논리연구 23-2(2020) pp. 71-115

엄격한 시험과 증거-가설 생성의 메커니즘*

전 영 삼

【국문요약】 어느 가설에 대해 동일한 증거라 할지라도 그 증거가 어떤 방식으로 얻어진 것인가에 따라 분명 해당 가설에 대한 지지의 정도가 다른 것으로 보인다. 이러한 점에서 증거를 얻는 절차에 관한 메이요의 ‘엄격한 시험’ 개념과 그에 대한 기법의 개발은 주목할 만하다. 그럼에도 불구하고 그에 대한 비판들이 여러 측면에서 제기되었는데, 여기서는 그 가운데 특별히 정동욱 (2018)과 Iseda (1999)에서 제기된 비판에 초점을 맞춰, 메이요를 대신해 그에 답해 보기로 한다. 이를 위해 본 논문에서는 특히 ‘증거-가설의 메커니즘’이라는 새로운 개념이 제안된다. 이 과정에서 또한 메이요 자신의 잘못에 대해서도 지적하게 될 것이다.

【주요어】 엄격한 시험, 증거-가설 생성의 메커니즘, 메이요, 정동욱, 이세다

투고일: 2020.05.16. 심사 및 수정완료일: 2020.06.22. 게재확정일: 2020.06.24.

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A5B5A07084835). 문제를 좀더 넓은 관점에서 지적해 준 심사위원들께 감사드리며, 필요한 만큼 각주로 답해 보았다. 그럼에도 불구하고 여전히 남아 있을 수 있는 불명료함에 관해서는 여기서보다는 또 다른 맥락에서 한층 더 잘 답할 수 있으리라 보고 후일을 기약하기로 한다.

과학의 근본 목표 중 하나는 ‘과학적 지식’의 확대이다. 그리고 그러한 지식의 확대에 ‘가설’의 역할이 중요함은 분명하다. 문제는 그 가설을 얼마나 믿을 수 있느냐 하는 점이다. 물론 경험 과학에서 그 가설에 대한 믿음은 궁극적으로 경험적 증거에 기반을 둘 수밖에 없다. 이러한 관점에서, 어떤 과학적 가설이 믿을 만한 것이기 위해서는 무엇보다 주어진 경험적 증거와 부합되어야 함은 쉽사리 짐작할 수 있다. 하지만 해당 가설이 주어진 경험과 부합된다 할지라도, 그에 대한 믿음의 정도에 여전히 차이가 있을 수 있다. 그러한 차이가 발생하는 중요한 한 가지는 해당 가설에 대해 문제의 증거가 과연 어떤 절차에 따라 얻어졌느냐 하는 점이다.

이를 위해 사소하긴 하나 관련성 깊은 다음 예를 보기로 하자. 예컨대 눈으로 보아서는 확인하기 어려운, 어느 사람의 키에 관해 ‘그의 키가 (이전에 비해) 컸다’라는 가설을 세운다고 해 보자. 이 가설이 사실에 부합되는지의 여부를 알 수 있는 가장 자연스러운 방법은 자를 이용해 그의 키를 직접 재 보는 일일 것이다. 그런데 이처럼 자를 이용해 그의 키를 재 본 결과, 이전에 비해 키가 크지 않은 것으로 나타났다고 해 보자. 그렇다면 이러한 증거에 의해 우리는 곧 앞서의 가설을 기각하고 대신 ‘그의 키는 크지 않았다’는 상반된 가설을 믿어야만 하는 것일까? 이에 대한 답은 단순히 그 주어진 증거를 넘어 그와 같은 증거가 어떠한 ‘절차’에 의해 얻어진 것인가에 따라 달라질 수 있다. 예컨대 문제의 자가 2mm 단위의 눈금을 가진 경우와 1mm 단위의 눈금을 가진 경우, 해당 가설에 대해 문제의 증거가 보여 주는 바는 동일하다 할지라도, 그 증거가 어느 자에 의해 얻은 것인가에 따라 그 믿음의 정도가 달라질 수 있는 것이다. 이때 전자보다는 후자의 경우에 해당 가설에 대한 믿음의 정도가 더 높아 보임은 물론이다. 왜냐하면 2mm 눈금의 자로 재어 키가 크지 않았다는 증거를 얻었을지라도 1mm 눈

금의 자료는 키가 큰 결과로 나타날 수도 있기 때문이다. 직관적으로 후자야말로 전자에 비해 그 가설을 잘못 받아들일 가능성이 훨씬 더 적은 절차인 것이다.

이처럼 어떤 가설을 잘못 받아들일 가능성이 적은 증거를 얻는, 일종의 테스트 절차를 흔히 ‘엄격한 시험’(severe test)이라 부르며, 그 개념을 잘 정립한 인물이 누구보다 바로 메이요(Deborah Mayo)이다. 이와 관련한 그녀의 주장들은 그 자체 매우 분명할 뿐만 아니라 그를 뒷받침하는 기법들도 나름대로 잘 개발되어 상당히 주목을 받고 있는 형편이다. 그럼에도 불구하고 물론 그것들이 가설과 증거 사이의 관계에 관한 여러 다른 접근 방식들로부터 비판을 받아 오지 않은 것은 아니다. 따라서 그와 관련해 논란이 계속되고 있긴 하나, 다행히 그 중 많은 부분에 관해 메이요 자신이 나름의 답을 제시해 왔고, 그럼으로써 서로간의 입장이 분명해지는 한편 그 장단점도 많이 드러났다고 생각한다. 다만 그녀의 엄격한 시험 방법에 대한 규정이 안고 있는 어떤 문제나 그로부터 파생되는 결과들에 대한 비판으로서 아직까지 그녀로부터 명료한 답을 얻고 있지 못한 경우들이 남아 있다. 그 가운데 이 글에서 주목하고자 하는 바는 바로 정동욱 (2018)과 Iseda (1999)의 비판이다. 이 두 비판은 그 동기와 과정은 다소 상이할지 몰라도 그 주요한 비판점에서는 공통점이 있다. 바로 메이요식의 시험에서 그 엄격성 판정에 비일관성이 존재한다는 것이다. 이 비판은 적어도 나의 생각으로는 매우 치명적인 비판 중 하나로 보인다. 하지만 매우 이상하게도, 내가 아는 하는 한, 그것은 적어도 메이요로부터 직접적인 대응을 받아 본 적이 없다.[1)]

이러한 상황에서 본 논문에서는 바로 그와 같은 비판들에 대해

---

1) 예컨대 자신에 대한 여러 비판들에 응수・해명하며 그 동안의 성과들을 잘 정리하고 있는, 그녀의 최근 저술의 하나인 Mayo (2018)에서만 하더라도 정동욱과 이세다식의 비판에 대한 직접적인 해명을 찾기란 어렵다.

나 자신의 새로운 개념을 제안하고, 그에 의해 그녀를 대신해 그러한 비판들에 대한 답을 시도해 보는 데 목표가 있다. 이때 내가 새로이 제시하고자 하는 개념은 바로 '증거-가설 생성의 메커니즘' 개념이다. 개략적으로 말해, 이는 어느 증거와 가설을 서로 연관된 방식으로 생성해 준다고 여겨지는 확률적 메커니즘을 뜻한다. 어쩌면 그 동안, 나 자신은 미처 발견하지 못했을지라도, 혹 메이요 자신이 이미 간접적으로(아니면 만에 하나 직접적으로) 그들의 비판에 답을 했을지도 모른다. 하지만 이 경우일지라도 위와 같은 비판에 답하는 방식은 나의 새로운 제안에 의해 여전히 새로울 수 있다. 게다가 이러한 과정에서 메이요 자신의 잘못에 대해서도 지적하게 될 것이다.

이러한 대응이 단순히 '메이요'라는 한 연구자를 위한 옹호에 그치는 것이 아님은 물론이다. 단지 과학적 가설에 대한 테스트라는 매우 중차대한 과학 일반의 문제에 대한 좀더 믿을 만한 근거를 확보하려는 노력의 일환일 뿐이다.

1. 메이요의 엄격한 시험

본 논문의 목적을 위해서는 자연히 가장 먼저 메이요가 생각하는 엄격한 시험이 어떠한 것인지부터 소개할 필요가 있다. 다만 이후 이루어지는 그에 대한 비판들을 이해할 수 있는 정도로만 한하기로 한다.

