논리연구 12-2(2009) pp. 89~114

의학의 뉴컴 문제와 인과적 결정 이론* †

여 영 서

【요약문】 우리는 여러 종류의 인과적 믿음을 지니고 있으며, 인과적 믿음은 합리적인 결정을 내리는 과정에서 중요한 역할을 한다. 이 직관을 발전시킨 인과적 결정 이론은 행위자의 결정이 합리적이라는 설명을 제시하기 위해서 그 행위자의 결정이 의존하는 인과적 믿음을 명확하게 밝히는 것이 필요하다고 주장한다. 그럴 필요가 없다는 입장의 증거적 결정 이론은 뉴컴 문제를 통해 반박된다. 그러나 뉴컴 문제의 다양한 형태 중에서 의학의 뉴컴 문제가 증거적 결정 이론을 반박하는 데에 가장 성공적이라는 일반적 판단은 잘못이라는 점이 본 논문에서 논증된다. 본 논문은 의학의 뉴컴 문제는 인과 관계를 명료하게 진술함으로써 오히려 증거적 결정 이론을 반박하기 어려워진다는 점을 지적한다. 이 과정에서 본 논문은 증거적 결정 이론과 인과적 결정 이론 사이의 차이점을 드러내고, 합리적 결정 과정에서 인과적 믿음이 정확하게 어떤 역할을 하는지를 밝힌다.

【주요어】 인과, 의학의 뉴컴 문제, 증거적 결정 이론, 인과적 결정 이론, 쌍둥이 죄수의 딜레마

---

* 접수완료: 2009. 7. 2 심사 및 수정완료: 2009. 8. 11

† 본 논문은 2005년도 한국학술진흥재단의 지원(인문사회분야지원심화연구 KRF-2005-079-AS0034)에 의해 연구되었음.

1. 들어가는 말

우리는 여러 종류의 인과적 믿음을 지니고 있으며, 대개는 이러한 인과적 믿음에 의존하여 어떤 결정을 한다. 예를 들어, 우리가 담배를 끊어야겠다고 결정한다면, 그 결정은 적어도 부분적으로는 "흡연은 폐암을 유발한다"와 유사한 종류의 인과적 믿음에 의존한 것이다. 흡연이 원인이 되어 나타날 결과들을 고려해 볼 때, 그 결과들이 폐암과 같이 내가 원하는 바가 아니기 때문에 또는 흡연으로 인해 얻을 수 있는 즐거움보다 흡연으로 인해 나타날 피해가 더 크기 때문에 그 원인이 되는 행위 즉 흡연을 포기하는 결정을 한다.

이와 같이 인과적 믿음이 합리적인 결정을 내리는 과정에서 중요한 역할을 한다는 직관을 발전시킨 것이 인과적 결정 이론(Causal Decision Theory, 이후 CDT로 약칭함)이다. CDT는 행위자의 결정이 합리적이라는 설명을 제시하기 위해서 그 행위자의 결정이 의존하는 인과적 믿음을 명확하게 밝히는 것이 필요하다고 주장한다. 이와는 달리 행위자의 결정이 합리적이라는 설명을 제시하기 위해서 인과적 믿음을 특별히 밝힐 필요가 없다는 입장이 증거적 결정 이론(Evidential Decision Theory, 이후 EDT로 약칭함)이다. EDT는 인과적 믿음이 그 행위자의 비인과적 믿음에 의해 충분히 드러난다는 입장이다.

CDT는 EDT를 비판하면서 제시되는데, 현재 다수의 철학자들은 CDT를 지지하고 있는 상황이다. 하지만 EDT와 CDT 사이의 논쟁이 마무리된 것은 아니며, EDT를 옹호하려는 시도 또한 계속되고 있다. 제프리(R. Jeffrey), 프라이스(H. Price), 히치콕(C. Hitchcock) 등을 중심으로 EDT를 옹호하는 여러 가지 시도가 있다.[1] 또 이러

한 시도를 반박하고 CDT를 옹호하는 논문들도 계속 출판되고 있는 실정이다.[2] 최근에는 CDT도 뉴컴 문제와 유사한 난점에 직면해 있다는 주장이 제기되면서 기존의 EDT 및 CDT 이외의 새로운 합리적 결정 이론을 요청하는 논의도 나타나고 있다.[3]

이처럼 합리적 결정 이론에 대한 관심이 다시 한 번 증폭되고 있는 상황에서 이 글은 EDT와 CDT 사이의 논쟁의 핵심에 놓여 있는 뉴컴 문제(Newcomb's Problem)를 다시 검토해 보는 것을 목적으로 한다. 뉴컴 문제는 여러 가지 형태로 제시될 수 있는데, 그 다양한 형태는 노직(R. Nozick)이 소개한 두 상자 문제, 코코의 문제나 흡연의 문제와 같은 의학의 뉴컴 문제(Medical Newcomb's Problem), 쌍둥이 죄수의 딜레마(Twin Prisoners' Dilemma) 등 크게 세 가지 유형으로 구분할 수 있다.[4] 그런데 이러한 뉴컴 문제의 다양한 형태 중에서도 특히 의학의 뉴컴 문제가 EDT를 명확하게 반박한다고 알려져 있다. 하지만 이 글에서는 의학의 뉴컴 문제가 노직이 제시한 원래의 뉴컴 문제에 비해 인과 관계를 명료하게 진술함으로써 오히려 EDT를 반박하는 힘이 약화된다고 주장할 것이다. 이 과정에서 이 글은 EDT와 CDT 사이의 차이점을 드러내며 합리적 결정 과정에서 인과적 믿음이 어떤 역할을 하는지를 재고할 수 있는 기회를 제공할 것이다.

---

1) EDT도 두 상자를 선택한다고 하는 tickle 논변, 뉴컴 문제는 결단의 문제가 아니라는 제프리의 논변, 행위자의 확률의 의존하는 프라이스의 논변 등의 시도가 있다.

2) Ahmed, A. (2005)

3) Egan, A. (2007)

4) Price (1986) p.195.

2. 노직의 뉴컴 문제와 증거적 결정 이론

합리적 결정 이론은 행위의 기대 효용성(Expected Utility)을 극대화하는 선택지를 계산해낸다. 기대 효용성(EU)은 행위(A)로부터 기대하는 모든 결과(O_i)의 효용성(Ut)을 그 결과가 발생할 확률(Pr)에 곱한 것의 평균값이다. 이것은 다음처럼 기호화할 수 있다.

$$EU(A) = \sum Pr(O_i \text{ if } A)Ut(A\&O_i)$$

이때 $Pr(O_i \text{ if } A)$을 어떻게 계산할 것인가가 EDT와 CDT를 구분짓는 점이다.[5] EDT에 따르면, $Pr(O_i \text{ if } A)$는 $Pr(O_i \mid A)$라고 해석되고, $Pr(A) \neq 0$의 경우 $Pr(O_i \mid A) = Pr(O_i\&A)/Pr(A)$라고 정의된다. 반면 CDT에 따르면, $Pr(O_i \text{ if } A)$는 $Pr(O_i \Box\!\!\rightarrow A)$ 즉 인과적 확률값으로 해석된다. 이 경우 $Pr(O_i \Box\!\!\rightarrow A)$는 A 및 O의 확률값으로부터 확률론에 의해 계산해서 얻을 수 있는 값이 아니다. 그 자체로 확률값을 부여받아야 하는 것이다.[6]

이처럼 $Pr(O_i \text{ if } A)$의 값을 어떻게 계산할 것인가에 따라 구분되는 EDT와 CDT는 노직이 소개한 뉴컴 문제[7]에 대해 서로 다른

---

5) $Pr(O_i \text{ if } A)$를 $Pr(A \Box\!\!\rightarrow O_i)$ 즉 가정법적 조건문(subjunctive conditional)으로 해석하려는 시도는 스톨네이커(R. Stalnaker)에 의해 제시되었다. 기버드(A. Gibbard)와 하퍼(W. Harper)는 스톨네이커의 제안을 정교화하여 기대 효용성을 EDT처럼 $Pr(O_i \text{ if } A)$를 $Pr(O_i \mid A)$라고 해석하는 것보다는 CDT처럼 $Pr(O_i \text{ if } A)$를 $Pr(A \Box\!\!\rightarrow O_i)$로 해석하여 계산해야 한다고 주장한다.

