 차의 시동은 걸렸을 것이고, 장인환의 총에 소음기가 달렸더라도 스티븐슨은 죽었을 것이다. 데이비드슨의 무법칙적 일원론에서 원인이 되는 사건은 개별자로서의 사건이다. 그래서 거기서도 그 사건이 정신적인 것으로서 사건인지 물리적인 것으로서 사건인지 하는 점이 문제가 된다.⁶⁾

이제 F와 G가 각각 c와 e가 사례화하는 속성이라고 할 때, 반사실주의자들은 (7)을 다음과 같이 바꾼다.

- (10) 다음과 같은 조건을 만족하면 그리고 오직 그 경우에만 c가 e를 일어나게 하는데 c의 속성 F가 인과적으로 힘이 있다: c가 F가 아니면 e가 안 일어날 것이다.(Heil and Mele 1991: 68쪽. 원문의 c*, b를 이 글의 논의와 맞추기 위해 c, e로 바꾸었다.)

나는 반사실주의자들의 이 정의가 바로 두 원인의 문제, 다시 말해서 같은 결과가 두 원인에 의해 과잉 결정된다는 문제에 걸리기 때문에 날자마자 떨어진다고 생각한다. (10)은 정신 인과에도 그대로 적용된다. M이 정신 속성이고 P가 그것을 실현하는 물리 속성이라고 하자. 그리고 M은 행동으로 된 결과 B를 일으킨다고 하자. 그때 c가 M이 아니라면 B가 일어나지 않을 것이지만, 또한 P가 아니라도 B가 일어나지 않을 것이다. 따라서 (10)이 맞다면 M과 마찬가지로 P도 결과의 원인이 되고, 그러면 두 원인의 문제에 걸린다.

김재권은 반사실주의자에게 방금 설명한 것처럼 그들의 시도도 인과적 배제의 위협에 속수무책임을 지적하는 방식으로 대응한다(Kim 1995: 133-4쪽; Kim 1990: 42-5쪽). 나는 인과적 배제의 위협이 꽤 세다고 생각하며, 따라서 김재권 식의 대응이 옳다고 생각한다. 그러나 나는 내가 생각하기에는 아주 심각한 문제인 이 점을 일단은 무시하고 반사실주의자들의 주장을 최대한 따라가 보기로 하겠다.

(10)은 여전히 몇 가지 점에서 다듬어야 한다. 우선 c가 ‘자동차 열쇠를 돌림’이란 사건이고 e는 ‘차의 시동이 걸림’이란 사건이라고 할 때, F가 ‘차의 시동이 걸림의 원인이 되는 속성’이라고 해 보자. 그러면 이 F는 위 (10)을 너무 쉽게 만족시켜 인과적으로 힘이 있는 속성이 되고 만다. 반사실주의자들은 이런 허망한 상황이 생기는 것을 막기 위해 (10)에 조건 하나를 더 덧붙여야 한다.

6) 그래서 Terence Horgan은 사건이 특별히 정신적인 것으로서 원인이 될 때를 ‘정신 자격인과’(mental quausation)라는 신조어를 만들어 쓴다(Horgan 1989).

(11) 다음과 같은 조건을 만족하면 그리고 오직 그 경우에만 c가 e를 일어나게 하는데 c의 속성 F가 인과적으로 힘이 있다.

- (a) c가 F가 아니면 e가 안 일어날 것이다.
- (b) c가 F임과 e가 일어남은 논리적으로 그리고 형이상학적으로 독립적이다.⁷⁾

그 다음에 (a)를 다시 손봐야 한다. 위 문장 (9)가 참이고 거기에 나오는 소프라노가 신영옥이라고 하자. 그런데 같은 방에 조수미가 있어서 신영옥의 고음의 진폭은 달랐지만 조수미의 고음의 진폭 때문에 유리컵이 깨졌다고 하면 (9)는 거짓이 된다. 그러면 우리의 예상과 달리 신영옥의 고음이 유리컵이 깨지게 한 원인이 아니게 되어 (11)은 인과적 효력의 검사로 좋지 못하다.⁸⁾ 이런 일을 막으려면 (9)는 다음과 같이 표현해야 한다.

(12) 소프라노의 고음의 진폭이 달랐다면 그 소프라노의 고음은 유리컵이 깨지게 하지 않았을 것이다.

이 점을 반영하기 위해서 (11)은 다시 이렇게 고쳐야 한다.

(13) 다음과 같은 조건을 만족하면 그리고 오직 그 경우에만 c가 e를 일어나게 하는 데 c의 속성 F가 인과적으로 힘이 있다.

- (a) c가 F가 아니면 c는 e가 안 일어나게 할 것이다.
- (b) c가 F임과 e가 일어남은 논리적으로 그리고 형이상학적으로 독립적이다.

이 정의에서 핵심이 되는 조건은 두 말할 것도 없이 반사실적 조건문인 (a)이다.⁹⁾ 그러나 반사실주의자들이 (a)를 구체적으로 어떻게 읽어야 하느냐에 모두 일치하는 것은 아니다. 반사실적 조건문을 읽는 방법은

- (가) 규범-도출적 접근법
- (나) 가능 세계 접근법

7) Braun(1995): 449-450쪽, LePore and Loewer(1997): 635쪽, Kafez(1995): 82 쪽, Horgan(1989): 58쪽은 자세한 설명 없이 조건 (b)를 언급하고 있다. 한편 Ludwig(1994): 338-40쪽, Dardis(1993)는 이 조건에 대해 깊이 있게 논의하고 있다. 특히 Ludwig는 ‘빨강’과 ‘색깔’의 관계까지도 논리적 의존 관계로 보아 거기에 근거해 기능주의 이론을 비판하고 있다.

8) 이 경우는 조수미의 고음이 신영옥의 고음을 선취한 거라고 말할 수 있다. (Braun 1995: 450-1쪽 참조.) 결국 (21)번에서 말할 조건도 선취를 배제하는 조건이다. 다만 지금 경우는 서로 다른 사건들 사이에서 일어나는 선취이지만, 거기에서는 한 사건의 다른 속성들 사이에서 일어나는 선취가 중요하게 거론될 것이다.

9) (a)가 여전히 덜 다듬어졌다고 주장하는 이가 있다. 다음 문장을 보자.

(*) 소프라노의 고음에 ‘소프라노와 유리컵 사이에 방음벽이 없음’이라는 속성이 없다면 소프라노의 고음은 유리컵이 깨지게 하지 않았을 것이다.

이 문장은 참이다. 그렇다면 소프라노의 고음에서 소프라노와 유리컵 사이에 방음벽이 없음은 유리컵이 깨지게 하는데 인과적으로 효력이 있다고 말해야 할 것이다. 심각한 문제는, 이와 마찬가지로 방식으로 ‘유리컵에 담요가 덮여 있지 않음’, ‘유리컵을 누군가가 잡고 있지 않음’ 등과 같은 무수한 부정적 속성이나 또 ‘뉴턴의 제1 법칙이 성립함’, ‘소프라노의 고음이 일어난 1초 후에 세상은 존재함’과 같은 무수한 보편 속성도 유리컵이 깨짐에 인과적인 효력이 있다고 말해야 할 것이다. 그래서 Braun(1995)는 속성 F는 이상하지 않은 자연스러운 속성에만 한정해야 한다고 주장한다(458쪽). 그런데 내가 보기에는 위와 같은 ‘이상한’ 속성들에 인과적인 효력이 있다는 게 왜 문제가 되는지 모르겠다. 소프라노와 유리컵 사이에 방음벽이 없음은 소프라노의 고음이 유리컵을 깨게 하는 데 인과적인 효력이 있는 것 중 분명히 하나다. 다만 우리의 주된 관심이 아닐 뿐이다. 인과적 관련과 인과적 설명을 구분한다면 이진 좋은 인과적 설명은 못 되지만 인과적 관련은 분명히 있다. 그리고 ‘자연스러운 속성’이라는 개념은 얼마나 자의적인가? 이 논문 4절에서 지적할 ‘익숙함’의 문제일 뿐이다.