이를 위해 우선 앞서 살펴본 키 측정의 사례로 되돌아가 보기로 하자. 이 사례에서 어떤 사람의 키를 재는 과정이 하나의 시험 절차라면, 그 절차가 해당 가설을 잘못 통과시키는 일이 적으면 적을수록 그것은 더욱 '엄격'해 보일 것이다. 문제는 이와 같은 직관을 어떻게 엄밀하게 정식화할 수 있느냐 하는 것이다. 이와 관련해 메

이요는 여러 버전을 제시한 바 있으나, 우리의 목적을 위해서는 다음으로 충분하다.[2)]

(ST) 가설 h는 임의의 증거 e로써 엄격한 시험 T를 통과하였다 ⇔
(i) e가 h와 부합된다. 그리고
(ii) 만일 h가 거짓이라면, 시험 T는 e보다 h에 더 잘 부합되지 않는 결과를 (매우 높은 확률로) 산출했을 법하다.

우선 여기서 'e가 h와 부합된다'(agree with or fit)는 말은 문제의 증거가 해당 가설과 충돌하지 않는다는 의미이다. 만일 가설 h가 옳은 가설로서 엄격한 시험 T를 통과했다면, 이는 일차적으로 분명한 요건이다. 예컨대 위의 키 측정 사례에서 해당 가설이 1mm 자의 시험을 통과했다면, 그때의 증거는 무엇보다 그 가설에 반하는 결과여서는 안 될 것이기 때문이다. 다만 그것의 정확한 의미를 어떻게 밝히는가는 문제가 될 수 있는데, 이에 관해 메이요는 관련 각주에서 적어도 다음과 같이 규정할 수 있는 것으로 본다. 즉 증거 e가 주어졌을 때 가설 h의 우도(가능도, likelihood)가 그 가설 아닌 여타 가설의 그것보다 더 커야 한다는 것이다. 그리고 이를 $P(e;h)>P(e;\text{not-}h)$와 같이 나타내었다.[3)] 이는, 만일 가설 h가 옳

---

2) Mayo (2005), p. 99. 여기서는 지금의 맥락상 불필요한 부분을 생략하고, 우리의 용어와 기호법에 맞게 고쳐 인용하였다. 메이요의 경우, 이러한 용어 사용과 기호법은 이하 마찬가지이다.

3) Ibid., n.3. e가 주어졌을 때 가설 h의 우도로서 흔히는(특히 베이즈주의에서는) '$P(e;h)$' 대신 '$P(e/h)$'와 같이 표기하곤 한다. 하지만 이는 가설에 의한 조건부 확률 개념에 따른 것으로, 가설의 사전 확률에 반대하는 메이요는 후자 대신 전자의 표기 방식을 강조하고 있다(p. 102). 만일 문제의 우도를 가설에 의한 조건부 확률로 이해한다면, 그 확률을 구하기 위해 필경 가설의 사전 확률을 필요로 하기 때문이다(하지만 메이요가 가설에 의한 조건부 확률에 반대할 뿐 조건부 확률 자체에 반대하는 것은 아님에 주의!—이에 관해서는 Mayo 2010, p. 192 참조). 그리고 이때의 확률 의미는 요건 (i)과 관련해서는

다면, 증거 e가 여타 가설에서보다는 바로 가설 h가 가정하는 확률 분포 하에서 더 나타날 법하다는 관점에서 쉽사리 이해 가능하다.

하지만 이제 엄격한 시험의 관점에서 좀더 중요한 요건은 바로 위의 요건 (ii)이다. 왜냐하면 엄격한 시험에서 좀더 본질적인 것은 우리가 문제의 가설을 잘못 받아들일 가능성이 적어야 한다는 점이기 때문이다. 사실상 위의 조건 (i)의 경우에는 엄격성이 떨어지는 시험에서도 충족될 수 있는데, 예컨대 키 측정의 사례에서 어느 사람의 키가 사실상 3mm 정도 커진 결과를 얻는 경우, 해당 증거는 엄격성이 상대적으로 떨어지는 2mm 자의 시험에서도 문제의 가설과 잘 부합될 수 있기 때문이다. 그러나 위의 요건 (ii)의 정확한 의미란 무엇인가?

키 측정의 사례를 보았을 때, 엄격한 시험의 요체는 그 시험의 절차가 문제의 가설을 잘못 받아들일 가능성이 가능한 한 적은 증거를 산출해야만 한다는 것이다. 그렇다면 먼저 가설을 '잘못 받아들인다'는 것은 무슨 의미인가? 그것은 문제의 가설이 사실은 거짓인데도 그것을 잘못 참인 것으로 여겼음을 의미한다. 그렇다면 엄격한 시험일수록 그 시험에서 산출되는 증거는 그러한 사실을 제대로 밝혀 줄 수 있어야만 할 것이다. 하지만 어떻게 그러할 수 있는가? 이에 대한 답이 바로 위의 요건 (ii)에 제시돼 있다. 만일 어떤 가설이 사실상 거짓이라면, 문제의 엄격한 시험에서 산출될 결과들은 가능한 한 해당 가설과 잘 부합되지 않는 것이어야만 한다. 그

여러 가지로 열려 있으나, 특히 요건 (ii)와 관련해서는 상대 빈도 또는 그것의 극한으로서의 확률로 이해할 필요가 있다. 또한 'not-h'라 할 때, 이는 베이즈주의에서 말하는 '잔여 총괄 가설'(catch-all hypothesis)을 의미하기보다 h와 대립되는 가설, 예컨대 통계학에서 말하는 '귀무 가설'(null hypothesis)을 뜻한다(Mayo 2014, pp. 81-82 참조). 이상의 점들은 이하의 논의들을 통해 좀더 분명해질 것이다.

러므로 문제의 시험에서 산출된 증거는 그러한 결과 중 하나로, 메이요는 그와 같은 결과들은 적어도 실제 산출된 증거 e보다는 가설 h와 더 잘 부합되지 않을 것이라 보고 있다. 달리 말해, 그 결과들은 적어도 실제 산출된 증거 e만큼 가설 h와 부합되지는 않을 것이라 보고 있다. 예컨대 키 측정의 사례에서 1mm 눈금의 자로 재어 얻은 증거 e가 3mm 정도의 키 차이를 보여 주는 한, 그것은 실상 어떤 사람의 키가 컸다는 가설 h와 잘 부합된다. 이는 그 증거가 이전 키와의 차이에서 1mm보다 더 큰 어떤 측정 결과를 보여 주었음을 의미한다. 하지만 만일 그 가설 h가 거짓이라면, 그러한 측정 결과들은 실제 주어진 증거 e보다 가설 h와 더 잘 부합되지 않는 결과, 즉 이전 키와의 차이에서 1mm보다 작은 결과들을 산출했을 가능성이 크다.

이러한 원리에 따라 메이요는 어떤 시험의 엄격성 정도를 다음과 같이 수치화하는 일도 가능하였는데,[4] 이 또한 위의 논의를 통해 쉽사리 이해할 수 있을 법하다.

(SEV) 가설 h가 거짓일 경우 시험 T가 h를 탈락시킬 확률
= 1 — (가설 h가 거짓일 경우 시험 T가 h를 통과시킬 확률)

이처럼 어떤 시험 T의 엄격성 정도를 수치화하기 위해 사용되는 확률은 문제의 시험을 동일하게 계속 적용하였을 때 나타나는 결과들의 상대 빈도로서의 확률을 뜻한다. 그러므로 이러한 의미의 확률을 결정해 주는 확률 함수를 여전히 P로 두고 앞서와 같은 메이요식의 표기법에 따르면, 위의 (SEV)는 1—P(시험 T가 가설 h를 통과시킴; h가 거짓)와 같은 식으로 표현 가능하다. 하지만 이러한 식을 구체적으로 어떻게 적용할 수 있을 것인가? 메이요는 이러한

[4] Mayo (1996), pp. 180-181.

식이 구체적으로 적용될 수 있는 가장 전형적인 경우를 네이먼-피어슨 통계학에서 발견할 수 있다고 본다. 따라서 앞서 키 측정의 사례를 이와 같은 통계학의 틀 내에서 예시해 보면 다음과 같다.