6) $Pr(O_i \Box\!\!\rightarrow A)$를 어떻게 계산할 것인가의 문제에 대해서는 다음 참조.

7) 1969년 노직이 처음 소개한 뉴컴 문제는 결정 이론의 두 원리, 즉 지배의 원리(the principle of dominance)와 기대 효용 극대화의 원리(the principle of expected utility maximization)가 서로 충돌하는 사례로 제시된다. 두 결단 원리는 간단히 다음처럼 정의된다.

지배의 원리: 세계를 상태에 따라 구분할 때 그 각각의 상태에 상대적으

답변을 제시하면서 차이점을 드러낸다. 노직이 소개한 뉴컴 문제는 다음 상황에서 발생한다.

> 당신의 선택을 예측하는 힘에 대해 당신이 매우 신뢰하는 존재자를 가정해보자. (이 존재자에 대해 다른 행성에서 왔고, 진보한 과학기술을 지녔으며 호의적이라는 등등의 공상소설을 쓸 수도 있을 것이다.) 당신은 이 존재자가 과거에 당신의 선택을 정확하게 예측한 경우가 많다는 사실(그리고 당신이 아는 한 당신의 선택에 대해 잘못된 예측을 한 적이 없다는 사실)을 알고 있다. 나아가 당신은 이 존재자가, 아래에서 기술하게 될 특별한 상황 아래에서, 대부분 당신과 비슷한 다른 사람의 선택을 정확하게 예측한 경우가 많다는 사실을 알고 있다. 더 길게 이야기를 할 수도 있겠지만 이 모든 것은 당신으로 하여금 거의 분명하게 논의하게 될 상황에서의 당신의 선택에 관한 이 존재자의 예측이 옳다고 믿게 할 것이다.
>
> (B1)과 (B2)의 두 상자가 있다. (B1)에는 $1,000가 들어있다. (B2)에는 $1,000,000가 들어있거나 비어 있다. (B2)의 내용물이 무엇인가는 곧 기술하게 될 조건에 달려있다.
>
> (B1) {$1,000}　　(B2) {$1,000,000 또는 $0}
>
> 당신은 두 가지 행동 중에 선택을 할 수 있다.
>
> (1) 두 상자의 내용물을 모두 가진다
>
> (2) 두 번째 상자의 내용물만을 가진다.
>
> 나아가 당신은 다음 사실을 알고 있고, 존재자 역시 당신이 다음 사실을 알고 있는 것을 알고 있는 등등이다.
>
> (1) 만약 당신이 두 상자의 내용물을 모두 가질 것으로 존재자가 예측하면, 그 존재자는 $1,000,000을 두 번째 상자에 넣어두지 않는다.
>
> (2) 만약 당신이 두 번째 상자의 내용물만을 가질 것으로 존재자가 예측하면, 그 존재자는 $1,000,000을 두 번째 상자에 넣어둔다.
>
> 상황은 다음과 같다. 먼저 존재자가 예측을 한다. 그리고 나서 그 존재자는 그의 예측에 따라 $1,000,000을 두 번째 상자에 넣거나

로 행위 A가 행위 지배한다면, B 대신 A가 수행되어야 한다.

기대 효용성 극대화의 원리: 취할 수 있는 행위 중에 기대 효용을 극대화하는 행위를 수행해야 한다. Nozick, R. 1969. p.118.

이 두 원리에 대한 상세한 설명은 이종권(2006) 참조.

넣지 않거나 한다. 그 후 당신은 선택을 한다. 당신은 어떻게 할 텐가?[8]

위의 문제에서 선택 상황은 다음과 같은 표로 제시될 수 있다.

<표 1> 노직의 뉴컴 문제

선택지	존재자는 당신이 한 상자(두 번째 상자)의 내용물만을 가질 것으로 예측한다(Pone)	존재자는 당신이 두 상자의 내용물을 모두 가질 것으로 예측한다(Ptwo)
한 상자(두 번째 상자)의 내용물만을 가진다(One)	$1,000,000	$0
두 상자의 내용물을 모두 가진다(Two)	$1,001,000	$1,000

여기에서 당신은 One의 경우 Pone일 가능성이 매우 높다고 보지만 Two의 경우 Pone일 가능성은 매우 낮다고 본다. 즉 Two의 경우 Ptwo일 가능성이 매우 높다고 본다. 이것은 명제 (One&Pone) ∨ (Two&¬Pone)에 대한 당신의 주관적 믿음의 정도가 확률값 1에 가깝다는 것이다.

이 때 EDT는 한 상자의 내용물만을 가지는 것이 합리적인 선택이고, 두 상자의 내용물을 모두 가지는 것이 비합리적인 선택이라고 말한다. 이와는 달리 CDT는 두 상자의 내용물을 모두 가지는 것이 합리적인 선택이고 두 번째 상자의 내용물만을 가지는 것이 비합리적인 선택이라고 말한다. 이렇게 서로 다른 답변을 제시한 결과 많은 철학자들은 EDT를 포기하고 CDT를 지지하게 된다. EDT가 틀린 답변을 제시한 반면 CDT가 옳은 답변을 제시하기 때문이다. 두 상자의 내용물을 모두 가지는 것이 합리적인 선택이다.

---

[8] Nozick, R. 1969. pp.114-115.

CDT의 판단이 옳은 이유는 문제의 선택 행위가 두 번째 상자에 $1,000,000이 있거나 없거나 하는 데에 아무런 영향을 끼치지 않기 때문이다. 두 사건이 인과적으로 독립적이라는 이러한 믿음에 근거할 때, 지금 $1,000이 들어있는 첫 번째 상자의 내용물을 가지지 않기로 결정하는 것은 $1,000을 더 얻을 수 있는데도 그렇게 하지 않는 것이기 때문에 비합리적이다.