두 가지가 있다(Kim 1996: 141-2쪽; 김기현 1998: 77-83쪽 참조). 규범-도출적 접근법에서 반사실적 조건문이 참이 되는 경우는 그 반사실적 조건문의 후건이 법칙과 보조 조건들의 도움을 받아 전건에서 도출될 때이다. 이를테면 다음 반사실적 조건문을 보자.

(14) 만약 이 성냥을 그었다면, 불이 붙었을 것이다.

이 조건문은 참인데, 왜냐하면 “이 성냥을 그었다”라는 전제에 “산소가 있을 때 쪼이지 않은 성냥을 그으면 성냥에 불이 붙는다”라는 법칙과 “이 성냥은 쪼이지 않았다”, “산소가 있다”라는 보조 조건이 합해질 때 “불이 붙었다”라는 결론이 도출되기 때문이다.

그러나 앞서 인과 관계의 법칙에 의한 분석은 인과적 관련과 법칙적 관련이 관계가 없을 수 있기 때문에 문제점이 있음을 지적했다. 그렇다면 규범-도출적 접근법은 훌륭한 대안이 될 수 없다. 반사실적 조건문의 분석에 법칙이 필요로 한다는 점은 문제를 다시 법칙에 의한 분석으로 되돌리기 때문이다. 더구나 전건이 정신 속성 또는 사건인 정신 인과의 경우에는 정신 법칙의 존재가 의심스럽기 때문에 문제를 더 깊게 끌고 나아갈 수가 없다.

그렇다면 반사실적 조건문에 가능 세계 접근법으로 다가서는 수밖에 없다. 이 접근법에서는, 가령 “P이면 Q이다”라는 조건문은 P가 참인 가능 세계에서 Q도 참인 오직 그 경우 참이라고 할 수 있다. 이때 가능 세계의 범위를 주의해서 잡아야 하는데, 우리가 사는 현실 세계보다는 넓고 논리적으로 가능한 모든 세계보다는 좁은, 현실 세계에 가장 가까운 가능 세계이다. 다음 반사실적 조건문을 보자.

(15) 만약 내가 새라면 하늘을 날 것이다.

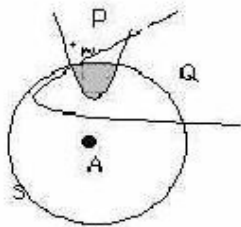


그림 4 반사실적 조건문의 가능 세계적 해석

이 조건문을 실질적 조건문(material conditional)으로 해석하면 (전건이 거짓이 되므로) 너무 쉽게 만족되고, 엄격한 조건문(strict conditional)으로 해석하면 만족하기가 너무 어렵다. 그래서 반사실적 조건문으로 해석하는 게 가장 낫다. 위 조건문은 반사실적이므로 “내가 새이다”라는 전건은 현실 세계(그림 1에서 A)에서 거짓이기 때문에 그 전건이 참이 되는 가능 세계로 가서 거기서도 내가 하늘을 난다가 참인지 살펴봐야 한다. 그런데 “내가 새이다”가 참이 되는 가능 세계(그림 1에서 P) 중에는 지상의 대기가 공기가 아니라 유리로 채워진 세계도 있을 수 있다(W가 있는 곳). 그 세계에서는 설령 내가 새라고 해도 나는 날 수가 없다. 이런 세계는 논리적으로 얼마든지 가능하긴 하지만, 반사실적 조건문의 참거짓값을 고려할 때 대상이 되지 않는다. 고려해야 할 가능 세계는 내가 새라는 가정만이 현실 세계에 덧붙여지고 현실 세계에서 성립하는 자연 법칙은 대부분 성립하는 세계이다¹⁰⁾(그림

1에서 그림자진 곳). 그런 가능 세계들에서 내가 하늘을 난다가 모두 참이 된다면 (15)는 참이 될 것이다. 그림 1에서 내가 하늘을 나는 가능 세계는 Q의 원호로 표시되어 있다. 그리고 우리가 사는 현실 세계와 가장 가까운 세계는 원 S가 나타내고 있다. P가 성립하는 가능 세계 중에서 우리 세계에 가장 가까운 세계(그림자진 곳)는 모두 Q가 성립하는 가능 세계이기도 하므로 반사실적 조건문 (15)는 참이 된다. 다른 말로 하면 P가 성립하는데 Q가 성립 안 하는 세계가 없는 것은 아니지만(가령 W) 우리 세계와 가장 가까운 세계 안에는 없다.

현재까지 반사실적 조건문을 읽는 방법으로서의 규범-도출적 접근법보다 가능 세계 접근법이 더 나아 보인다. 그림 가능 세계 접근법을 채택하자. 그런데 반사실적 조건문을 가능 세계 접근법으로 읽는 방법도 다시 두 가지가 있다. 위 (13)의 (a) “c가 F가 아니면 c는 e가 안 일어나게 할 것이다”를 가능 세계 접근법으로 읽는다고 해 보자. 그 때도

- (다) c가 애초에 안 일어난 경우
- (라) c가 일어났는데 F가 아닌 경우

로 나눌 수 있다.¹¹⁾ (다)는 반사실적 세계에서 c가 아예 안 일어나기 때문에 c의 속성이라는 것도 있을 수 없고 그래서 c가 e가 안 일어나게 하는 경우를 말한다. 가령

(16) 자동차 충돌이 빠른 속도로 일어나지 않았다면, 사상자가 없었을 것이다

에서 자동차 충돌(c)이라는 사건이 아예 일어나지 않았다고 한다면 (물론 빠른 속도라는 속성[F]도 없게 되고) 사상자가 없었을 것이므로 (16)은 참이다.

존 헤일(John Heil)과 밀(Alfred Mele)은 (13)의 (a)를 (다)처럼 해석하고 있다. 그들은 이렇게 말한다.

우리는 그런 역사가 없는 c가 일어나는 가능 세계를 상상할 수 있고, 그 세계 c에서 e가 일어날 수도 있다. 그러나 대부분의 경우에 그런 세계는 정말 너무 멀리 있다. c의 원인이 되는 역사가 없으면서 우리 세계와 가까운 세계는 c도 역시 없는 세계이다.