이제 어떤 사람의 키를 측정하는 일이 완전히 정확치는 않아, 앞서와 같은 1mm 눈금의 자로 재는 시행을 100번 행하고, 그 가운데 60%가 이전 키와의 차이가 1mm보다 큰 것으로 나타났다고 해 보자. 그렇다면 이러한 증거 e를 바탕으로, 우리는 방금과 같은 100번의 시행 결과 그 가운데 적어도 60%가 이전 키와의 차이가 1mm보다 큰 결과를 보인 경우, '어떤 사람의 키가 컸다'는 가설 h를 통과시키는 절차를 우리의 시험 T로 삼을 수 있을 것이다. 이때 만일 시험 T가 엄격하다면, 문제는 그와 같은 결과가 단지 우연에 의해 나타나기란 어렵다는 점을 보여 주는 일이다. 하지만 이를 어떻게 보여 줄 수 있을 것인가? 이에 대해 메이요는, 만일 가설 h가 거짓일 경우, 그리하여 문제의 결과가 단지 우연에 의해 나타났다고 볼 경우, T가 h를 통과시킬 확률을 구함으로써 그에 답할 수 있다고 본다.

이를 위해 우선 문제의 결과가 단지 우연에 의해 나타난 것이라는 가설을 h_0라 해 보자. 이는 곧 우리가 받아들이고 싶어 하는 가설 h와 대립되는 가설, 그리하여 우리가 기각하고 싶어 하는 가설로서의 귀무 가설에 해당한다. 그런데 만일 어떤 사람의 키 성장에 관해 아무런 측정 없이 단지 '그 사람의 키가 크기 않았다'고 주장한다면, 그것이 완전히 임의로 이루어지는 한, 그 결과가 1mm를 넘지 않을 비율은 0.5, 그리하여 또한 그 결과가 1mm를 넘을 비율은 p=0.5로 간주할 수 있다. 즉 가설 h_0는 h_0: p=0.5를 주장하는 것으로 설정 가능하다. 반면 가설 h는 h: p>0.5를 주장하는 셈이다.

그렇다면 이제 가설 h가 거짓인 경우, 곧 h_0가 참일 경우, T가

h를 통과시킬 확률은 다음처럼 구할 수 있다. 먼저 시험 T를 행할 때, 그 결과가 보여 주는 확률 분포는 p=0.5의 확률로 1mm를 넘을 사건들의 계열이 보여 주는 분포이다. 이러한 분포는 그 사건들을 낳는 n번의 시행에서 평균이 np, 분산이 npq=$np(1-p)$인 이른바 '이항 분포' $B(np, npq)$를 따른다. 우리의 예에서라면 그 평균은 100(0.5)=50, 분산은 100(0.5)(0.5)=25이다. 하지만 이 경우 n=100이나 되므로, 그 분포는 정규 분포 $N(np, npq)$에 근사할 수 있다. 따라서 이에 준해 그를 다시 표준화된 정규 분포인 $N(0, 1)$으로 변환할 수 있는데, 이 표준 정규 분포상에서 어떤 확률 변수 X에 관해 우리가 직접 관찰한 결과값 X=60이 그 평균 50으로부터 얼마나 떨어져 있는가는 변환식 Z=(X-np)/$\sqrt{npq}$로 (60-50)/$\sqrt{25}$=2와 같이 구할 수 있다. 이는 가설 h_0에 따라 가정된 상대 빈도와 우리가 실제 관찰해 얻은 상대 빈도, 즉 증거 e 사이의 거리가 표준 정규 분포상에서 2만큼 떨어져 있음을 의미한다. 이를 '표준 편차 2'(2sd)라 부른다.

그렇다면 이러한 분포상에서 우리가 실제 관찰해 얻은 증거 e보다 평균으로부터 더 멀리 떨어져 있는 결과 모두를 얻을 확률을 얼마인가? 달리 말해 해당 분포상에서 우리가 적어도 증거 e를 얻을 확률은 얼마인가? 이 확률은 물론 미리 마련된 표준 정규 분포표를 통해 쉽사리 얻을 수 있다. 하지만 이때 흔히 표준 편차로서 2에 가장 가까운 값으로 널리 알려져 있는 것이 바로 1.96이며, 이때 문제의 확률이 0.025임도 잘 알려져 있다. 따라서 앞으로 논의의 편의상 표준 편차 2를 즐겨 이용하고, 이때 문제의 확률 역시 0.025로 정해 두기로 하자. 그렇다면 이야말로 바로 지금의 예에서 P(시험 T가 가설 h를 통과시킴; 가설 h가 거짓)에 해당하는 확률이다. 이로써 우리는 최종적으로 (SEV)에 따라 시험 T의 엄격성의 정도를 1-0.025≒0.97로 구할 수 있게 된 것이다.

위의 확률 0.025는 가설 h_0 하에서는 우리가 얻은 증거 e와 같은 결과를 얻기가 매우 어렵다는 것을 뜻한다. 반면 엄격성의 정도 0.97은 매번 크기 100의 표본으로 T와 같은 시험을 계속 반복할 경우 가설 h_0 하에서는 그 97%가 e에 나타난 60%보다 더 작은 퍼센트의 결과를 보일 것임을 의미한다. 이로써 예의 시험 T는 실천적인 맥락에서 상당히 엄격한 것이라 하기에 충분하다.[5]

이상으로 엄격한 시험에 관한 메이요의 개념은 직관적으로 호소력이 있으며, 그를 수치화하려는 시도 역시 기존의 통계학을 배경으로 매우 탄탄한 기반을 갖고 있는 것으로 보인다. 그럼에도 불구하고 이에 대한 여러 비판이 계속되어 왔는데, 그러한 비판들 가운데 여기서는 앞서 언급한 대로 정동욱과 이세다의 비판에 초점을 맞추기로 한다. 단순히 시기적으로는 이세다의 비판이 정동욱의 그것에 앞서기는 하나, 나의 판단으로는 정동욱의 비판에 먼저 답하는 것이 이세다의 비판에도 답하기 훨씬 용이하다. 그러므로 다음 절에서 먼저 정동욱의 비판으로부터 시작해 보기로 하자.

2. 메이요에 대한 정동욱의 비판

메이요에 대한 정동욱의 비판 역시 여러 갈래이다. 하지만 그 핵심은 메이요의 엄격성 판정이 비일관적이라는 점이다.

이를 위해 앞서의 키 측정 사례에서와는 조금 다르게 다음의 사례를 생각해 보기로 하자. 우선 어느 측정이 각각 독립적으로 동일

---

5) 우리의 키 측정 사례에서라면 눈금의 단위가 클수록 키가 큰 결과가 나타날 비율이 더 작을 것으로 예상된다. 예컨대 눈금의 단위가 작은 경우에 비해 그 비율이 0.6보다 더 작은 0.52와 같은 수치를 보일 수 있다. 이 경우 X=52로, 그 Z-값은 이전보다 더 작아지게 된다. 이렇게 되면, 방금의 논의대로, 눈금이 큰 자로 측정할 때 그 시험의 엄격성이 눈금이 작은 자로 측정할 때에 비해 더 작아짐도 알 수 있다.

한 정규 분포 $N(\mu,\sigma^2)$에 따라 이루어지는 하나의 모집단을 생각해 보기로 하자. 이러한 방식으로 모두 n번의 측정이 이루어진다면, 우리는 그러한 분포를 따르는 확률 변수 $X=(X_1,\ldots,X_n)$를 고려할 수 있을 것이다. 그런데 이때 그 분산 σ^2은 미리 알려져 있으나, 그 평균 μ는 알려져 있지 않다고 가정해 보자. 그리하여 우리는 지금의 분포를 보이는 모집단의 평균 μ를 추정해 보고자 한다. 이를 추정할 수 있는 가장 자연스러운 방법은 물론 그러한 모집단으로부터 표본을 추출하여 그로부터 문제의 평균을 추정하는 방법이다. 하지만 어떻게 그럴 수 있는가?

이를 위해 네이먼-피어슨 통계학에서는 앞서 키 측정의 사례에서와 유사하게 다음과 같이 나아가고 있다. 먼저 키 측정의 사례에서처럼 n=100의 표본을 해당 모집단으로부터 추출하고, 그 표본의 평균 $\overline{x}$를 고려해 보자. 이 경우 그처럼 n=100의 표본들이 이루는 표본들의 평균, 즉 표본 평균 $\overline{x}$ 자체가 이루는 분포를 가정할 수 있는데, 그 분포상의 표본 표준 편차 sd는 $\sigma/\sqrt{n}$이 된다는 것이 잘 알려져 있다. 그렇다면 이러한 상황에서 우리는 가설 $h(x)$: ($\overline{x}$-2sd$\leq\mu<\overline{x}$+2sd)를 설정할 수 있고, 이 가설의 확률은 다음과 같이 주어진다.