한 상자의 내용물만을 가지는 것이 합리적인 선택이라는 EDT의 주장은 두 사건의 인과적 독립성을 무시한 결과이다. EDT는 과거의 예측이 모두 정확했고, 그에 따라 두 번째 상자만을 선택한 사람이 $1,000,000을 얻었다는 증거에만 초점을 맞춰, 한 상자의 내용물만을 가지기를 권한다. 이러한 권유를 좀 더 상세하게 분석해 보면 다음과 같은 추리 과정을 찾을 수 있다. 즉 당신은 존재자의 예측이 정확하게 들어맞을 것으로 확신하고, 따라서 존재자는 당신이 지금 두 상자의 내용물을 모두 가지기로 결정할 때 그 점을 정확하게 예측하여, 두 번째 상자에 $1,000,000을 넣지 않았을 것이고, 그 결과 당신은 $1,000밖에 못 가지게 될 것이라는 추리이다. 하지만 이러한 추리는 잘못되었다. 지금의 상황은 존재자가 이미 예측을 마친 후이고, 그에 따라 두 번째 상자에는 $1,000,000이 있거나 없거나 결정되어 있기 때문이다. 존재자의 예측은 이제 바뀔 수가 없다. 마찬가지로 두 번째 상자에 $1,000,000이 있거나 없거나 하는 것은 이제 바뀔 수가 없다. 그렇다면 $1000이라도 가질 수 있도록 지배의 원리에 따라 두 상자를 모두 선택하는 것이 합리적인 선택이다.[9]

---

9) 노직의 뉴컴 문제에서 EDT와 CDT 각각이 두 상자를 선택하는 Two의 기대 효용성을 계산하는 방식은 다음과 같다. EDT는 One과 Two의 기대 효용성을 다음처럼 계산한다.

EU(One) = Pr(Ptwo | One)Ut(One&Ptwo) + Pr(Pone | One)Ut(One&Pone)

EU(Two) = Pr(Ptwo | Two)Ut(Two&Ptwo) + Pr(Pone | Two)Ut(Two&Pone)

뉴컴 문제는 좋은 증거에도 불구하고 실제의 결과가 좋지 않을 수 있는 예외적인 경우이다. 증거가 그럴 듯 하지 않더라도 가장 좋은 결과가 발생할 수 있는 경우를 고안해 낸 것이다. 그럼에도 불구하고 EDT는 실제로 좋은 결과를 가져오는 선택을 하지 않고, 기분이 좋은 정보를 쫓아 선택을 하도록 권하기 때문에 비판받는 것이다. 반면 CDT는 그럴듯한 증거를 확보하기보다는 실제로 좋은 결과를 가져오는 데 초점을 맞춘다. 하나의 상자를 고르는 것은 $1,000,000이 있으리라는 길조일 뿐 막상 두 상자를 모두 선택함으로써 가장 좋은 결과 즉 추가적으로 $1,000을 가질 수 있는 결과를 낳지 않는다. 그래서 EDT는 효력이 부족한 잘못된 판단을 하는 것이라고 비판받는 것이다.

이상에서 드러나는 EDT와 CDT의 차이점은 길조(auspiciousness)와 효력(efficacy)이라는 선택 기준의 차이로 설명된다. EDT는 길조에 따라 선택을 하는 반면 CDT는 효력에 따라 선택을 한다. 길조는 좋은 일이 일어날 것으로 보이는 증거이다. 하지만 좋은 증거가 많은 선택지라고 해서 언제나 그 선택지가 효력을 발휘하는 것은 아니다. 대개의 경우에는 길조가 좋은 것과 효력이 좋은 것이

이 때 EDT는 존재자가 정확한 예측을 하였다는 증거를 근거로 Pr(Ptwo | Two)와 Pr(Pone | One)에 매우 높은 값을 부여한다. 또 Ut(One&Pone)와 Ut(Two&Pone)는 매우 높은 값을 지닌다. 따라서 EU(One) > EU(Two)의 결과가 도출된다. EDT가 하나의 상자 즉 One을 선택하는 이유는 이러한 기대 효용성의 계산에 따른 것이다.

EDT와 달리 CDT는 One과 Two의 기대 효용성을 다음처럼 계산한다.

EU(One) = Pr(One→Ptwo)Ut(One&Ptwo) + Pr(One→Pone)Ut(One&Pone)

EU(Two) = Pr(Two→Ptwo)Ut(Two&Ptwo) + Pr(Two→Pone)Ut(Two&Pone)

그런데 Pr(One→Ptwo), Pr(One→Pone), Pr(Two→Ptwo), Pr(Two→Pone)는 각각 Pr(Ptwo), Pr(Pone), Pr(Ptwo), Pr(Pone)과 같다. 지금의 선택이 존재자의 예측에 아무런 영향을 미치지 못하기 때문이다. 이 경우 Ut(One&Ptwo) < Ut(Two&Ptwo)이고, Ut(One&Pone) < Ut(Two&Pone)이기 때문에 CDT는 두 상자 즉 Two를 선택하는 것이 합리적이라고 한다.

일치하지만, 좋은 증거와 어긋나는 선택지가 효력을 발휘할 수도 있다. 따라서 뉴컴의 문제와 마찬가지로 길조와 효력이 충돌하여 서로 다른 선택지를 고르도록 제안한다면, 합리적 선택의 기준은 길조가 아니라 효력이 있는 선택지를 고르는 것이어야 할 것이다. 결국 올바른 결정 이론이라면 합리적인 결정을 옹호하고, 비합리적인 결정을 옹호하지 않아야 할 텐데, 뉴컴 문제에서 EDT는 효력이 아니라 길조를 따르는 비합리적인 결정을 옹호한다. 따라서 EDT는 올바른 결정 이론이 될 수 없다. 이러한 비판에 따라 많은 철학자들이 EDT를 포기하고 CDT로 입장을 바꾼 것이다.

3. 의학의 뉴컴 문제와 인과적 믿음

노직의 뉴컴 문제를 역설이라고 부르는 경우가 있는데 그것은 사실 꽤 많은 사람들이 CDT가 제시하듯이 두 상자를 모두 가지겠다고 선택하기보다는 하나의 상자를 선택하기 때문이다. 이러한 사람들의 선택이 비합리적인가? 뉴컴 문제를 변형한 다음 경우를 생각해보자.

마술사의 뉴컴 문제는 노직의 뉴컴 문제에 나타난 존재자를 마술사로 바꾼 것이다. 이 마술사는 조작과 속임수에 능하기로 유명하다. 이 마술사는 뉴컴 문제를 제시하며 사람들의 선택을 정확하게 예측한 전력을 가지고 있다. 지금껏 한 상자를 선택한 사람은 $1,000,000을 가져가고 두 상자를 선택한 사람은 $1,000을 가져갔다. 이제 이 마술사는 당신에게 두 가지 선택지 중 하나를 고르도록 한다. 마술사는 노직이 제시한 상황에서와 마찬가지로 먼저 당신의 선택을 예측하고, 그 예측에 따라 $1,000,000을 두 번째 상자에 넣거나 넣지 않으며, 그 후 당신으로 하여금 어떤 선택을 할 것