그들이 말하는 사건의 역사는 사건의 일종의 속성(F)이라고 생각할 수 있는데, 그러니까 헤일과 밀은 (라)처럼 해석된 세계보다 (다)처럼 해석된 세계가 우리 세계와 더 가깝다고 생각한 것이다.¹²⁾ 그들의 의도야 어찌 됐든 (13)의 (a)의 전건을 (다)처럼 해석하는 것은 심각한 문제를 낳는다. 그렇게 해석하면 전건이 거짓이 되므로 후건에 무엇이 오든 반사실적 조건문은 공허하게 참이 된다. 예를 들어서 위 (16)만 참이 되는 것이 아니라

10) 우리가 고려해야 할 세계(곧 우리 세계와 가장 가까운 세계)를 현실 세계에서 성립하는 자연 법칙이 대부분 성립하는 세계로 제한하는 것만 가지고는 안 된다. 가령 더 많은 사람들이 담배를 끊으면, 더 많은 사람들이 폐암에 안 걸릴 것이라는 반사실적 조건문은 법칙적인 세계 중에서 담배를 끊는 사람이 대부분 석탄 광산에서 일하는 사람이었다면 폐암에 걸리는 비율이 여전히 비슷하게 되므로 거짓이 될 수 있다. 그래서 더 많은 사람이 담배를 끊는다는 것을 제외하고는 실제 세계와 아주 비슷한 세계를 상상하라. 현실 세계에서는 최소한도로만 떠나라. Doring(1998) 참조.

11) 이 구분에 대해서는 Kazez(1995): 73-5쪽에서 자세하게 논의하고 있다. 그 외 LePore and Loewer(1987): 636쪽, Horgan(1989): 주 17에서도 이 구분에 대해 잠깐 언급하고 있다.

12) Heil과 Mele에서 또 중요한 것은 (5)의 (a)를 거짓으로 만드는, 곧 전건은 참이고 후건은 거짓이 되는 가능 세계가 있지만 그 세계는 우리 세계와 너무 멀리 떨어져 있다고 말한 것이다(같은 곳). 이 주장은 ‘너무 먼 세계 논변’으로서 곧 다루겠다.

(17) 자동차 충돌이 빠른 속도로 일어나지 않았다면, 자동차는 서울에 없었을 것이다.

도 역시 참이다. 따라서 (13)에 의해 자동차 충돌이 빠른 속도로 일어남은 사상자가 있음에 인과적으로 관련 있기도 하지만 자동차가 서울에 있음에도 인과적으로 관련 있다. 더 나아가 (17)의 후건에 어떤 사건을 넣어도 그 사건에 인과적으로 관련 있게 된다.

(13)의 (a)를 (다)처럼 해석하면 더 근본적으로는 다시 (7)의 정의로 되돌아 가게 되어 (7)을 기껏 (4)로 세련되게 만든 수고가 무색해지고 만다.¹³⁾ 한편 (7)의 (a)를 (다)처럼 해석해서는 안 되는 이유는 왜 (7)을 가능 세계 접근법으로 읽어야 하는가 하는 점과 연결되기도 한다. (10)에서 반사실적으로 가정한 빠른 속도로 일어나지 않은 자동차 충돌은 현실 세계의 자동차 충돌의 상대역(counterpart)이다. 그러나 자동차 충돌이라는 사건의 동일성을 현실 세계 기준으로만 본다면 빠른 속도로 일어나지 않은 자동차 충돌은 더 이상 우리가 관심 있어 하는 자동차 충돌이 아니다. 따라서 (10)의 전건이 (다)의 해석에서처럼 자동차 충돌이 아예 안 일어난 경우로 읽혀진다. 그러면 역시 전건이 거짓이므로 (11)과 같은 보기를 무수히 만들어내기 때문에 가능 세계 개념을 끌어들여야 한다.

그렇다면 반사실주의자는 (7)의 (a)를 (라)처럼 읽어야 한다. 그리고 중요한 반사실주의자들인 르포와 로워, 그리고 호건은 실제로 그렇게 읽는다.

클라이드가 냉장고로 간 까닭은 냉장고에 맥주가 있다는 그의 생각이라고 할 때 냉장고에 맥주가 있다는 내용을 소유한 생각이 클라이드를 냉장고로 가게 만든 사실과 어떤 관련이 틀림없이 있을 것 같다.

클라이드가 냉장고에 맥주가 있다고 생각하지 않았는데도 냉장고에 쌍둥이-맥주가 있다고 생각하기 때문에 여전히 냉장고로 갈 것인지가 걱정된다. (LePore and Loewer 1989: 176쪽, 190쪽)

르포와 로워는 클라이드의 생각이 냉장고에 맥주가 있다는 내용의 생각이 아니라면 (그리고 대신에 가령 냉장고에 쌍둥이-맥주가 있다는 내용의 생각이어도) 냉장고로 가지 않을 것이라는 반사실적 조건문을 염두에 두고 있다. 클라이드의 생각이 사건 c에 해당하고 냉장고에 맥주가 있다는 내용이 속성 F에 해당하므로 그들은 (라)처럼 해석한 셈이다. 한편 무늬만 (라)이고 실제 내용은 (다)일 수도 있다.

c가 다른 세계에서 [본질적 속성]이 없다면 실제로 일어나는 c와 같지 않다. 결국 F가 문제되는 사건 c의 본질적 속성이어서 c가 F를 소유하지 않은 사건과 같을 수 없다고 주장될 수 있다. 그렇기 때문에 실제 일어나는 원인-사건 그 자체인 사건을 전건으로 하는 반사실적 조건문만 가지고 자격인과적 관련을 이해하려 해서는 안 된다. (Horgan 1989: 57쪽. c와 F는 이 글의 논의에 맞게 바꾼 것이다.)

곧 반사실적 조건문의 전건을 c가 일어났는데 F가 아닌 경우로 해석해도 F가 본질적 속성이라면 c가 애초에 일어나지 않은 경우나 마찬가지이다. 이를 조심해야 한다.¹⁴⁾

13) (13)의 (a)를 (다)처럼 읽는 것과 (7)의 공통점은 그것들이 c가 F임이 인과적으로 효력 있음의 필요조건은 될지 몰라도 충분조건은 안 된다는 것이다.

14) Braun(1995): 453-4쪽에서 c의 F임이 e의 G임에 인과적으로 효력이 있으려면 c가 e의 원인이고 F가 c에 본질적 속성이면 된다고 말한다. 그리고 이 정의는 반사실적 조건문에 의한 정의와 같은 효력이 있다고 말한다. 그런데 Braun의 이 주장이 맞으려면 그의 본질적 속성에 의한 정의를 가능 세계 접근법으로 읽어야만 할 것이다. 현실 세계에서 c에 본질적 속성 F가 없다면 더 이상 c가 아니기 때문이다.

3. 너무 먼 세계 비판

이제 반사실주의자들은 주장 (13)이 어떤 내용인지 정비되었다. 그것을 다음과 같이 표현해 보자.