(2.1) $P(\overline{x}$-2sd$\leq\mu<\overline{x}$+2sd)=0.95

이 결과는 사실상 앞서 키 측정 사례에 관한 논의로부터 쉽사리 이해할 수 있는데, 앞서 표준 정규 분포상에서 그 평균으로부터 양의 방향으로(즉 그 분포 그래프상으로는 오른쪽으로) 2sd 이상 떨어져 있는 측정 결과를 얻을 확률이 0.025, 그리고 그와 대칭적으로 유사하게 음의 방향으로(즉 그래프상으로 왼쪽으로) 그러한 결과를 얻을 확률이 역시 0.025이므로, 1-2(0.025)=0.95로 위의 값을

얻을 수 있다.

그러므로 메이요는 지금처럼 어떤 증거로부터 통계적인 신뢰 구간의 가설을 구성하는 경우에도 앞서의 키 측정 사례에서와 마찬가지 원리로 해당 시험의 엄격성을 검토하고 그 정도를 제시할 수 있다고 주장한다.[6] 우선 이 경우의 시험 절차란 방금 설명한 방식에 따라 신뢰 구간의 가설을 구성하고 시험하는 과정 자체를 말한다. 좀더 정확히 말해, 어떤 자료 x가 주어졌을 때, 그 자료와 부합되게 가설 $h(x)$를 구성하거나 선택하고 동시에 그 자료로써 문제의 가설을 시험하는 절차를 말한다. 이제 이 시험 절차를 R(X)라 해 보자.[7] 그렇다면 이 경우에도 다음과 같은 확률을 구하는 일이 가능하고,

(2.2) P(R(X)가 $h(x)$를 산출함; $h(x)$가 거짓)=0.05

이로써 R(X)의 엄격성 정도는 1-0.05=0.95가 된다.

물론 이때 여기서 말하는 자료 x가 앞서 우리가 말한 증거 e에 해당한다. 논의를 더 구체화하기 위해, 실제 위와 같은 절차에 따라 n개의 측정 결과가 x=($x_1,\ldots, x_n$)으로 주어지고, 이때의 평균이 x_0라고 해 보자. 그렇다면 이에 대응해 우리의 구체적인 실제 증거를 e_0라 할 수 있다. 그렇다면 위의 식 (2.2)의 좌변은 이 증거에 맞게 다음과 같이 재구성 가능하다.

---

6) Mayo (2014), pp. 84-85.

7) 이처럼 이미 주어진 자료를 사용하여 가설을 구성하고 동시에 그 가설을 시험하는 데 동일한 자료를 이용할 경우, 과연 그 자료가 해당 가설을 온전하게 지지할 수 있는가에 관해 많은 논란이 있다. 이른바 ‘예측주의’(predictivism) 대 반예측주의 사이의 논란이다. 하지만 이는 본 논문의 주제에서 벗어나므로, 여기서는 더 이상 논하지 않기로 한다.

(2.3) P(R(X)가 $h(e_0)$를 산출함; $h(e_0)$가 거짓)

그렇다면 이때의 확률값 역시 0.05일까? 그렇지는 않다. 위의 식 (2.2)에 주어진 0.05의 확률값은 표준 정규 분포상에서 양과 음의 양 방향으로 각각 2sd 이상 및 이하의 결과가 나타날 모든 확률을 뜻한다. 따라서 그 결과 중 어느 특정한 하나에 대한 식인 위의 (2.3)의 확률값은 0.05 이하일 수밖에 없다. 따라서 이 경우의 엄격성 정도 역시 0.95 이상이게 마련이다.

이러한 결과를 두고 정동욱이 그 비판을 위해 전형으로 삼는 또 다른 사례가 있다. 원래 이 사례는 메이요 자신이 여러 곳에서 그 엄격성의 정도가 매우 낮은 사례의 하나로 제시한 것이나,[8] 아이러니컬하게도 그 비판자들에게는 메이요를 비판하는 사례로 전용(轉用)되고 있는 사례이다.

이제 동전 하나를 n번에 걸쳐 던지는 실험을 행한다고 해 보자. 이러한 실험의 결과를 두고 우리가 흔히 세울 수 있는 귀무 가설 h_0는 '그 동전은 (치우침 없이) 공정하다'는 것이다. 통계학적으로는 각 시행에서 그 동전의 앞면이 나올 확률이 p=0.5라는 주장이다. 그런데 이제 이러한 가설 대신, 이미 나타난 결과에 완전히 부합되게 다음과 같은 방식으로 새로운 가설 하나를 세워 본다고 해 보자. 즉 이미 앞면이 나온 결과에 대해서는 p=1이고 이미 뒷면이 나온 결과에 대해서는 p=0을 주장하는 가설을 말한다. 그러므로 이때의 실험 결과를 증거 e로 둔다면, 이로써 구성된 가설은 $G(e)$로 둘 수 있다. 여기서 주어진 기호 'G'는 매우 상징적인데, 이는 이미 초능력자로 널리 알려진 유리 겔러(Uri Geller)의 이름에서 따온 것으로, 그 초능력의 비밀이 결국엔 이처럼 사후 예언의 방식에 놓여 있는 것으로 여겨졌기 때문이다. 따라서 메이요는 이와 같은

8) 예컨대 Mayo (1996), pp. 201-203.

가설을 “겔러식 가설”(gellerized hypothesis)이라 불렀다. 그렇다면 이러한 가설에 대한 시험의 엄격성은 어떻게 될까?

이를 위해 메이요는 우선 시험 절차 R_G로서 ‘바로 증거 e를 기반으로, 가설 h_0는 탈락시키고 최대로 우도가 큰(maximally likely) 가설 $G(e)$는 통과시킨다’는 절차를 가정한다. 그렇다면 이 시험의 엄격성을 계산하기 위해 먼저 가설 $G(e)$가 거짓일 경우 시험 R_G가 $G(e)$를 통과시킬 확률을 구할 필요가 있다. 그런데 메이요의 생각에 “$G(e)$가 거짓이고 h_0가 참일지라도(즉 문제의 동전이 ‘공정하다’고 할지라도) 그 어떠한 실험 결과이든 그 모두가 귀무 가설 h_0는 탈락시키지 못하고 [증거] e에 부합되게 구성된 가설 $G(e)$는 통과시키게 되어 있다.” 그러므로 그녀는 “[이때] $G(e)$를 잘못 통과시킬 확률은 최대가[즉 1이] 되고[…] 그 시험의 엄격성은 최소가[즉 0이] 된다”[9])고 보았다.

메이요의 이러한 추론 과정은 직관적으로 매우 그럴 법해 보인다. 그럼에도 불구하고 정동욱은 이 과정을 앞서 고려했던 신뢰 구간 가설에 대한 추론 과정과 비교해 보면 치명적인 비일관성이 드러난다고 비판한다. 어떤 점에서 그러한 것일까?

우선 문제의 두 과정에는 두 가지 공통점이 존재한다. 첫째, 그 과정에서 구성되는 가설 모두 이미 해당 증거가 주어진 이후에 그에 부합되게 구성되었다는 점이다. 즉 가설 $h(x)$와 $G(e)$ 모두 그에 해당하는 증거 x와 e가 미리 주어진 이후 그에 따라 구성된 가설이라는 점이다. 둘째, 두 과정 모두의 시험에서 그 엄격성을 따지는 경우, 문제의 가설이 거짓이라는 가정 아래, 그리하여 그와 대립되는 가설이 참이라는 가정 하에 해당 시험이 그러한 가설을 산출하거나 통과시키는 확률을 계산하고 있다는 점이다. 하지만 이와 같은 공통점들의 존재에도 불구하고, 적어도 정동욱이 보기에,

[9]) Ibid., p. 202.

여기에는 중요한 차이점이 존재한다.