인지 결정하도록 한다. 당신은 일단 당신의 선택이 두 번째 상자에 $1,000,000이 있거나 없거나 하는 데에 아무런 영향을 끼치지 않을 것이라고 판단한다. 하지만 곰곰이 생각해보니 당신은 이 판단을 신뢰하기 어렵다는 점을 깨닫는다. 그 마술사는 조작 및 속임수를 잘 쓰기로 유명할 뿐만 아니라, 지금까지 천 번이 넘는 유사한 상황에서 당신 및 다른 사람의 선택을 정확하게 예측하였기 때문이다. 당신은 마술을 믿지 않기에 마술사가 그렇게 정확하게 예측하는 것은 불가능하다고 생각한다. 이에 당신은 당신의 선택이 실제로 두 번째 상자에 $1,000,000이 있거나 없거나 하는 데에 영향을 끼친 것이라고 의심한다. 예를 들어 그 마술사는 당신이 두 박스의 내용물을 모두 가지겠다는 선택을 하고 그 선택에 따른 버튼을 누르는 순간 두 번째 상자 속에 놓여 있던 $1,000,000이 사라져 버리도록 조작을 하고 있는지도 모른다고 의심한다. 그 마술사의 예측이 언제나 정확했던 것으로 보이는 이유는 결국 그와 같은 조작에 의한 것이었고, 그렇다면 당신의 선택에 따라 두 번째 상자에 $1,000,000이 있거나 없거나 할 수도 있는 것이다. 간단하게 두 상자를 모두 선택하는 것이 합리적으로 보이는 것은 당신의 선택이 두 번째 상자에 대해 아무런 영향을 끼치지 않을 것이라고 믿기 때문인데, 마술을 믿지 않는 당신으로서는 마술사가 과거에 천 번 넘게 정확한 예측을 했다는 사실이 이러한 믿음을 의심하는 근거가 된다. 이제 당신은 어떻게 할 텐가? 두 상자를 모두 가지겠다고 선택하는 것이 합리적인 결정처럼 보이는가?

위의 문제에서 선택 상황은 다음과 같은 표로 제시될 수 있다.

<표 2> 마술사의 뉴컴 문제

선택지	선택 결과 두 번째 상자에 $1,000,000이 들어있다	선택 결과 두 번째 상자에 $1,000,000이 들어있지 않다
한 상자(두 번째 상자)의 내용물만을 가진다	$1,000,000	$0
두 상자의 내용물을 모두 가진다	$1,001,000	$1000

위의 <표 2>의 분석이 앞의 <표 1>의 분석과 달라진 부분은 세계의 상태를 기술할 때 <표 1>은 존재자의 예측을 따졌지만 <표 2>는 상자에 들어 있는 돈을 따진 것이다. 이러한 변화는 존재자의 예측에 관한 기술이 행위자의 결정 시점 이전에 발생한 것을 분명하게 밝힌 것과 달리 상자에 들어 있는 돈에 관한 기술이 행위자의 결정 시점 이후에 발생한 것일 수 있게 한 것이기에 커다란 차이점으로 나타난다. 그것은 바로 노직의 뉴컴 문제를 마술사의 뉴컴 문제처럼 해석할 수 있는 여지를 주기 때문이다. 또 많은 사람들이 CDT가 제시하듯이 두 상자를 모두 가지겠다고 선택하기보다는 하나의 상자를 선택하는 이유는 그들이 비합리적이기 때문이기보다는 노직의 뉴컴 문제를 마술사의 뉴컴 문제처럼 이해하기 때문이다.

이제 상황은 두 상자를 고르는 선택만이 합리적이라고 말하기 어렵게 됐다. 마술사의 전력과 천 번이 넘는 정확한 예측의 성공이라는 증거는 하나의 상자를 선택하는 것이 합리적이라고 말할 수 있는 근거가 된다. 당신의 선택이 두 번째 상자에 대해 아무런 영향을 끼치지 않을 것이라는 인과적 믿음에 대해 의심하게 된 것이다. 바로 이 점이 노직의 두 상자 문제 상황과 예측이 100% 확실하다고 가정하는 상황을 구분해준다. 예측이 100% 확실하다고 가정할 때 선택

의 문제는 간단하게 $1,000,000이냐 $1,000이냐로 환원될 수 있다. 하지만 그렇다고 해도 소벨(H. Sobel)이 반박하듯이 여전히 더 좋은 선택지 즉 $1,000을 더 가져갈 수 있는 선택지가 존재한다.[10] 이와는 달리 마술사의 사례는 당신의 선택이 두 번째 상자에 대해 영향을 끼칠 수 있다고 보기 때문에 더 좋은 선택지 즉 $1,000을 더 가져갈 수 있는 선택지가 존재하지 않을 수도 있게 된다.

사실 노직은 뉴컴 문제를 제시함에 있어서 사건들 간의 인과적 관계에 관해 명확하게 언급하지 않는다. 다만 당신이 선택하려는 순간에 존재자는 이미 두 번째 상자에 $1,000,000을 넣어 두었거나 넣어두지 않았거나 행동을 마친 상황이라는 사실을 밝혔을 뿐이다. 이 단서는 대개 당신의 선택 행위가 두 번째 상자에 $1,000,000을 넣거나 넣지 않거나 하는 존재자의 행위에 아무런 인과적 영향력을 미치지 못한다는 점을 밝히고자 한 것으로 이해할 수 있다. 하지만 그 점은 명확하지 않다. 꼭 그렇게 해석해야만 하는 것은 아니다. 노직의 사례를 마술사의 사례와 유사하게 해석하는 것이 완전히 불가능하다고 말하기는 어렵다. 많은 사람들이 두 상자가 아니고 한 상자를 선택하는 이유는 노직의 뉴컴 문제를 마술사에서의 경우처럼 이해하기 때문일 수 있다. 사람들은 자신의 선택이 존재자의 행위에 아무런 인과적 영향력을 미치지 못하리라는 믿음보다는 마술사의 정확한 예측이 불가능하다는 믿음을 더 그럴듯하다고 판단한 것이다.

물론 노직의 사례를 마술사의 사례처럼 이해하는 것이 가능하다고 해서 한 상자를 선택하는 것만이 합리적인 선택이라고 말할 수 없다. 또 그렇게 이해하는 것은 뉴컴 문제를 오독하는 것일 수 있다. 오독이라면 그것은 예측에 따라 상자에 돈을 넣어 두었다는 존재자의 말을 믿지 못하는 것이 뉴컴 문제에서 제기하고자 하는 논

---

10) Sobel (1994) pp.100-107.

점을 벗어난 것이기 때문이다. 상자의 선택과 돈이 들어있다는 사건은 인과적으로 독립적이라는 것이 뉴컴 문제의 핵심이라고 할 때 그런 인과적 독립성을 의심하는 것은 문제의 의도에 부합하지 않는 것일 수 있다. 하지만 시간을 역행하여 인과적 영향력을 행사할 수 있다는 믿음이 흔들리기 어려운 것처럼 노직의 뉴컴 문제이든 마술사의 문제이든 간에 그렇게 정확한 예측을 할 수 있다는 사실을 받아들이기 어렵다고 판단하는 것이 잘못은 아니다. 그렇게 정확한 예측을 할 수 있다는 가정 역시 대단히 비현실적이다.

의학의 뉴컴 문제(Medical Newcomb Problem)는 인과 관계를 명확히 밝힘으로써 노직의 뉴컴 문제가 지닌 비현실적 가정과 여러 가지 오해의 소지를 제거하고 EDT를 확실하게 반박하고자 한다. 의학의 뉴컴 문제 역시 여러 가지 사례로 제시될 수 있는데, 가장 많이 논의되고 있는 다음의 두 사례를 살펴보자.