- (18) 다음과 같은 조건을 만족하면 그리고 오직 그 경우에만 c 가 e 를 일어나게 하는 데 c 의 속성 F 가 인과적으로 힘이 있다.
- (a) 우리 세계와 가장 가까운 임의의 세계 w 에서, c^* 가 c 의 상대역이고 c^* 가 F 가 아니면 c^* 는 e 의 상대역 e^* 가 안 일어나게 할 것이다.
 - (b) c 가 F 임과 e 가 일어남은 논리적으로 그리고 형이상학적으로 독립적이다.

c^* 가 c 에 대해 상대역 관계에 있다는 것은 c^* 와 c 가 중요한 측면에서 상당한 정도로 질적 유사성이 있다는 것을 의미한다.¹⁵⁾ 그렇다면 정신 인과에 (18)을 적용해서 정신 속성에 인과적인 힘이 있는가 검사해 보자. 내가 냉장고에 맥주가 있다고 믿기 때문에 냉장고로 갔다고 하자. 이때 냉장고에 맥주가 있다는 믿음에 냉장고로 가는 행동을 일어나게 하는 인과적인 힘이 있기 위해서는, 일단 위 (b)는 만족하므로 (a)만 만족하면 된다. c^* 는 맥주가 있다고 믿는 실제 세계 사건과 물리적으로 똑같은 사건이다. 그 사건은 냉장고로 걸어감이라는 실제 세계 행동을 이루는 동작(e)과 물리적으로 똑같은 동작(e^*)의 원인이 된다. 그런데 그 c^* 에 우리가 인과적인 효력이 있는지 관심 있어 하는 “맥주가 있다고 믿음”이라는 정신 속성 F 가 없다고 해 보자. 그때 e^* 가 안 일어날까? e^* 가 안 일어나면, (18)에 따라, 정신 속성에는 냉장고로 가는 행동이 일어나게 하는 인과적인 효력이 있다. 그러나 여전히 e^* 가 일어나면 정신 속성에는 인과적인 효력이 없다고 말해야 한다.

물론 반사실주의자들은 e^* 가 안 일어난다고 말한다. 다음과 같은 세계를 상상해 보자.

- (19) 내가 냉장고에 맥주가 있다고 믿는 실제 세계의 사건과 물리적으로 가장 가까우면서, 동시에 냉장고에 맥주가 있다는 믿음이라는 정신 속성은 없는 가능 세계.

그들은 이구동성으로 이런 세계가 있을 수는 있지만 그 세계는 우리 실제 세계와 너무 동떨어져 있기 때문에, 곧 우리 세계와 가장 가까운 세계¹⁶⁾의 밖에 있기 때문에 (19)의 존재 여부는 “냉장고에 맥주가 있다고 믿음”이라는 정신 속성이 인과적으로 효력이 있느냐는 우리의 문제에 적절하지 않다고 생각한다.¹⁷⁾ 이런 논증을 너무 먼 세계 논변(Too Distant World Argument, Kazez 1995: 73쪽)이라고 해 보자.

너무 먼 세계는 몇 가지 방식으로 상상할 수 있다. 가장 널리 쓰이는 방법은 퍼트남 식의 쌍둥이 지구를 상상하는 거다.

- (20) c^* 가 일어나는 세계는 단 두 가지 점에서만 우리 현실 세계와 다르다. 하나는 냉장고에 맥주가 있다는 믿음

15) Lewis(1968), 선우환(1993) 참조.

16) Horgan(1989)은 이런 세계를 ‘적절하게 비슷한 세계’(pertinently similar world)라고 부른다. 그는 적절하게 비슷한 게 어느 정도인지는 이론 이전의 강한 직관에 의존하면 충분하다고 말한다(60쪽).

17) LePore and Loewer(1989): 188-190쪽, Horgan(1989): 61-4쪽, Braun(1995): 464-7쪽, Heil and Mele(1991): 68쪽, Baker(1993): 93쪽. 그런데 LePore and Loewer의 논변은 약간 이상한 쪽으로 흘러간다. 그들은 맥주가 냉장고에 있다는 믿음이 안 일어나는 먼 세계는 여기서 배제하므로 대신에 그런 믿음이 안 일어나는 가까운 세계에서 는 냉장고 대신에 찬장 같은 곳으로 갈 것이라고 생각한다(190쪽). 그러면 (12)의 (a)는 참이 될 것이다. 그러나 맥주가 냉장고에 있다는 믿음이 없는 사건 c^* 가 과연 맥주가 찬장에 있다는 생각일까? 그들은 그렇다는 증거를 못 들고 있다.

이 없고 또 하나는 물이 H_2O 가 아니라 XYZ로 이루어져 있다. 성분 중에 물이 H_2O 가 아니라 XYZ로 이루어져 있는 맥주를 쌍둥이 맥주라고 불러 보자. 그러면 이 세계에는 냉장고에 맥주가 있다는 믿음이 있다 하더라도 그 믿음의 내용은 현실 세계의 것과 같지 않다. 그것은 냉장고에 쌍둥이 맥주가 있다는 믿음이기 때문이다.

이 세계를 가정하면 (19)가 가능하다. 냉장고에 맥주가 있다는 믿음이 없어도 냉장고에 쌍둥이 맥주가 있다는 믿음이 나를 냉장고로 가게 하기 때문이다. 그러나 이 가능 세계는 우리 세계와 가장 가까운 세계로 간주하기에는 “너무 터무니 없이 다른”(Horgan 1989: 63쪽) 세계이다.

(19)와 같은 세계는 좀비(zombie)의 세계에서도 가능하다. 냉장고에 맥주가 있다는 믿음과 같은 의식 상태가 전혀 없지만 현실 세계와 똑같은 행동은 일어난다. 그러나 이런 세계는 쌍둥이 지구보다 지구에서 더 멀리 떨어져 있는 세계라고 사람들은 생각할 것이다.

그 다음으로는 전도된 스펙트럼(the inverted spectrum) 논변과 비슷한 구조로 (13)과 같은 세계를 상상해 보는 거다. 가령 이 가능 세계에서는 ‘냉장고에 맥주가 있다는 믿음’을 사례화하는 신경 상태와 ‘맥주가 애인보다 좋은 일곱 가지 이유’가 있다는 믿음을 사례화하는 신경 상태가 서로 바뀌어 있다. 그래서 ‘냉장고에 맥주가 있다는 믿음’은 없지만 물리적인 상태는 똑같고, 그리고 여전히 내가 냉장고로 가는 행동이 일어난다. 그러나 역시 이 세계는 우리 현실 세계와 가장 가까운 세계라고 하기에는 너무 이상한 세계라고 사람들이 말할 것이다.

어떤 세계를 상상하든 간에 (19)와 같은 세계는 가능하고 상상할 수는 있지만 너무 멀리 있는 세계라는 게 너무 먼 세계 논변의 요지이다. 나는 세 가지 점에서 이 논변에 문제가 있다고 생각한다. 첫번째는 진 카젯(Jean Kazez)이 지적한 점으로서(1995: 77쪽) (19)와 같은 세계가 없다면 (18)의 (a)에서 c^* 라는 사건이 일어날 수 없으므로 (18)의 (a)의 전건은 거짓이 된다. 따라서 (18)의 (a)는 공허하게 참이 되고 그러면 위에서 (특히 헤일과 밀처럼) 반사실적 조건문을 (다)처럼 해석해서 생기는 문제점이 도지게 된다.

두번째는 (19)와 같은 세계가 너무 멀리 떨어져 있는 세계라는 주장에 대한 불만이다. 냉장고에 맥주가 있다는 믿음이 없으면 냉장고로 가지 않을 것이라는 반사실적 조건문이 거짓이면 정신 속성에는 인과적 효력이 없다. 그게 성립하는 가능 세계는 너무 먼 세계라고 비판한다. 그러나 무슨 근거로 너무 먼 세계라고 단정짓는가? 다음 두 세계를 견주어 보자.

(21) 나는 냉장고에 맥주가 있다는 믿음이 없었고 냉장고로 안 갔다.

(22) 나는 냉장고에 맥주가 있다는 믿음이 없는데도 냉장고로 갔다.