그는 우선 다음과 같이 추론해 나아간다.[10)]

> [⋯ 시험 R(X)의 엄격성을 이해함에 있어] 우리는 자료가 아니라 **자료 생성 방식**이 고정된 것으로 보아야 한다. 즉 $h(x)$가 거짓인데도 주어진 자료 생성 방식에 의해 $h(x)$로 추론되는 자료가 만들어질 확률을 따져야 하는 것이다. [⋯] 물론 이번의 표본 자료 x_0에 대해 $h(x_0)$를 산출했듯이, [시험] R(X)는 새로운 표본 자료를 구하더라도 새 표본 자료 x_1에 대해 $h(x_1)$을 산출할 것이다. [⋯] 따라서 그 $h(x_1)$이 거짓임에도 불구하고 [시험] R(X)가 $h(x_1)$을 산출할 확률은 역시 1인 것처럼 보인다. 그렇다면 어떻게 R(X)가 $h(x_1)$을 산출할 확률이 1보다 작을 수 있는가? 그 계산은 바로 이번에 생성된 특정한 표본 x_0를 통해 만들어진 신뢰구간 가설 $h(x_0)$가 만약 거짓이라면 새로 표집된 표본 x_1은 $h(x_0)$를 산출하지 않을 가능성이 높다는 임의 표집 방식에 대한 수학적 분석에 의존한다. (강조는 원문)

이어 이상의 상황을 겔러식 가설의 구성에 관한 시험 R_G와 비교해 다음과 같이 지적한다.

> 그러나 엄격한 시험을 신뢰구간 사례에 적용한 위의 방식을 동전 던지기 사례에 일관되게 적용하면, 앞서 겔러식 가설로 분석된 $G(e)$ 역시도 엄격한 시험을 통과할 수 있다. [시험 R_G에서] 다음과 같은 특정한 자료 e_0=(h,t,t,h)를 얻은 경우, $G(e)$는 첫 번째와 네 번째는 앞면이 나올 확률이 1이었고, 두 번째와 세 번째는 앞면이 나올 확률이 0이었다고 말한다. $G(e_0)$가 거짓이고 만약 동전이 앞면이 나올 확률이 매번 0.5라는 대립 가설 h_0가 참이라면, 똑같은 자료 생성 방식에 의해 자료를 새로 구할 경우,

10) 정동욱 (2018), pp. 37-39. 단 그의 용어와 기호들은 우리의 용어와 기호법에 맞게 바꾸었으며, 정동욱의 경우 이하 마찬가지이다.

> 새롭게 구한 자료 e_1은 $G(e_0)$를 산출하지 않을 가능성이 15/16으로 매우 높다.11) […] 그렇다면 R_G도 엄격한 가설 구성 절차로 보아야 하는 것이 아닌가? (각주는 필자)

이로써 그는 두 경우의 비일관성을 다음과 같이 비판한다.

> 만약 R_G가 어떠한 자료 e_k에 대해서도 무조건 $G(e_k)$를 산출한다는 이유로 R_G의 엄격성이 0이라고 말한다면, 신뢰구간 가설 구성 절차인 R(X) 역시 똑같은 처지에 놓이게 된다. R(X)가 엄격한 시험을 구성한다면, 동전 던지기 결과로부터 가설을 구성하는 절차 R_G 역시 엄격한 시험을 구성하게 [되는 것 아닌가!]

게다가 그의 비판은 여기서 그치지 않는다. 만일 사정이 위와 같다면, 그로부터 좀더 확대돼, 어느 가설이든 그에 대립되는 가설에 의해 원가설의 증거가 무력화될 수도 있다는 매우 놀라운 비판을 가하기도 한다.12)

그는 먼저 “메이요에 따르면, 시험의 엄격성에 대한 판단은 항상 특정한 대립 가설에 비추어 이루어진다”고 지적한다. 그런데 “메이요가 언급한 겔러식 가설을 대립 가설로 고려하면 시험의 엄격성은 어떻게 될까?”라고 의문을 제기한다. 이를 위해 정동욱은 커피 맛 구별과 관련된 전통적인 예를 들고 있으나, 여기서는 그를 앞서 동전 던지기의 예로 바꿔 생각해 보기로 하자.

동전 던지기의 사례에서 우리가 어떤 동전을 100번 던져 본 결과 그 가운데 70번 앞면이 나왔다고 해 보자. 그리하여 그때의 특정한 증거가 e_0=(h,t,h,t,h,h,…,h)라고 해 보자. 이 정도 되면 우리는 문제의 동전이 앞면으로 치우침이 있을 것으로 예상하고, 그로써

---

11) 이항 분포상에서 p=0.5인 경우, 특정한 e_0가 나타날 확률은 $(0.5)^2(0.5)^2$ =1/16이고, 이로써 문제의 확률은 1—(1/16)=15/16와 같이 구할 수 있다.
12) 정동욱 (2018), pp. 39-40.

그 동전을 던질 때 앞면이 나타날 확률이 $p>0.5$임을 주장하는 가설 h를 제시할 법하다. 그런데 이제 그 대립 가설로서 **p=0.5을 주장하는 가설 h_0 대신** 바로 겔러식 가설 $G(e_0)$를 고려하고, 이에 대해 앞서 제2절에서 소개한 시험 T와 같은 유형의 시험을 행한다고 해 보자. 이 경우 정동욱은 그 시험의 엄격성 정도에서 매우 이상한 결과가 초래된다고 지적한다. 왜냐하면 앞서 (SEV)에 따라 그 엄격성 정도를 구하면 P(시험 T가 h를 탈락시킴; h가 거짓, 즉 $G(e_0)$가 참)=1—P(시험 T가 h를 통과시킴; $G(e_0)$가 참)=1—1=0이 되기 때문이다. 이에 대해 그는 다음과 같은 해설을 덧붙인다. "[바로 앞면이 나오는] 횟수가 60번이 넘는 자료에 대해서, h 외에 그것을 설명할 수 있는 대립 가설이 있는 것이다. $G(e_0)$가 참일 경우에도, 시험절차 T가 가설 h를 통과시킬 수 있는 확률이 1이기 때문이다."

그리하여 그는 이렇게 주장한다. "[이로써] 표준적인 통계적 시험절차조차 $G(e_0)$와 같은 종류의 대립 가설을 배제함으로써만 엄격한 시험으로 간주될 수 있다는 것에 주목할 필요가 있다. [⋯] '엄격한 시험'을 이용한 증거 개념에서는, 통상적인 관점에서는 가설에 대한 좋은 증거로 간주될 수 있는 자료가 그에 대한 겔러식 대립 가설에 비추어 보는 순간 더 이상 좋은 증거로 간주될 수 없게 될 수 있다는 것이다." 이러한 관점에서 그는 예컨대 "다윈주의 진화론을 뒷받침한다고 여겨지는 상당수의 자료들이 그에 대한 겔러식 대립 가설인 세련된 버전의 개별 창조 가설에 의해서도 설명될 수 있다고 [가정하는 순간], [⋯ 메이요의] 엄격한 시험 개념에 따르면, 그 자료들은 더 이상 다윈주의 진화론의 증거가 될 수 없을 것"이라는 매우 놀라운 주장을 펴고 있기도 한다. 그리하여 그는 이처럼 묻는다. "엄격한 시험 개념은 겔러식 대립 가설에 의해 [이처럼] 증거가 무력화되는 것을 어떻게 막을 수 있겠는가?"

이상의 비판들에도 불구하고, 물론 정동욱은 메이요의 편에서 고려하는 일도 잊지 않는다. 이러한 고려에서 메이요를 위해 그가 가장 유력하게 보는 대응 방식은 다음과 같다.[13)]

우선 첫 번째 비판과 관련하여, 그는 시험 절차 R_G와 R(X)의 차이가 무엇에서 비롯되는가를 묻고, 그 한 가지 가능성을 $G(e_0)$가 동전만에 대한 가설이 아니라 동전과 시점(時點) 모두를 포함한 가설이라는 데에서 찾아볼 수 있다고 말한다. 그리하여 “새로운 자료 생성을 통해 $G(e_0)$를 산출하지 않는 e_1이 새로 나온다고 해도, 그것은 새 시점에 대한 가설 $G(e_1)$을 산출할 뿐, $G(e_0)$에 아무런 영향을 주지 않는다”고 말한다. “즉 동전 가설 $G(e_1)$과 $G(e_0)$는 서로 다르더라도 양립가능하다”고 생각한다. “반면 신뢰구간 가설 $h(e_0)$와 $h(e_1)$은 서로 다를 경우 양립불가능하다”고 보고, “이러한 차이가 두 가설 구성 절차에 대한 차이를 가져왔을 가능성이 있다”고 진단한다.