<코코의 초코렛 사례>
코코는 초코렛을 먹을지 말지를 결정하려 한다. 초코렛을 좋아하긴 하지만 코코는 초코렛을 먹고 나면 편두통 때문에 고생하는 경우가 많다. 코코는 또 자신이 편두통이 일어나기 전 상태, 즉 감각장애, 운동장애, 기분장애 등의 증상을 동반하는 PMS를 경험하는 경우가 많은데, 이때 PMS는 코코로 하여금 초코렛을 먹고 싶은 생각이 들게 하는 원인이며 동시에 편두통 때문에 고생하게 하는 원인이다. 이러한 사실을 잘 알고 있는 코코는 초코렛을 먹고 싶은 생각은 있지만 편두통을 앓을 생각은 없다. 그렇다면 코코는 초코렛을 먹는 것이 좋을까 아니면 먹지 않는 것이 좋을까?[11]

<명수의 흡연 사례>
명수는 흡연을 할 것인가 말 것인가를 고민하고 있다. 명수는 흡연이 폐암과 높은 상관관계에 놓여 있지만 그 상관관계는 흡연과 폐암을 모두 일으키는 제 3의 공통원인 때문이라고 믿고 있다. 이 제 3의 공통원인이 있다고 하든가 아니면 없다고 하든가 둘

11) 이 사례는 프라이스가 검토하고 있는 사례를 조금 변형한 것이다. H. Price (1991) p.162.

> 중 하나로 고정시켜 놓으면 더 이상 흡연과 폐암은 상관관계에 놓여 있지 않게 된다. 명수는 물론 폐암에 걸리지 않으면서 동시에 흡연을 하지 않기보다는 폐암에 걸리지 않으면서도 흡연할 수 있기를 바란다. 또한 명수는 폐암에 걸릴 바에야 흡연을 하지 않기보다는 차라리 흡연할 수 있기를 바란다. 그렇다면 명수는 흡연을 하는 것이 좋을까 아니면 하지 않는 것이 좋을까?[12]

위의 두 사례와 같은 의학의 뉴컴 문제는 노직의 뉴컴 문제보다 현실적 사례이면서 동시에 더욱 분명한 정답을 제시한다는 점에서 EDT와 CDT 사이의 논쟁을 확실하게 CDT의 승리로 정리해주는 사례로 알려져 있다.[13] 노직의 뉴컴 문제가 큰 돈을 나눠주며 정확한 예측을 할 수 있는 존재자를 가정해야 했던 것처럼 비현실적인 것과 달리 의학의 뉴컴 문제는 꽤 그럴듯한 현실적인 상황이다. 또 노직의 뉴컴 문제에서는 한 상자를 선택하는 것이 합리적인 선택이라는 주장이 있기에 무엇이 합리적인 선택인가의 문제에 있어 논란이 있고 직관의 차이가 존재하는 것으로 보인다.[14] 이와 달리 코코가 직면한 문제에서 초코렛을 먹는 것이 코코에게 합리적인 선택이라는 점에 동의하지 않을 사람이 없고, 명수의 경우에도 흡연을 하는 것이 합리적인 선택이라는 점에 동의하지 않을 사람이 없다. 이런 두 가지 측면은 일반적으로 의학의 뉴컴 문제가 노직의 뉴컴 문제보다 EDT를 반박하기에 더 적절한 근거로 평가된다.

---

12) 이 사례는 이건이 검토하고 있는 사례를 조금 변형한 것이다. A. Egan (2007) p.94.

13) 이 점은 CDT 옹호자뿐만 아니라 EDT 옹호자인 프라이스(H. Price 1986, 1991) 등도 동의하는 부분이다.

14) T. Horgan(1981), P. Horwich(1987) 등이 한 상자를 고르는 것이 실제로 더 많은 돈을 벌게 된다는 이유로 이러한 주장을 펼친다. 와이릭(P. Weirich)에 따르면, CDT는 뉴컴 문제가 바로 그런 비합리적인 선택이 돈을 더 벌수 있도록 예외적인 상황을 고안해낸 것이라고 답변한다. 한 상자 선택이 돈을 더 벌게 하더라도 한 상자 선택은 비합리적이라는 것이다. Weirich, P., (2008)

그러나 이러한 평가가 정확한가의 문제는 논란의 소지가 있다. 먼저 코코의 상황은 원래의 뉴컴 문제보다 현실적이라고 할 수 있지만, 합리적 결정 이론의 도움을 필요로 하는 대부분의 문제는 의학의 뉴컴 문제처럼 인과 관계를 명확하게 제시하지 않는다. 사실 인과 관계가 명확하다면, 선택의 문제가 그렇게 어렵지는 않을 것이다. 그런 측면에서 코코의 경우 초코렛을 먹는 것이 합리적인 선택이라는 결정에 모두 동의하는 것은 EDT를 반박함에 있어 장점으로 작용하기보다는 오히려 단점으로 작용할 수 있다. 즉 원래의 뉴컴 문제가 여러 가지 가정을 필요로 한다는 점에서 비현실적이긴 하지만 의학의 뉴컴 문제처럼 인과 관계를 명료하게 밝히지 않았기 때문에 합리적 결정 이론의 논의에 오히려 더 적절하고 현실적인 사례로 다가올 수 있다는 것이다. 이러 측면에서 뉴컴 문제를 제시할 때 인과 관계를 명료하게 밝힐 필요가 있는지가 의심스럽다. 명료하게 밝히지 않더라도 인과 관계를 한정하는 방식은 노직의 경우처럼 시간의 선후 관계를 명확하게 한다거나 하는 것처럼 여러 가지 방법이 있을 것이기 때문이다.

더욱 중요한 점은 의학의 뉴컴 문제가 원래의 뉴컴 문제보다 더 확실하게 EDT를 반박하는 것인지 분명치 않다는 것이다. 코코의 사례에서 중요한 점은 노직의 뉴컴 문제 상황과 달리 선택 행위를 하는 행위자가 인과 관계에 대해 명확하게 알고 있다는 것이다. 즉 코코는 PMS가 초코렛 및 편두통의 공통 원인이며 초코렛을 먹는 행위가 편두통을 앓는 결과를 야기하지는 않는다는 사실을 분명히 알고 있다. 이 점은 의학의 뉴컴 문제를 노직의 뉴컴 문제처럼 마술사의 문제로 오독하지 않도록 확실하게 차단한다. 이 점은 의학의 뉴컴 문제의 장점이다.

그러나 이처럼 관련된 인과적 판단을 정확하게 알고 있다고 가정할 경우 코코가 초코렛을 먹지 말아야겠다고 결정할 근거는 무

엇인가? 명수가 흡연을 하지 말아야겠다고 결정할 근거는 무엇인가? 그것은 모두 통계적 상관 관계이다. 그러나 코코와 명수는 그 통계적 상관 관계가 아무런 인과적 힘을 지니지 않는다는 사실 또한 알고 있다. 이것이 치명적이다. 의학의 뉴컴 문제는 아무런 인과적 힘을 지니지 않는 통계적 상관 관계에 대해 정확하게 알고 있으면서도 그 상관 관계를 증거로 삼아 결정을 내리는 행위자를 비판한다. EDT를 따르는 행위자가 초코렛을 먹지 말아야 한다고 결정하고 흡연을 하지 말아야 한다고 결정하는 것이 잘못인 것은 이처럼 아무런 인과적 힘을 지니지 않는 통계적 상관 관계에 따른 결과라는 것이다. 그 결정이 잘못이라는 점은 분명하다. 하지만 과연 이와 같은 결정은 EDT에 따른 결정인가가 문제이다. 노직의 뉴컴 문제에서는 과거에 존재자가 많은 수의 성공적인 예측을 했다는 증거가 있었다. 그러한 증거가 EDT로 하여금 한 상자를 고르는 것이 합리적 선택이라는 유혹을 했다. 그것은 인과적 힘이 없다고 확신하기 어렵게 유혹할 수 있었다. 그런데 지금 의학의 뉴컴 문제에서는 그런 유혹을 할 수가 없이 증거를 너무 미약하게 바꿔버렸다. 인과 관계가 없다는 것을 명료하게 진술함으로써 인과 관계가 있는 것으로 기대하고 행동하는 것이 모순적임을 분명하게 했다. 이에 의학의 뉴컴 문제는 EDT를 따르는 행위자가 모순적 결정을 한다고 비판하는 것이 된다. 이것은 허수아비 논증의 오류를 범하는 꼴이다.