(21)은 현실 세계에서 가장 가까운 세계에서 성립하지만, (22)가 성립하는 세계는 현실 세계에서 가장 가까운 세계에 있지 않다고 반사실주의자들은 말한다. 왜 (21)이 (22)보다 우리 현실 세계에 더 가까운가? 그들은 직관이 그렇게 말해준다고 대답한다. 그러나 김재권에 의하면 그 직관 속에는 이런 대답이 숨어 있다.

한편으로는 [나의] 욕구, 믿음 등과, 다른 한편으로는 [나의] 행동 패턴 사이를 지배하는 어떤 법칙적 규칙성과 성향이 있다는 것을 우리는 알거나 믿는다 우리는 자세하고 정확하게 말해야 할 필요가 있는 부분에 대해서는, 대강의 생각밖에 없긴 하지만, 그런 규칙성과 성향이 믿을 만하고 법칙적이라고 생각하고 반사실적 조건문을 평가할 때 거기에 의존한다. (Kim 1996: 143쪽)

그러므로 나에게 냉장고에 맥주가 있다는 믿음이 없다면 그때 내가 냉장고로 안 간다는 가

정이 간다는 가정보다 법칙같은 규칙에 더 들어 맞는다. 그래서 (21)이 (22)보다 현실 세계에 더 가까운 세계이다. 그러나 법칙에 의존해서 우리 세계와 가장 가까운 세계를 결정한다면 반사실주의자들이 애초에 반사실적 조건문을 해석할 때 규범-도출적 접근법과 가능 세계적 접근법의 두 갈림길에서 가능 세계적 접근법을 골라 지금에 이르렀는데, 이제 다시 규범-도출적 접근법으로 되돌아 가는 꼴이 된다. 더구나 (14) “만약 이 성냥을 그었다면, 불이 붙었을 것이다”라는 반사실적 조건문의 참거짓값을 결정해 주는 “산소가 있을 때 켜지 않은 성냥을 그으면 성냥에 불이 붙는다”라는 법칙에 비해 우리의 “나의 욕구, 믿음 등과, 다른 한편으로는 나의 행동 패턴 사이를 지배하는 어떤 법칙적 규칙성과 성향”은 훨씬 더 의심스러운 것이다. 결국 반사실주의자들이 등장하게 만든 원점, 곧 데이비드슨의 무법칙적 일원론으로 되돌아 가고 만다.

셋째 비판도 (19)와 같은 세계가 너무 멀리 떨어져 있는 세계라는 주장에 대한 불만이다. (19)와 같은 세계가 너무 멀리 떨어져 있다는 근거 중에 하나는 (20)과 같은 세계가 우리 세계와 너무 터무니 없이 다르다는 것이었다. 그렇다면 다음 세계를 보자.¹⁸⁾

(23) ♀가 일어나는 세계는 단 두 가지 점에서만 우리 현실 세계와 다르다. 하나는 동물원에 오랑우탄이 있다는 믿음이 없고 또 하나는 오랑우탄이 단백질이 아니라 나무섬유로 이루어져 있다. 성분 중에 오랑우탄이 단백질이 아니라 나무섬유로 이루어져 있는 오랑우탄을 쌍둥이 오랑우탄이라고 불러 보자. 그러면 이 세계에는 동물원에 오랑우탄이 있다는 믿음이 있다 하더라도 그 믿음의 내용은 현실 세계의 것과 같지 않다. 그것은 동물원에 쌍둥이 오랑우탄이 있다는 믿음이기 때문이다.

실제로 이런 세계는 있을 수 있다. 오랑우탄을 실제로 한 번도 못 보고 그림으로만 본 사람에게서는 오랑우탄이 있는 동물원에 처음으로 가기 전에 품고 있는 오랑우탄 믿음은 얼마든지 나무섬유로 이루어진 쌍둥이 오랑우탄 믿음일 수 있다. (23)과 같은 세계가 가능하면 내가 동물원에 오랑우탄이 있다고 믿는 실제 세계의 사건과 물리적으로 가장 가까우면서, 동시에 ‘동물원에 오랑우탄이 있다는 믿음’이라는 정신 속성은 없는 가능 세계—(19)와 짝이 되는 세계—가 가능하다. 동물원에 오랑우탄이 있다는 믿음이 없어도 동물원에 쌍둥이 오랑우탄이 있다는 믿음이 나를 동물원에 가게 하기 때문이다. 그런데 (23)은 (20)과 똑같은 구조로 이루어져 있는 세계이다. 그렇다면 (20)도 가능한 세계라고 인정해야 한다. 오랑우탄보다 물이 우리에게 훨씬 낯익기 때문에 쌍둥이 오랑우탄이 있는 세계보다 쌍둥이 물이 있는 세계가 그만큼 터무니 없이 이상한 세계라고 여겨지겠지만, 그런 낯익음을 씻어내고 바라보면 (20)과 (23)은 똑같은 구조이므로 (20)은 얼마든지 우리와 가깝게 있는 세계라고 인정해야 한다.

4. 익숙함의 문제와 선택 안 됨 조건

나는 (19)와 같은 세계가 가능할지 몰라도 너무 멀리 있는 세계라는 비판이 가능한 것은 순전히 익숙함의 문제 때문이라고 생각한다. 나는 다음과 같은 네 가지 사례를 견주어 봄으로써 익숙함 때문에 생기는 문제를 지적하려고 한다. 그래서 현실 세계에서 먼 세계에서 일어나는 사건은 인과적 효력의 검사에서 고려 대상이 안 된다는 너무 먼 세계 비판이 잘못되었다고 주장하겠다.

18) 이 보기가 순전히 나 스스로 만든 것이라면 좋으련만, 아쉽게도 Kazez(1995): 73쪽에 착상 수준으로 나온 것을 구체적인 보기로 만들어 본 것이다.

내가 자동차 열쇠를 돌림은 자동차의 시동이 걸림의 원인이다. 이때 내 자동차 열쇠가 은빛이라는 속성은 자동차의 시동이 걸림에 인과적으로 효력이 있는지 없는지 알고 싶다.

(24) 내 자동차 열쇠가 은빛이 아니면 자동차의 시동이 안 걸릴 것이다.

이 반사실적 조건문이 참인지 아닌지 알기 위해서 정의 (18)이 가르쳐준 바에 따라 자동차 열쇠가 은빛이 아닌 세계를 상상해 봐야 하는데, 다음과 같은 두 세계가 가능하다.

(24가) 자동차 열쇠가 은빛이 아니면서 동시에 자동차 열쇠의 모양도 달라졌다.

(24나) 자동차 열쇠가 은빛이 아니면서 동시에 자동차 열쇠의 모양은 그대로다.