하지만 정동욱은 이러한 가능성에 대해 곧 다음과 같이 비판한다. “이런 해석 방식에서 $G(e_0)$는 동전 던지기 결과 앞면이 나올 확률이 항상 0.5라는 가설 h_0과도 양립가능할 수 있다. 특정 시점을 고려하지 않았을 때에는 앞면이 나올 확률이 0.5이면서도, 특정 시점의 수많은 변수를 고려하면 그 확률이 0이나 1로 확정될 수 있기 때문이다. 그런데 만약 두 가설이 양립가능하다는 해석을 받아들이는 경우, h_0를 대립 가설로 가정하여 R_G를 엄격하지 않은 겔러화된 시험의 사례로 제시한 메이요의 분석은 의미를 잃게 된다.” 그리하여 그는 다음과 같이 통렬하게 묻고 있다. “이러한 해석 하에서 R_G의 엄격성을 따지기 위해서는 다른 종류의 대립 가설을 가정해야 할 텐데 도대체 어떤 대립 가설이 가능하겠는가?”

두 번째 비판과 관련해서는, 정동욱은 메이요의 편에서 있을 수

13) Ibid., pp. 38-39.

있는 세 가지 선택지를 고려한다.[14] 그러나 그 가운데 처음 두 가지는 곧바로 문제가 있음을 지적하고, 그나마 마지막 선택지가 어느 정도 타당성이 있을 것이라 본다. 바로 위에서 고찰한 방식대로 "$G(e_0)$가 h와 양립불가능한 가설이 아니므로 대립 가설로서 배제해야 한다고 말하는" 선택지이다. 하지만 이 세 번째 선택지를 따른다 할지라도 다른 선택지들과 마찬가지로 공통된 문제를 안게 된다고 그는 지적한다. 곧 "[이로써] 표준적인 통계적 시험 절차조차 $G(e_0)$와 같은 종류의 대립 가설을 배제함으로써만 엄격한 시험으로 간주될 수 있다는 것[…]"이다. 그런데 이는 바로 앞서 우리가 인용했던 한 대목이다. 그러므로 자신의 두 번째 비판과 관련해 메이요의 편에 선다 할지라도 결국엔 원점으로 돌아갈 수밖에 없음을 정동욱은 다시 한번 확인시켜 준 셈이다.

만일 정동욱의 위와 같은 비판들이 옳다면, 이는 언뜻 매우 그럴듯해 보이는 메이요의 추론 과정에 매우 치명적이지 않을 수 없다. 그러나 과연 그러한가?

3. 증거-가설 생성의 메커니즘과 정동욱에 대한 대응

방금의 질문에 나의 답은 '그렇지 않다'이다. 이를 위해 다음과 같은 예로부터 출발해 보기로 하자.[15] 이제 잘 섞인 어느 트럼프 카드 한 벌을 생각해 보기로 하자. 그 한 벌은 총 52장의 카드로 이루어져 있으며, 4가지 '슈트'(suit)'로 구성된다. 각 슈트는 클럽, 다이아몬드, 하트, 스페이드로 나뉘고, 각 슈트는 13장이다. 이 13장은 다시 숫자 A(Ace)로부터 10, 그리고 '그림 카드'[face card]라

---

14) Ibid., p. 40.

15) 이 예는 전영삼 (2018), pp. 22-23에서의 예를 지금의 맥락에 맞게 수정한 것이다.

부르는 J(Jack), Q(Queen), K(King)로 구성된다. 또한 클럽과 스페이드는 검은색, 다이아몬드와 하트는 빨간색으로 되어 있다.

그렇다면 이제 이와 같은 구성을 염두에 두고 그 카드 한 벌로부터 무작위로 어느 한 장의 카드를 뽑아 본다고 해 보자. 이때 그렇게 뽑힌 카드가 스페이드일 것이라는 사건을 주장하는 명제를 a라고 해 보자. 나아가 그것이 스페이드 에이스일 것이라 주장하는 명제를 b라고 해 보자. 이 경우 만일 a가 주어지지 않은 상태에서, (이후 상대 빈도적으로 해석 가능하긴 하나 여기서는 일단 편의상) 고전적인 수학적 확률을 가정할 경우, b의 확률은 $P(b)$=1/52이 될 것이다. 만일 a가 주어진 경우라면, 그 조건 하에서 b의 확률은 조건부 확률 $P(b/a)$=1/13로 주어진다.

그렇다면 앞으로의 논의를 위해 위의 사례로부터 다음과 같은 일반화를 시도해 보기로 하자. 먼저 어느 개체가 일정한 속성을 갖게 되는 사건을 고려해 보자. 위의 사례에서라면, 예컨대 어느 한 카드가 스페이드라는 속성을 갖게 되는 사건을 말한다. 이와 유사하게 그 카드는 4가지 슈트 중의 다른 속성을 가질 수도 있다. 바로 이때 그러한 속성들의 범주를 일정한 변항 내지 변수 X로 두고, 우리는 그 변수의 값으로서 그 구체적인 속성 하나하나를 제시할 수 있다. 아니면 그 속성 하나하나에 대응하는 별도의 수를 제시해도 좋다. 우리의 카드 예에서라면, 슈트를 나타내는 X가 그 값들로서 스페이드, 클럽 등등을 갖는 경우를 상상하면 좋을 것이다. 이제 그 값들 역시 변항 내지 변수 r로 대신한다고 해 보자. 그렇다면 어느 대상이 속성 범주 X 가운데 어떤 값을 갖는가를 '$X=r$'과 같이 나타낼 수 있을 것이다.

그런데 이제 문제의 그 대상이 어떤 속성 범주의 값들을 확정적으로 갖는 대신 단지 불확실하게 확률적으로만 가질 수밖에 없는 상황이라고 해 보자. 우리의 예에서라면, 임의로 뽑힐 어느 한 장

의 카드의 경우가 그러하다. 이러한 경우 통계학에서는 그때의 X를 '확률 변수'(random variable)라 부른다. 그렇다면 여기서 우리는 그 변수의 각 값에 구체적인 확률도 부여하길 원할 수 있다. 물론 가능한 일이다. 그리하여 위의 예에서 우리는 사실상 X=스페이드에 1/52의 확률을 부여한 셈이다. 그렇다면 유사한 방식으로 우리는 X의 다른 값들에 관해서도 그에 적절한 다른 확률을 부여할 수도 있을 것이다. 이처럼 어느 확률 변수의 각 값에 일정한 확률을 부여하여, 그 값마다 어떤 확률을 갖는가를 보여 준 결과를 통계학에서는 '확률 분포'(probability distribution)라 부른다. 그런데 이러한 확률 분포상에서 나타난, 확률 변수의 값과 그에 대응하는 확률 사이의 관계를 일정한 함수로 나타낼 수도 있다. 그러한 함수를 '확률 함수'(probability function)라 부른다. 위의 예에서 제시된 $P(\cdot)$야말로 하나의 확률 함수인 셈이다.

그렇다면 이제 이 대목에서 우리의 카드 예에서 문제의 확률 함수 P가 근원적으로 어디에 연원하는가를 물어 보자. 이렇게 묻고 들어간다면, 결국 그 함수는 트럼프 카드 한 벌의 구성 방식이나 구조에 연원함을 알 수 있다. 그리고 그 구조도 결정적이기보다는 확률적이다. 이처럼 어느 확률 함수를 가능케 하는, 대상들의 근본 구조를 통계학에서는 '확률 모델'(probability model)이라 부른다. 이는 우리의 논의를 위해 대단히 중요한 개념이므로, 좀더 엄밀하게 규정할 필요가 있다.

위의 카드 사례로 볼 때, 확률 모델을 구성하는 핵심 요소는 두 가지임을 알 수 있다. 그 하나는 일정한 확률이 부여될 대상으로서 어떤 사건들이나 그를 주장하는 명제들이다. 이러한 사건들이나 명제들은 더 이상 분해될 수 없는 근원 사건이나 명제들로부터 그것들에 적절한 집합이나 논리 연산을 행해 여러 가지로 다양하게 만들어 낼 수 있다. 이제 이와 같은 연산이 가능한 하나의 집합이

'불 대수 집합'(Boolean algebra)이며, 이를 B라 두어 보자. 다음으로, 확률 모델을 구성하는 또 다른 하나의 핵심 요소는, 집합 B의 각 원소에 대해 확률 공리들에 따라 0으로부터 1까지의 실수를 부여하는 함수, 즉 확률 함수 P이다. 따라서 우리는 이제 어떤 확률 모델을 가장 넓게 일반화해, 그를 집합 B와 함수 P로 이루어진 순서쌍 M=<B, P>로 규정 가능하다.