인과적 믿음이 EDT에서 어떤 위치를 차지하게 되는지에 대한 정확한 이해가 필요하다. EDT는 인과적 믿음이 그 행위자의 비인과적 믿음에 의해 충분히 드러난다는 입장이다. 그것은 EDT가 인과적 믿음을 인정하지 않는다는 의미가 아니다. EDT가 인정하는 인과적 믿음은 행위자의 배경지식에 포함될 수 있다. 즉 $Pr(O_i \text{ if } A)$를 $Pr(O_i \mid A)$라고 해석할 때 $Pr(O_i \mid A)$는 좀 더 엄밀하게 말하

면 배경지식 K를 포함하여 $Pr(O_i \mid A\&K)$에 의해 계산된다는 것이다. 이 때 배경지식에 포함된다는 것은 충분한 증거에 기반하고 있다는 것을 의미할 수 있다. 그렇지 않더라도 일부 증거에 의해 그 배경지식의 내용이 반박될 것으로 보기는 어렵다. 기준이 불분명하지만 충분한 증거가 있어야만 배경지식의 내용이 변경될 수 있다. 따라서 그러한 배경지식을 무시하고 배경지식과 어긋나는 일부 증거에 따라 행위자가 결정을 내리는 것은 EDT를 잘못 적용하는 것이다. 전체 증거를 모두 공평하게 고려하지 않는 판단이다. 그것은 EDT가 틀린 것이 아니라 EDT를 올바르게 적용하지 못한 것이다. 따라서 인과관계를 무시하고 상관관계에 따라 결정을 내리는 것은 EDT를 올바르게 적용하지 못한 것이라고 할 수 있다.

코코의 경우도 마찬가지이다. 코코는 PMS, 초코렛 및 편두통 사이의 인과적 관계에 대해 잘 알고 있다. 따라서 코코는 그 인과적 관계에 대한 믿음을 지니기 위해 필요한 비인과적 믿음을 충분한 증거로 확보하고 있다고 볼 수 있다. 적어도 이 믿음이 쉽게, 갑자기 변경될 수 있는 것은 아니다. 이에 비하면 초코렛과 편두통 사이의 상관관계에 관한 증거는 초코렛을 먹으면 편두통을 앓을 위험이 있다는 믿음을 지지하더라도 매우 미약하게 지지할 것이다. 이 사실을 코코는 잘 알고 있다. 코코는 초코렛과 편두통이 인과관계에 놓여 있지 않다는 믿음을 지니고 있는 것이다. 따라서 EDT가 이러한 인과적 믿음을 무시하고 당장 눈 앞에 제시된 상관관계만을 고려하여 코코에게 초코렛을 먹지 말 것을 권한다면 그것은 전체 증거를 공평하게 고려하지 않고 일부 증거만을 고려하여 추리하는 논리적 오류를 범하는 것이다.

의학의 뉴컴 문제를 통해 EDT를 비판하는 입장에서는 아마도 코코의 비인과적 믿음이 단지 초코렛과 편두통 사이의 상관 관계에 대해 알려줄 뿐이고, PMS, 초코렛 및 편두통 사이의 인과적 관

계에 대해 알려 주지 않는다고 판단하는 듯 하다. 이러한 지적은 EDT가 인과적 관계를 단지 증거를 통해 정확하게 포착할 수 있는지를 의심하기 때문에 제시될 수 있다. 그러나 이러한 지적은 CDT 역시 인과적 관계를 어떻게 알 수 있는지에 대해서는 침묵하고 있기에 적절한 반론이 될 수 없다.[15] 의학의 뉴컴 문제를 성립시키기 위해 코코는 PMS, 초코렛 및 편두통 사이의 인과적 관계를 잘 알고 있다는 전제를 필요로 하고, 명수는 흡연이 폐암의 원인이 아니라는 사실을 잘 알고 있다는 전제가 요청된다. 그러한 전제가 없다면 의학의 뉴컴 문제는 노직의 뉴컴 문제와 다른 점이 없다. 하지만 그러한 인과적 믿음을 명확하게 전제할 때 반박 가능한 EDT는 매우 미약한 형태의 EDT일 뿐이다.[16] 의학의 뉴컴 문제가 EDT를

---

15) 카트라이트(N. Cartwright)는 뉴컴 문제와 유사한 종류의 결단 문제들이 인과 관계의 존재를 증명하는 증거라고 본다. 인과가 환원불가능하고 제거될 수 없다는 것이다. 그러나 이러한 주장은 인과 관계를 원초적인 것으로 만들뿐 그것이 무엇인지에 대한 설명을 제시하는 것이 아니다. N. Cartwright (1979)

16) 아래 제시된 사례는 노직의 뉴컴 문제를 좀 현실적인 경우로 바꾸었지만 의학의 뉴컴 문제와 마찬가지로 인과적 믿음을 명확하게 밝혔기 때문에 반박의 대상이 되는 EDT를 미약한 이론으로 만드는 한계를 지닌다. 이 사례는 원래 소벨의 사례를 조이스가 변형한 것인데, 이 글에서도 논자의 의도에 맞게 조금 더 변형을 가하였다. Joyce, J., (1999)

철수는 아주 똑똑한 갑부 심리학자를 친구로 두고 있다. 이 심리학자는 철수와 오랜 친구사이로 철수를 잘 알고 있어서 철수가 어떤 행동을 할지에 대해 매우 정확하게 예측할 수 있다. 그러던 어느 날 철수는 은행에 가던 길에 심리학자를 만난다. 심리학자는 10만원 현금을 손에 들고서 철수에게 이렇게 말한다. "원한다면 이 돈을 가지게. 하지만 이 점을 잘 생각해서 선택하게나. 내가 어제 은행에서 자네가 어떤 선택을 할지에 대해 예측을 했다네. 자네가 이 돈 10만원을 가지지 않을 것으로 예상했다면 나는 자네 은행 계좌에 100만원을 이체시켰을 것이네. 하지만 자네가 이 돈 10만원을 가질 것으로 예상했다면 나는 자네 은행 계좌에 한 푼도 이체시키지 않았을 것이네. 물론 자네 계좌에 이체된 돈을 내가 달라고 할 일은 없을 걸세. 내가 지금 내 마음대로 자네 계좌의 돈을 이체할 수도 없는 것이고 말일세. 지금 자네 계좌에 있는 돈은 모두 자네 것이네. 100만원이 지금 자네 계좌

반박함에 있어 노직의 뉴컴 문제보다 더 좋은 사례라는 결론은 EDT를 이미 틀린 이론이라고 판단하기 때문이다.