우리는 이 두 세계 중에서 (24나)가 현실 세계에 더 가까운 세계인지 안다. 그 세계에서는 여전히 자동차의 시동이 걸리므로 (24)는 거짓이고, 따라서 자동차 열쇠의 빛깔은 시동이 걸리게 하는 데 인과적으로 효력이 없다. (24가)의 세계라면 (24)가 참이 되어 자동차 열쇠의 빛깔에 인과적인 효력이 있을 텐데, 그 세계에서 (24)의 참거짓값을 검사하는 사람은 없다. 왜 그럴까? 일단은 루이스-스틀레이커가 말한 대로 (24가)보다 (24나)가 현실 세계에서 최소한의 변화만 있는 세계이기 때문에, (24가)가 (24나)보다 현실 세계에 더 가깝다고 생각할 사람은 없으므로 거기서 (24)의 참거짓값을 검사하지 않는 것이다. 그러나 또 다른 이유는 (24가)는 자동차의 시동이 걸리는 데 인과적으로 효력이 있는지 궁금한 후보, 곧 열쇠의 빛깔과 인과적으로 효력이 있다고 생각되는 유력한 경쟁 후보, 곧 열쇠의 모양이 동시에 변화하고 말았다. 그러므로 그 세계에서 자동차 시동이 안 걸린다 하더라도 그 원인이 열쇠의 빛깔에 있는지 열쇠의 모양에 있는지 알 수 없다는 난점이 생긴다. 그래서 (24가)는 인과적 효력을 검사하기에 좋은 세계가 아니다. (24)의 경우에는 우리에게 익숙한 세계와 인과적 효력을 검사할 세계가 일치하였다.

그 다음 보기를 보자(Kazez 1995: 84쪽). 의학자들이 기형아 출산의 원인을 연구하고 있다. 스트레스녀라는 여인이 기형아를 출산했다고 해 보자. 의학자들은 그 여자가 임신 기간 중에 꽤 큰 스트레스를 받았고, 술을 마셨으며, 담배를 피웠고, 영양 결핍이 있었다는 것을 알아냈다. 의학자들은 이 원인들 중에 음주만이 기형아 출산의 원인이라고 주장한다. 그런데 위 (18)에 의하면 스트레스녀의 과도한 스트레스도 기형아 출산에 인과적으로 책임이 있다고 말할 수 있다.

(25) 만약 스트레스녀가 스트레스를 안 받았다면 기형아를 출산하지 않았을 것이다.

반사실주의자들은 이 반사실적 조건문이 참임을 알기 위해 스트레스녀가 스트레스를 안 받는 세계를 상상해야 하는데 다음과 같이 두 가지 세계가 가능하다.

(25가) 스트레스녀가 스트레스를 안 받았고 술도 덜 마셨다.

(25나) 스트레스녀가 스트레스를 안 받았고 술은 여전히 마셨다.

(25가)가 우리 세계와 더 가까운 세계이다. 우리는 정말로 스트레스를 덜 받을수록 술도 덜 마신다. 그 세계에서는 (25가)가 참이다. 그렇다면 의학자들의 주장과는 달리 스트레스가 기형아 출산에 인과적 효력이 있다는 결과가 나온다. 그러나 (25가)는 스트레스의 인과적 효력을 검사해 보기에 좋은 세계가 아니다. 기형아 출산의 잠정적 후보인 스트레스 외에 유

력한 경쟁 후보인 음주까지 함께 없는 세계이기 때문에 스트레스의 인과적 효력을 제대로 검사해 볼 수 없다. 의학자들은 (25나)의 세계에서 검사해 보아야 제대로 된 대답을 내놓을 수 있다. (25)에서는 인과적 효력을 검사해야 할 세계가 우리에게 익숙한 세계가 아니다. (24)와 (25)를 견주어 보자. (24가)와 (25가) 모두 인과적 효력을 검사해 보기에는 좋은 세계가 아니었다. 인과적 효력의 검사를 위한 잠정적 후보와 유력한 경쟁 후보가 함께 없는 세계이기 때문이다. 그런데 우리는 (24가)가 검사를 위한 좋은 세계가 아니라는 것을 금방 알면서도 (25가)도 좋은 세계가 아니라는 것을 알지 못하고 잘못된 검사를 하곤 한다. 왜 그럴까? (24가)는 (24나)에 비해 덜 익숙한 세계이기 때문에 그 세계에서 검사를 하면 안 된다는 것을 잘 알지만, (25가)는 (25나)보다 더 익숙한 세계이기 때문에 그 세계에서 검사하는 데 전혀 거리낌이 없다는 이유가 있다. 익숙한 세계란 곧 우리 세계와 가까운 세계이다. 따라서 인과적 효력을 검사할 세계가 꼭 우리 현실 세계와 가장 가까운 세계여야 한다는 주장은 잘못이다. 그리고 우리 세계와 먼 세계이기 때문에 인과적 효력을 검사할 때 고려 대상이 되지 않는다는 비판은 옳지 못하다. 우리 현실 세계와 가깝다는 기준보다는 인과적 효력이 있다고 생각되는 속성이 제대로 검토되고 있느냐는 기준이 더 중요한 것이다. 정신 인과에 대해서도 마찬가지로 이야기를 할 수 있다.

(26) 만약 정신 속성이 없다면, 행동은 일어나지 않았을 것이다.

이 반사실적 조건문이 참임을 알기 위해 정신 속성이 없는 가능 세계를 다음과 같이 두 가지로 상상한다.

(26가) 정신 속성이 없고 거기에 해당하는 물리 속성도 없는 세계

(26나) 정신 속성은 없으나 물리 속성은 여전히 있는 세계

(21)과 (22)가 각각 (26가)와 (26나)에 해당할 것이다. (26가)가 (26나)보다 우리 세계에 더 비슷하다. 거기서는 (26)이 참이 되어 정신 속성에는 인과적 효력이 있게 된다. 그러나 (26가)는 (24가)와 (25가)처럼 정신 속성의 인과적 효력을 검사하기에 좋은 세계가 아니다. 인과적 효력의 잠정적 후보인 정신 속성과 결정적인 경쟁 후보인 물리 속성이 함께 없는 세계이기 때문이다. 따라서 진정한 검사를 위해서는 덜 가까운 세계라도 (26나)로 가야 하고 거기서 (26)은 거짓이 되어 정신 속성에는 인과적 효력이 없다는 결론이 나오게 된다. 그렇다면 마찬가지로 쌍둥이 물의 세계가 물의 세계에 비해 덜 익숙하기 때문에 맥주가 있다는 믿음이 인과적으로 효력이 있는지 따질 때 고려 대상이 안 된다는 주장은 잘못이다.

나는 (24), (25), (26)의 보기는 우리에게 두 가지 점을 말해 주고 있다고 생각한다. 하나는 (25가)와 (26가)에 비해 (24가)는 인과적 효력의 검사를 위해 좋은 세계가 아니라는 것을 얼른 알게 된 데에는 (24가)가 덜 익숙한 세계라는 까닭이 있다는 점이다. 그러나 인과적 효력이 있는지 여부를 검사할 때 익숙함의 정도는 본질적인 요소가 아니다. 따라서 (25나)와 (26나)가 덜 익숙한, 먼 세계이기 때문에 인과적 효력을 검사하기에 좋은 세계가 아니라고 말해서는 안 된다. 두 번째는 어떤 속성이 인과적 효력이 있는지 알기 위해서는 그 속성과 경쟁하는 속성이 있는지를 살펴보아야 한다는 점이다. 같은 결과에 대해 경쟁하는 원인이 두 개 있고 하나가 다른 하나보다 우세한 원인이라면 그 우세한 원인이 다른 원인을 인과적으로 배제한다는 것이 인과적 배제 논변이다. 그렇다면 우리 현실 세계에서 가장 가까운 세계가 반사실적 조건문의 참거짓값 검사에서 가능 세계의 범위라고 한다면, 가장 가까운

세계가 어디까지인지 결정할 때 경쟁하는 속성의 존재 여부가 고려 요소 중에 하나가 될 것이다.