그런데 흥미롭게도 어떤 확률 모델의 경우에는 그 유형에 있어 상당히 전형성(典型性)이 있어 그 적용 사례가 매우 광범위할 뿐만 아니라 그 확률 함수도 수학적으로 매우 잘 형식화되어 있는 경우가 있다. 이산적(離散的)인 확률 변수에 관해 그러한 모델의 대표적인 것이 바로 이항 분포이고, 연속적인 확률 변수에 관해 그처럼 대표적인 것이 바로 정규 분포이다. 그러므로 이항 분포의 경우, 확률 변수 X에 대해 그 값이 r로 주어질 때, 그 확률 함수는 $P(X=r)={}_nC_r p^r q^{n-r}$로 주어진다. 정규 분포의 경우라면, 확률 변수 X에 대해 그 값이 x로 주어질 때, 그 확률 함수는 $P(X=x)$ $=(1/\sqrt{2\pi}\,\sigma)\exp[(-1/2\sigma^2)(x-\mu)^2]$이다. 이로써 앞서 우리가 제1~2절에서 여러 사례들에 대해 적용한 확률 함수들도 결국엔 이러한 함수들을 특징으로 하는 적절한 확률 모델을 가정한 상태에서 적용 가능했음을 알게 된다.

그렇다면 이 경우 앞서 제1~2절에서 제시했던 가설 h_0나 h(또는 $h(x)$)는 무엇에 대한 가설인가? 그것은 결국 이항 분포나 정규 분포상에서 그 확률 함수의 매개 변수[16] p나 μ에 대한 가설임을 알 수 있다. 그러므로 예컨대 제1절에서 가설 h_0는 이항 분포나 정규 분포상에서 p=0.5를 주장하는 반면, 가설 h 역시 같은 분포상에서

16) '매개 변수'란 원래는 하나의 상수이나 그것을 일반화시켜 변수로 삼은 것을 말한다. 달리 '파라미터'(parameter), 혹은 모집단과 관련된 경우 '모수'(母數)라 칭하기도 한다.

p>0.5를 주장하고 있는 셈이다.[17] 이때 우리가 주목할 점은 바로 그러한 가설들은 **동일한 매개 변수**에 대해 **동일한 확률 분포 유형**을 전제로 서로 대립되는 가설들이라는 점이다. 즉 그 가설들은 동일한 매개 변수에 대해 동일하게 이항 분포면 이항 분포, 정규 분포면 정규 분포상에서 서로 양립 불가능한 가설들이라는 점이다. 나아가 그 가설들을 테스트하기 위해 얻어지는 증거들 역시 그러한 분포상에서만 해당 가설들과 서로 의미 있게 연관될 따름이다. 예컨대 제1절의 키 측정 사례에서 증거 e 역시 동일한 매개 변수에 대해 이항 분포나 정규 분포를 가정한 상태에서만 가설 h_0나 h에 대한 증거 역할을 할 따름이다. 이러한 점들은 우리의 논의를 위해 매우 중요하므로, 이제 이를 잘 반영할 수 있는 새로운 개념 하나를 제시해 보기로 하자.[18]

먼저 앞서와 같이 집합 B를 대상으로, 같은 형식의 확률 함수들을 산출하는 어느 특정한 확률 분포 유형을 D라고 해 보자. 예컨대 $P(X=r)={}_nC_r p^r q^{n-r}$와 같은 형식의 확률 함수는 매개 변수 p 값에 따라 여러 가지로 달리 주어질 수 있으나, 그 모두는 동일하게 '이항 분포'라는 하나의 확률 분포 유형에서 만들어지는 확률 함수들일 따름이다. 이제 이를 이용해 구성할 수 있는 순서쌍 <B, D>를 고려해 보자. 그렇다면 우리는 이와 같은 순서쌍으로 만들어

17) 이러한 관점에서라면, 이제 앞서의 카드 사례에서 한 벌의 카드로부터 무작위로 하나의 카드를 뽑는다고 하는 것 자체가 이미 하나의 가설로서, 그 사례에서 주어진 수학적 확률들 역시 그와 같은 가설이 주어진 상태에서 상대 빈도적 확률로 재해석 가능하다(Mayo 2010, p. 192 참조). 즉 그 카드 한 벌에서 어느 하나의 카드가 무작위로 추출된다고 가정하는 가설을 h라 할 때, 그 카드 나름의 확률 분포(구체적으로는 이른바 '다항 분포')상에서 예로 주어진 확률은 각각 $P(b;h)$, $P(b/a;h)$를 의미한다. 이 경우 가설 h 하에서는 망라적이고 상호 배타적인 사건들에 대해 균등하게 확률이 배분되므로, 그 확률값 자체는 각각 예의 확률과 동일하다.

18) 이 새로운 개념은 원래 Fitelson (2006)으로부터 촉발되어 전영삼 (2018)에서 처음 제시된 것이나, 여기서는 그를 좀더 수정하고 일반화해 보았다.

지는 추상적인 하나의 대상들의 구조를 생각해 볼 수 있을 것이다. 그런데 만일 우리가 이러한 대상들의 구조를 가정한다면, 지금까지 우리가 문제삼아 왔던 여러 증거와 가설들은 서로 아무런 관련 없는 독자적인 것이 아니라 어디까지나 바로 그 구조 내에서 서로 연관된 것들임을 알 수 있다. 예컨대 앞서 p=0.5를 주장했던 가설 h_0, 그리고 그를 테스트하기 위한 증거 e_0 역시 모두 동일하게 이항 분포상의 가설이자 증거였던 셈이다. 이를 떠나 그 가설이나 증거는 아무런 의미도 갖지 못한다. 그러므로 이제 방금의 구조를 중심으로 말한다면, 그 구조가 문제의 증거와 가설을 모두 '생성'한 것으로 간주할 수 있을 법하다. 나는 이처럼 어떤 증거와 가설을 함께 연관지어 확률적으로 생성해 내는 것으로 간주할 수 있는 대상들의 구조를 **'증거-가설 생성의 메커니즘'**(mechanism generating an evidence-hypothesis)이라 부를 것을 제안한다. 그리고 이하의 논의를 편리하게 하기 위해 이처럼 임의의 어느 증거 e와 가설 h를 함께 생성해 주는 증거-가설 생성의 메커니즘을 'MGeh'와 같은 식으로 표현하기로 한다. 이 경우 만일 동일한 메커니즘 내에서 단지 그 증거나 가설만이 다를 경우에는 'MG' 부분은 동일하게 유지하고, 단지 그 변화된 증거와 가설만을 달리 표현하면 좋을 것이다. 예컨대 동일한 이항 분포상에서 증거는 동일하게 e_0이지만, 서로 대립하는 두 가설 h_0와 h를 고려한다면, 우리는 이때 'MGe_0h_0'와 'MGe_0h'로 그 두 메커니즘을 나타낼 수 있을 것이다. 이때 특히 주의할 점은 이처럼 주어진 증거와 가설뿐 아니라 그를 생성하는 메커니즘마저 다른 경우이다. 즉 그 증거와 가설을 연관시켜 주는 확률 분포 유형이 서로 다른 경우이다. 이때에는 그 메커니즘이 다르다는 점을 분명히 나타내기 위해 'MG' 아래에 아래첨자를 써서 구별하면 좋을 듯하다. 예컨대 'MG$_1e_0h_0$'와 'MG$_2e_0h_0$'와 같은 식이다. 물론 메커니즘이 동일한 경우에도 이처럼 아래첨자를 사용해

동일하게 표현할 수 있을 것이나, 특별한 이유가 없는 한, 이런 경우에는 편의상 위에서처럼 아래첨자 없이 동일하게 'MG'로 표현하기로 한다.