4. 쌍둥이 죄수의 딜레마

의학의 뉴컴 문제가 EDT를 반박함에 있어 지닌 단점은 쌍둥이 죄수의 딜레마를 검토하는 과정에서 더욱 분명하게 제시될 수 있다. 죄수의 딜레마는 다음처럼 제시될 수 있다. 두 죄수는 서로 다른 방에 들어간 후 검사에게 자백을 할 것인지 말 것인지를 결정하게 된다. 둘 다 서로를 배신하여 자백할 수도 있고, 서로 협조하여 묵비권을 행사할 수도 있다. 둘이 서로 협조하는 것이 서로 배신하는 것보다 좋은 결과를 낳는다. 둘의 이익과 손해를 한꺼번에 따지면 그렇다. 하지만 각각 개인으로 보면 서로를 배신하여 자백하는 것이 더 낫다. 다른 죄수가 어떤 결정을 내리더라도 언제나

에 있다면 있을 것이고 없다면 없을 것이라는 점을 기억하게나. 그 점은 자네가 지금 어떤 선택을 한다고 해서 바뀌는 것이 아니라네. 자 그럼 이 돈 10만원을 가질텐가?"

철수는 이 심리학자가 200명의 사람에게 이 실험을 수행했다는 사실을 잘 알고 있다. 철수는 그 200명 중 100명은 10만원을 가져갔고 나머지 100명은 10만원을 가져가지 않았으며, 단 한 사람을 빼고 나머지 199명에 대해서는 심리학자가 정확하게 예측했다는 사실을 알고 있다. 철수는 심리학자가 미래를 볼 수 있는 어떤 신비한 기구를 가지고 마술을 부리는 것이 아니라는 점 역시 잘 알고 있다. 그 심리학자는 철수를 만나기 이전까지 습득한 지식에 근거해 철수의 행위를 예측한 것이다. 철수가 어떤 선택을 할지를 보여주는 유전자가 있다면 그 심리학자는 철수가 그 유전자를 지녔다는 사실을 알고 있을 수도 있다. 그 심리학자는 이런 선택의 순간에 철수가 어떻게 선택하는지를 세세하게 연구해 왔을 수도 있다. 어쨌거나 중요한 점은 지금 철수가 어떤 선택을 하는 것이 그 심리학자가 철수의 계좌에 100만원의 돈을 넣거나 넣지 않거나 하는 데에 아무런 인과적 영향력을 행사하지 못한다는 것이다.

더 좋은 결과를 가져오게 된다.

이제 죄수의 딜레마를 변형하여 두 죄수가 쌍둥이라는 가정을 해보자. 둘은 생각하는 방식이 같다. 많은 결정 상황에서 각자 독자적으로 판단을 했음에도 불구하고 둘의 결정은 똑같았다. 둘은 또한 자기들이 생각하는 방식이 같다는 사실을 잘 알고 있다. 그래서 그 중 한 사람이 묵비권을 행사하기로 결정한다면 그 사람은 자신의 쌍둥이도 또한 묵비권을 행사하기로 결정할 것이라고 결론짓는다. 즉 자신이 묵비권을 행사하기로 결정하는 것은 자신의 쌍둥이가 묵비권을 행사하기로 결정할 것이라는 좋은 증거가 된다. 이제 당신의 그 쌍둥이 중 한 사람이라고 해 보자. 당신은 어떻게 행동하겠는가?

이처럼 제시할 수 있는 쌍둥이 죄수의 딜레마는 노직의 뉴컴 문제와 같은 구조를 지닌다. 이를 보기 위해 먼저 소벨(H. Sobel)에 따라 노직의 뉴컴 문제를 구성하는 핵심적인 조건을 다음 세 가지로 분석해 보자.[17] 첫째(NC1), 노직의 뉴컴 문제의 기대 효용값은 다음과 같은 구조를 지닌다.

<표 3> 노직의 뉴컴 문제의 기대 효용값 구조

선택지	상태 S1 (한 상자 선택 예측)	상태 S2 (두 상자 선택 예측)
행위 A (한 상자 선택)	x	w
행위 B (두 상자 선택)	z	y

17) 소벨(H. Sobel)이 제시한 노직의 뉴컴 문제의 넷째 조건 즉 노직의 뉴컴 문제의 셋째 조건 때문에 행위 A가 지니는 길조는 행위 B가 지니는 길조보다 크다는 조건은 셋째 조건에서 함축되는 것으로 보아 제외한다. Sobel (1994) pp.78-80.

이때 <표 3>에서 x=Ut(A&S1), w=Ut(A&S2), z=Ut(B&S1), y=Ut(B&S2)이고 z>x>y>w이다. 둘째(NC2), 상태 S1과 S2는 행위 A 및 B와 인과적으로 독립적이다. 셋째(NC3), 인식적 판단에 따르면 상태는 행위에 의존한다. 따라서 Pr이 주관적 믿음의 정도를 나타낼 때 Pr(S1 | A)≃1이고 Pr(S2 | B)≃1이다(여기서 ≃는 거의 같다는 의미이다). 또 Pr(S1&S2)=0이고 Pr(A∨B)=1이기 때문에 Pr(S2 | A)≃0이고 Pr(S1 | B)≃0이다.

이제 쌍둥이 죄수의 딜레마는 다음과 같은 조건을 가진 것으로 볼 수 있다. 첫째(PC1), 기대 효용값은 다음과 같은 구조를 지닌다.

<표 4> 쌍둥이 죄수의 딜레마의 기대 효용값 구조

선택지	상태 S1* 자백하지 않기	상태 S2* 자백하기
행위 A* 자백하지 않기	-1	-10
행위 B* 자백하기	0	-5

둘째(PC2), 상태 S1*과 S2*는 행위 A* 및 B*와 인과적으로 독립적이다. 셋째(PC3), 다른 죄수가 나와 똑같은 결정을 할 것이라는 점을 확신한다. 넷째(PC4), 각 죄수가 서로 다른 방에서 취조를 받는다.

이제 노직의 뉴컴 문제를 구성하는 조건들과 쌍둥이 죄수의 딜레마를 구성하는 조건들을 비교해보면, NC1과 NC2는 PC1과 PC2와 같다고 볼 수 있다. PC4는 노직의 뉴컴 문제를 구성하는 조건에서는 찾아볼 수 없는 내용이다. 이 조건은 두 죄수가 서로 협의를 할 수 없도록 하려는 것이다. 쌍둥이 죄수의 딜레마가 하나의 뉴컴 문제를 제시하는 것이 아니라 각각의 죄수에게 뉴컴 문제를

일으키는 것이기에 PC4는 노직의 뉴컴 문제와 쌍둥이 죄수의 딜레마를 비교하는 측면에서는 무관한 조건이라고 할 수 있다. 따라서 PC4는 논외로 할 수 있다.