나는 이 두번째 점이 정의 (18)에 반영되어야 한다고 생각한다. 그 작업을 위해 다음 보기를 생각해 보자. 내가 목이 말라 손을 뻗어 컵을 잡으려고 했다고 하자. 나는 어떤 신경학적인 사건 때문에 내 손을 뻗는다고 할 수 있다. 그 사건은 목마름을 가시게 하려는 나의 욕구이기도 하다. 따라서 나의 움켜잡기가 그 신경학적인 사건이 원인이 되어 일어난다면 나의 움켜잡기의 원인은 다른 게 아니라 바로 나의 욕구다. 소자(Ernest Sosa)는 다음과 같이 말한다.

정신의 무법칙성을 가정한다면, 내 손을 뻗음은 어떤 의미에서는 목마름을 가시게 하려는 나의 갑작스런 욕구가 원인이 되어 일어나긴 했지만, 내 욕구가 욕구의 자격으로서 원인이 된 것이 아니라 어떤 종류의 신경학적인 것의 자격으로서 원인이 된 것이다 (정신적인 것이 정말로 무법칙적이라면) 내 욕구의 욕구임은 내가 손을 뻗음에 인과적인 관련이 전혀 없다. 만약 내 욕구인 사건이 내 욕구가 아니지만 어떤 종류의 신경학적 사건으로는 그대로 남아 있다면, 그 사건은 여전히 내 손을 뻗음의 원인이 될 것이다.(Sosa 1984: 277-8쪽)

내 욕구가 없더라도 손을 뻗는다는 행동은 여전히 일어난다. 어떤 종류의 신경학적 속성이 그 행동을 일어나게 했기 때문이다. 그리고 앞에서 말한 인과적 배제 주장을 받아들인다면 내 욕구라는 정신 속성은 신경학적 속성에 의해 선택되었고, 따라서 인과적 힘을 쓸 수 있는 지위가 배제되었다. 이제 우리는 (18)을 다음과 같이 바꾸어야 할 것 같다.

(27) 다음과 같은 조건을 만족하면 그리고 오직 그 경우에만 c 가 e 를 일어나게 하는데 c 의 속성 F 가 인과적으로 힘이 있다.

- (a) 우리 세계와 가장 가까운 임의의 세계 w 에서, c^* 가 c 의 상대역이고 c^* 가 F 가 아니면 c^* 는 e 의 상대역 e^* 가 안 일어나게 할 것이다.
- (b) c^* 가 F 가 아니면서 F 인데도 e^* 가 일어나는 그런 속성 F^* 가 없어야 한다.
- (c) c 가 F 임과 e 가 일어남은 논리적으로 그리고 형이상학적으로 독립적이다.

(b)를 선택 안 됨 조건이라고 불러 보자. 위 보기에서 목마름을 가시게 하려는 나의 욕구가 F 이고 거기에 해당하는 신경학적 속성이 F^* 이다. F 가 없으면 내가 손을 뻗는 행동은 안 일어난다. 그래서 (a)는 만족된다. F 가 없으면서도 F^* 가 있는 세계는 익숙하지는 않은 세계이지만 그 점 때문에 인과적 효력을 검사할 때 고려 대상에서 빼서는 안 된다고 이미 말했다. c^* 에서 F 가 없으면서 F^* 가 있다면 e^* 는 여전히 일어난다. 그렇다면 (b)는 만족되지 못한다. 따라서 F , 곧 목마름을 가시게 하려는 나의 욕구에는 손을 뻗는 행동을 일으키는 인과적 효력이 없다.¹⁹⁾

그러나 반사실주의자들은 다음과 같이 비판할 것이다(LePore and Loewer 1987: 638-40 쪽을 보라). ((27)은 반사실적 조건문으로 정의되긴 했지만 나는 반사실주의자가 반사실적 조건문으로 부수현상론을 피하려는 이들로 정의했기 때문에 (27)은 반사실주의자들의 주장이 아니다.) 그들은 선택 안 됨 조건까지 보탠 (27)은 인과적 효력의 조건으로서는 너무 강한 조건이라고 말한다. 이 정의에 따르면 인과적으로 효력있다고 생각되는 물리 속성마저도 효력이 없게 되는 결과가 생기기 때문이다. M 이 목마름을 가시게 하려는 나의 욕구, P 가 거기에 해당하는 신경생리학적 속성이라고 하자.

19) Horgan(1989)도 주 34에서 선택 안 됨 조건을 인지하고 있다. 그러나 그에게서 F^* 의 대상이 될 수 있는 것들은 쌍둥이 맥주처럼 너무 먼 세계에 있는 것이므로 해석하게도 실제 그 조건이 쓰여지지 않는다.

(28) c^* 가 M이 아니면서 P인데도 e^* 가 일어나는 그런 속성 P가 있다.

이게 가능하기 때문에 정신 속성에는 인과적인 효력이 없다는 게 정의 (27)의 요지였다. 르포와 로워같은 반사실주의자들은 그렇다면 다음도 가능하기 때문에 물리 속성에도 인과적 효력이 없게 된다고 말한다.

(29) c^* 가 P가 아니면서 M인데도 e^* 가 일어나는 그런 속성 M이 있다.

그들은 이렇게 말한다.

(29)는 무법칙적 일원론 그리고 (28)과 양립가능함이 보여질 수 있다. 더구나 (29)는 실제로 참이라고 할 만하다. 만약 c 가 목마름을 가시게 하려는 욕구이지만 P는 아니라면 거기에는 어떤 다른 속성 P^* 가 있을 것이다 곧 소자가 목마름을 가시게 하려는 욕구가 있지만 그 욕구가 P가 아닌 가장 가까운 세계에서 그는 여전히 손을 뻗는다. (LePore and Loewer 1987: 639쪽, 원문 강조, (28)과 (29)은 이 글의 논의에 맞게 바꾼 것이다.)

그들이 이렇게 주장할 수 있는 까닭은 정신 속성 M이 복수실현가능하기 때문이다. M을 실현하는 물리 속성이 P 말고 P^* 도 가능하기 때문에 그 때도 역시 손을 뻗는 행동의 원인이 되는 것이다.²⁰⁾

나는 이런 비판이 얼른 보기에 그럴 듯하면서도 자세히 보면 잘못된 비판이라고 생각한다.²¹⁾ 그 잘못된 이유는 (28)과 (29)의 차이를 알지 못한 데 있다. (28)에서는 M이 P에 의해 선택되면 그걸로 끝이다. 어떤 다른 정신 속성이 나타나서 물리 속성 P에 의해 선택됨을 막지 못한다. 그러나 (29)에서는 사정이 다르다. (29)가 맞다고 하자. 그럼 르포와 로워도 친절하게 말한 것처럼 P를 선택하는 것은 정신 속성 M이 아니라 또 다른 물리 속성 P^* 이다. 따라서 P가 M을 선택하지 못한다 하더라도 P^* 가 다시 M을 선택하므로 정신 속성에 인과적 효력이 없다는 주장에는 전혀 흠이 없다. 더구나 나는 어떤 가능 세계에서 P와 P^* 는 터무니 없이 다르지는 않을 것이라는 믿음이 있다.²²⁾

5. 맺음말

지금까지 논의를 요약해 보자. 반사실주의자들은 데이비드슨의 엄격한 법칙을 반사실적 조건문으로 이해하여 무법칙적 일원론이 부수현상론을 함축하지 않는다는 것을 보여주려고 한다. 반사실적 조건문을 통해 정신 속성에 인과적 효력을 찾아주려는 반사실주의자들의 시도는 천신만고 끝에 최종적으로 (27)에 이르렀다. 그러나 그 정의에서 정신 속성은 인과적 효력이 없다는 결론에 이르게 된다. 그들의 시도는 실패한 것이다.²³⁾

20) 하종호 교수는 LePore와 Loewer의 이 주장에 대해 다음과 같이 비판하고 있다. 만일 M이 발생할 때 항상 P만이 발생하고 다른 신경생리학적 속성은 발생하지 않는다고 할 때(이러한 가능성을 배제할 수 있는 근거는 전혀 없다) 그들은 어떻게 답변할 것인가?(하종호 1984: 387쪽). 이 비판은 복수실현 가능성 입론을 무시하고 있다.

21) 나는 Leiter and Miller(1994):226쪽과 Antony(1991): 306-7쪽, Block(1990): 주 6에서도 나와 비슷한 비판을 하고 있음을 알게 되었다.

22) LePore and Loewer도 허리케인이 홍수의 원인인 보기와 관련해서 다음과 같이 인정하고 있다. “정말로 [허리케인]에 P와 충분히 비슷한 어떤 속성 P^* 가 있을 것이고, P^* 는 (적절한 조건에서) 홍수를 일으킨다.” (1989: 640쪽, 나의 강조).

23) Block(1990)은 반사실주의적 접근을 살릴 수 있는 길로 기능적 역할 의미론(functional role semantics)을 소개하

K C I

고 있다. 그것에 관한 논의는 다음 과제로 남겨두겠다.

참고문헌

- 김기현(1998), 『현대 인식론』, 서울: 민음사.
- 선우환(1993), “양화양상논리 의미론과 루이스의 상대역이론: 양상의미론의 두 대안과 또 다른 대안”, 서울대학교 철학과 석사학위 논문.
- 최훈 (2000), 『정신 속성의 인과적 힘』, 서울대학교 박사학위 논문.
- 최훈 (2004), “데이비드슨의 무법칙적 일원론과 백도형 교수의 심신 유명론”, 미발표.
- 하중호(1994), “무법칙적 일원론과 심성 인과”, 김재권 외 지음, <수반의 형이상학>: 379-397쪽.
- Antony, Louise M. (1991), “The Causal Relevance of the Mental: More on the Mattering of Minds”, *Mind and Language* 6: 295-327 쪽.
- Block, Ned (1990), “Can the Mind Change the World?”, in George Boolos(ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge: Cambridge University Press: 137-170쪽.
- Braun, David (1995), “Causally Relevant Properties”, *Philosophical Perspectives* 9: 447-475쪽.
- Burge, Tyler (1993), Mind-Body Causation and Explanatory Practice, in J. Heil and A. Mele(eds.), *Mental Causation*, Oxford: Oxford University Press: 97-120쪽.
- Davidson, Donald (1980a), “Mental Events”, in *Essays on Actions and Events*, Oxford: Oxford University Press: 207-227쪽.
- Davidson, Donald (1980b), “Psychology as Philosophy”, in *Essays on Actions and Events*, Oxford: Oxford University Press: 229-239쪽.
- Davidson, Donald (1980c), “The Material Mind”, in *Essays on Actions and Events*, Oxford: Oxford.
- Doring, Frank (1998), “Counterfactuals”, Edward Craig(ed.), *Routledge Encyclopedia of Philosophy Vol. 3*, London: Routledge.
- Fodor, Jerry(1990a), “Making Mind Matter More”, in *A Theory of Content and Other Essays*, Cambridge: MIT Press: 137-159쪽.
- Follesdal, Dagfinn(1985), “Causation and Explanation: a Problem in Davidson’s View on Action and Mind”, in Ernest LePore and Brian McLaughlin(eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell: 311-323쪽.
- Hess, P. (1981), “Action, Reasons, and Human Causes”, *Analysis* 41: 77-81쪽
- Heil, John and Alfred Mele(1991), Mental Causes, *American Philosophical Quarterly* 28: 61-71쪽.
- Honderich, Ted (1982), “The Argument for Anomalous Monism”, *Analysis* 42: 59-64쪽.
- Horgan, Terence (1989), “Mental Quaustation”, *Philosophical Perspectives* 3: 47-76쪽.
- Johnston, Mark (1985), “Why Having a Mind Matters”, in Ernest LePore and Brian McLaughlin, (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell: 408-426쪽.
- Kazez, Jean R. (1995), “Can Counterfactuals Save Mental Causation?”, *Australasian Journal of Philosophy* 73: 71-90쪽.
- Kim, Jaegwon (1990), “Explanatory Exclusion and the Problem of Mental Causation”, in E. Villanueva(ed.), *Information, Semantics, and Epistemology*, Oxford: Oxford University Press: 36-54쪽.
- Kim, Jaegwon (1993), “The Myth of Nonreductive Materialism”, in *Supervenience and Mind: Selected*

Philosophical Essays, Cambridge: Cambridge University Press: 265-284쪽.

Kim, Jaegwon (1995), "Mental Causation: What? Me Worry?", in E. Villanueva(ed.), *Philosophical Issues* 6, Content, Atascadero: Ridgeview Pub. Co.: 123-51쪽.

Kim, Jaegwon (1996), *Philosophy of Mind*, Boulder: Westview Press.

Leiter, Brian and Alexander Miller (1994), "Mind Doesn't Matter Yet", *Australasian Journal of Philosophy* 72: 220-228쪽.

Lepore, Ernest and Barry Loewer (1987), "Mind Matters", *Journal of Philosophy* 84: 630-642쪽.

Lepore, Ernest and Barry Loewer (1989), "More on Making Mind Matter", *Philosophical Topics* 17: 175-192.

Lewis, David(1968), "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy* 65: 113-26쪽.

Ludwig, Kirk(1994), Causal Relevance and Thought Content, *The Philosophical Quarterly* 44: 334-353쪽.

McLaughlin, Brian (1989), "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical," in *Philosophical Perspectives* 3: 109-135쪽.

Skillen, A. (1984), "Mind and Matter: A Problem that Refuses Dissolution", *Mind* 93: 514-526쪽.

Sosa, E. (1984), "Mind-body Interaction and Supervenient Causation", *Midwest Studies in Philosophy* 9: 271-282쪽.

Stephan, Achim(1997), "Armchair Arguments Against Emergentism", *Erkenntnis* 46: 305-314쪽.

Stoutland, Fred (1985), "Davidson on Intentional Behavior", in Ernest LePore and Brian McLaughlin, (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell: 44-59쪽.

Tomberlin, James (1995), "Mental Causation: A Query for Kim", in E. Villanueva(ed.), *Philosophical Issues* 6, Content, Atascadero: Ridgeview Pub. Co.: 152-9쪽.

Counterfactuals Can't Help

Hoon, Choi

The anomalous monism proposed by Donald Davidson is criticized that it will be implies mental epiphenomenalism. One way to get out of that difficulty is to explain the causal relation by replacing the strict law of Davidson by counterfactual conditionals. This paper tracks the definition to endow mental properties with causal power, and conclude that mental properties cannot have mental property in any definition.

Key Words counterfactual argument, mental causation, epiphenomenalism, familiarity

K C I