그럼 이제 남은 것은 NC3가 PC3과 동일한 조건으로 볼 수 있는가 하는 것이다. PC3은 쌍둥이인 다른 죄수가 나와 똑같은 결정을 할 것이라고 확신하는 것인데, 이 조건은 내가 자백하면 나의 쌍둥이도 자백할 것이라고 믿는 정도가 매우 높다는 것을 의미한다. 즉 NC3를 함축하는 것으로 볼 수 있다. 역으로 NC3에서 $Pr(S1 \mid A) \simeq 1$이고 $Pr(S2 \mid B) \simeq 1$라는 것은 두 쌍둥이가 동일한 결정을 할 확률이 높다고 확신하는 것이기에 PC3을 함축하는 것으로 볼 수 있다.[18] 이상의 논의에서 쌍둥이 죄수의 딜레마는 노직의 뉴컴 문제와 동일한 구조를 지닌다고 볼 수 있다. 차이가 있다면 노직의 뉴컴 문제는 시간 관계로 인과 관계를 한정하지만 쌍둥이 죄수의 딜레마는 서로 다른 사람의 결정 관계로 인과 관계를 한정한다는 것이다. 이 점 역시 PC4처럼 쌍둥이 죄수의 딜레마가 뉴컴 문제인가라는 결론을 도출하는 데 무관한 문제이다. 한편 쌍둥이 죄수의 딜레마는 마술사의 사례처럼 해석할 수 있는 여지를 남기지 않게 설계할 수 있다. 나아가 쌍둥이 죄수의 딜레마는 의학의 뉴컴 문제와 달리 인과 관계를 명확하게 밝히지 않으면서도 길조의 증거를 충분히 제시한다. 그런 측면에서 쌍둥이 죄수의 딜레마

---

18) 루이스(D. Lewis)는 그의 논문 제목 "죄수의 딜레마는 뉴컴 문제이다"가 시사하듯이 죄수의 딜레마를 뉴컴 문제라고 주장했다. 좀 더 정확하게 말하자면 루이스는 죄수의 딜레마는 각 죄수가 뉴컴 문제에 부닥치게 되기 때문에 두 개의 뉴컴 문제라고 주장했다. D. Lewis (1986). 그러나 소벨(H. Sobel)은 이와 같은 루이스의 주장을 조심스럽게 이해해야 한다고 주장한다. 즉 모든 죄수의 딜레마 문제가 뉴컴 문제라고 해석하는 것이 옳지 않다는 것이다. 소벨의 논의는 뉴컴 문제에 해당하지 않는 죄수의 딜레마 문제가 있을 수 있다는 것인데 그것은 쌍둥이 죄수의 딜레마와는 구분될 수 있다. 자세한 논의는 다음 참조. Sobel (1994) pp.78-80.

는 EDT를 반박하고 CDT를 옹호하기에 다른 유형의 뉴컴 문제보다 더 설득력을 지닐 수 있다.

5. 나가는 말

지금까지의 논의는 다음처럼 정리할 수 있다. 첫째, EDT와 CDT의 차이는 행위의 기대 효용성을 계산하는 방식에서 구분된다. 특히 이 차이점은 $Pr(O_i \text{ if } A)$을 계산하는 방식에서 드러나는데 이것은 인과적 믿음이 구체적으로 어떤 역할을 하는가에 대해 서로 다른 입장에서 비롯된다. 둘째, 노직의 뉴컴 문제는 EDT가 길조를 따라 선택하고 CDT가 효력에 따라 선택한다는 점을 제시하며 EDT의 제안이 비합리적이라는 사실을 밝힌다. 셋째, 노직의 뉴컴 문제는 마술사의 문제로 오해할 소지를 지니는데, 의학의 뉴컴 문제는 인과 관계를 명확하게 밝힘으로써 그런 오해의 가능성을 차단한다. 넷째, 의학의 뉴컴 문제는 인과 관계를 명확하게 밝힘으로써 상관관계라는 길조가 길조로서의 작용을 하지 못하도록 하여 EDT를 너무 허약한 허수아비 이론으로 바꿔 반박한다. 다섯째, 쌍둥이 죄수의 딜레마는 마술사의 문제로 오해할 수 있는 여지를 주지 않으면서 길조의 작용을 할 수 있도록 하여 EDT를 반박함에 있어 가장 설득력이 높은 사례가 된다.

이상의 논의에서 특히 강조하고자 한 점은 EDT는 합리적 결정을 설명하기 위해 인과적 믿음을 명확하게 밝힐 필요가 없다는 입장이지만 인과적 믿음을 지닌다고 할 때에는 관련된 증거를 충분히 지니고 있다는 입장이라는 것이다. 이러한 증거에 비추어 볼 때 의학의 뉴컴 문제가 제시하는 상관 관계의 증거는 EDT에서 미미한 역할을 할 수 밖에 없으며, 그에 따라 EDT가 인과 관계를 무

시하고 상관 관계에 따라 결정을 한다는 해석은 EDT를 잘못 이해하는 것이라고 주장했다. 그럼에도 불구하고 뉴컴 문제가 제시하는 길조와 효력의 구분은 여전히 중요한 지적으로 남는다. EDT가 효력이 아니라 길조를 쫓는다면 그것은 분명히 잘못이기 때문이다. 행위의 기대 효용성을 계산하는 방식에서 EDT는 배경지식의 역할을 구체적으로 제시할 수 있어야 할 텐데 그 한 가지 방법은 길조가 아니라 효력을 쫓도록 하는 CDT처럼 인과적 믿음의 역할을 인정하는 것이다.

참고문헌

이종권(2006), "뉴컴의 역설과 합리적 선택", 『논리연구』 vol. 9, no. 1, pp. 63~95.

Ahmed, A.(2005), "Evidential Decision Theory and Medical Newcomb Problems", *BJPS* 56. pp. 191-198.

Cartwright, N.(1979), "Causal Laws and Effective Strategy", *Nous* 13. pp. 419-437.

Egan, A.(2007), "Some Counterexamples to Causal Decision Theory", *Philosophical Review* 116. pp. 93-114.

Hitchcock, C.(1996), "Causal Decision Theory and Decision-Theoretic Causality", *Nous* 30. pp. 508-526.

Joyce, J.(1999), *The Foundations of Causal Decision Theory*, Cambridge Univ. Press.

Lewis, D.(1986), "Prisoners' Dilemma is a Newcomb Problem", in *Philosophical Papers Vol. II.* pp. 299-305. Oxford Univ. Press.

Nozick, R.(1969), "Newcomb's Problem and Two Principles of Choice" in *Essays in Honor of Carl G. Hempel*, N. Rescher, ed., pp. 114-146. Reidel.

Price, H.(1986), "Against Causal Decision Theory", *Synthese* 67. pp. 195-212.

Price, H.(1991), "Agency and Probabilistic Causality", *BJPS* 42. pp. 157-176.

Sobel, H.(1994), *Taking Chances*, Cambridge Univ. Press.

Weirich, P.(2008), "Causal Decision Theory", *The Stanford*

Encyclopedia of Philosophy (Winter 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2008/entries/decision-causal/>.

동덕여자대학교
Email: yyeo4@hanmail.net

Medical Newcomb Problem and Causal Decision Theory

Yeongseo Yeo

We have many causal beliefs, and they play an important role in our decision making. Unlike evidential decision theory, causal decision theory claims that an account of rational choice must use causal beliefs to identify the considerations that make a choice rational. I claim that evidential decision theory is refuted by the original Newcomb's problem but not by the medical Newcomb problem. The latter is taken to be the best example to point out the weakness of evidential decision theory. However, by the explicit statement about causal relations, I argue that the medical Newcomb problem loses its strength in refuting evidential decision theory. With this argument, this paper clarifies the difference between evidential decision theory and causal decision theory.

[Key Words] Causation, Newcomb's Problem, Medical Newcomb's Problem, Evidential Decision Theory, Causal Decision Theory, Twin Prisoners' Dilemma