이처럼 '증거-가설의 메커니즘'이라는 새로운 개념과 그 표현 기법을 이용하면, 이제 앞서 2절에서의 정동욱의 비판에 어떤 문제점이 있는가를 매우 선명하게 보여 줄 수 있다. 먼저 그 첫 번째 비판에서 그는 증거 e_0와 관련해 가설 h_0와 $G(e_0)$를 비교하고 있다. 이 두 가설은 모두 문제의 동전을 던질 때 앞면이 나타날 확률이 p>0.5임을 주장하는 가설 h에 대립되는 가설로 제시된 것들이다. 즉 가설 h가 거짓이라고 할 때 고려할 수 있는 가설들로 제시된 것이다. 그리하여 정동욱은, 가설 h_0 대신 $G(e_0)$를 대립 가설로 삼을 경우, 시험 R_G 역시 시험 R(X) 못지않게 엄격한 시험이 될 수 있다고 주장한다. 하지만 이 경우 두 가설 모두, h가 거짓이라 할 때, 그 h의 적절한 대립 가설이 될 수 있는 것인가? 위에서 제시한 증거-가설 생성의 메커니즘의 관점에서 볼 때, 이에 대한 답은 '아니다'이다. 그 메커니즘의 관점에서 보면, 가설 h와 h_0는 모두 동일한 메커니즘상의 두 가설이다. 확률 분포 유형으로 말하면, 두 가설 모두 **동일하게** 이항 분포나 정규 분포상에서 매개 변수 p에 대해서만 서로 다른 주장을 펴고 있는 가설들일 뿐이다. 물론 가설 $G(e_0)$ 역시 p에 대한 주장을 펴고 있기는 하다. 곧 이미 앞면이 나온 결과에 대해서는 p=1이고 뒷면이 나온 결과에 대해서는 p=0을 주장하고 있는 것이다. 하지만 이때 중요한 점은 그와 같은 가설이 주장되는 확률 분포 유형이 결코 이항 분포나 정규 분포는 아니라는 점이다. 만일 그때의 확률 분포에 대해 이름을 붙이자면, 그것은 어쩌면 '유리 겔러 분포'라고 해야만 할지 모른다. 중요한 점은 이러한 이름의 차이가 아니라, 그 이름의 차이 이상으로 그 양 분포에 분명한 차이가 있다는 점이다. 이를 앞서 증거-가설 생

성의 메커니즘의 표기법으로 나타내면, 위의 세 경우 각각 $MG_1 e_0 h$, $MG_1 e_0 h_0$, $MG_2 e_0 G(e_0)$가 될 것이다. 이로써 처음 두 경우에는 해당 증거와 가설이 모두 동일한 메커니즘 MG_1상의 그것이나, 마지막 경우에는 그것이 전혀 다른 메커니즘 MG_2에 놓여 있음이 분명하다.

이러한 나의 지적에 혹자는 어쩌면 그처럼 해당 메커니즘이 상이하다는 점이 문제의 두 가설 h_0와 $G(e_0)$를 비교하는 데 근본적으로 왜 문제가 되는가라고 물을지 모른다. 이에 대해서는 다음과 같은 스틸의 예가 도움이 될 수 있으리라 본다. 예컨대 어느 고고학적 유적지가 남겨지게 된 시기를 지금으로부터의 햇수로 나타낸 매개 변수 θ를 고려해, 그에 대해 일정한 값, 예컨대 3000, 3001, … 등등을 부여하는 명제를 각기 하나의 가설로 보기로 해 보자. 그렇다면 θ=3000과 θ=3001이라는 두 가설은 모두 동일한 매개 변수 θ에 대해 그 값을 부여하고 있는 셈이다. 하지만 이제 어느 가설은 θ=3000을 주장하고, 또 다른 어느 가설은 '에펠탑의 높이는 1000피트다'라고 주장한다면, 우리가 그 두 가설을 온전히 서로 비교할 수 있겠는가? 스틸은 어느 두 가설이 동일한 매개 변수에 대해 일정한 값을 부여하는 경우에만 그 가설들이 '구조적으로 동일하다'(structurally identical)라고 부르고, 이에 따라 방금의 예에서 전자의 경우에는 두 가설이 구조적으로 동일하나, 후자의 경우에는 그렇지 않다고 본다.[19] 이처럼 구조적으로 동일하지 않은 가설들을 비교한다는 것은, 그 주제가 전혀 다른, 물리학의 어느 가설과 예술의 또 다른 가설을 비교하는 것과 다르지 않지 않은가? 어느 두 가설이 구조적으로 동일하지 않다면, 그 각각이 서로 다른 증거-가설 생성의 메커니즘에서 생성된 것임도 분명하다. 그런데 이러한 식의 비교가 바로 정동욱이 제시한 가설들 사이에서도 이루어지고

19) Steel (2007), p. 68 참조.

있는 것이다. 앞서의 동전 가설 h와 h_0에서는 모두 그 매개 변수가 동전을 던지는 **각 시행마다** 그 앞면이 나올 확률에 대한 것이다. 반면 가설 $G(e_0)$에서는 그 매개 변수가 동전을 던지는 시행에서 **이미 앞면이 나온 경우와 이미 뒷면이 나온 경우**의 확률에 대한 것이다. 따라서 h와 h_0는 서로 구조적으로 동일하다. 하지만 그것들과 $G(e_0)$ 사이는 전혀 그렇지 않다.[20]

이렇게 본다면, 정동욱의 두 번째 비판에 대해서도 마찬가지 방식으로 대응이 가능하다. 그 비판에서 정동욱은 h 이외에 e_0를 설명할 수 있는 대립 가설로서 $G(e_0)$를 설정하고, $G(e_0)$가 참일 경우에도 시험 T가 가설 h를 통과시킬 수 있는 확률이 1인 까닭에 그 시험의 엄격성이 0이 됨을 지적한 바 있다. 이렇게 되면, 통상적 관점에서는 가설에 대한 좋은 증거가 그에 대한 겔러식 대립 가설에 비춰서는 더 이상 좋은 증거로 간주될 수 없는 파국에 이르게 된다는 것이다. 하지만 이때 앞서의 증거-가설 생성의 메커니즘에 따르면, $G(e_0)$는 애초에 h의 대립 가설이 될 수 없다. 그러므로 h가 거짓이라고 가정할 때, 이 경우 $G(e_0)$를 참으로 설정하는 일은 불가능하다.

물론 이렇다 할지라도 정동욱이 이와 같은 사실을 전혀 눈치채

---

20) 이와 관련해 한 심사위원은 "[필자가] 가설의 구조적 동일성이 '증거-가설 생성 메커니즘'의 동일성을 보증하는 것으로 보고 있는데, 반드시 그렇지는 않다"고 지적하며 양자의 관계에 대한 추가 논변이 필요하다고 말한 바 있다. 그러나 본 논문에서도 전자가 후자를 보증하는 것으로 보고 있는 것은 결코 아니다. 단지 증거-가설 생성의 메커니즘이 상이할 때 나타날 수 있는 기묘한 상황 하나를 예시하기 위해 스틸의 구조적 동일성 개념을 빌려 온 것뿐이다. 증거-가설 생성의 메커니즘이 동일한 경우, 매개 변수와 관련해 구조적으로 동일해야 할 뿐만 아니라 그 확률 분포 유형 역시 동일해야 한다는 점에서 증거-가설 생성 메커니즘의 동일성은 구조적 동일성을 넘어선다. 이러한 해명이 또한, 위의 본문 가운데 카드-동전-스틸의 사례가 함께 얽혀 매우 복잡해 보인다는 또 다른 심사위원의 지적에 대해서도 적절한 응답이 될 수 있기를 희망한다.

지 못했다고 보지는 않는다. 왜냐하면 메이오의 입장을 고려하는 과정에서 그는 이미 시험 절차 R_G와 R(X)의 차이가 무엇에서 비롯되는가를 살피는 중 $G(e_0)$가 동전만에 대한 가설이 아니라 시점까지를 포함한 가설일 수도 있다는 점에 주목했기 때문이다. 그리하여 새로운 증거 e_1이 주어진 경우, 가설 $G(e_1)$과 $G(e_0)$는 서로 다르더라도 양립 가능한 반면, $h(e_0)$와 $h(e_1)$은 서로 다를 경우 양립 불가능하다고 보고, 그러한 차이가 어쩌면 두 가설 구성 절차에 차이를 가져 왔을지도 모른다고 생각한 바 있다. 하지만 그는 곧 이때 $G(e_0)$는 가설 h_0와도 양립 가능할 수 있음을 들어, h_0를 대립 가설로 가정해 R_G를 엄격하지 않은 겔러화된 시험의 사례로 제시한 메이오의 분석이 의미를 상실한다고 지적하였다. 그리하여 위와 같은 양립 가능성의 해석 아래 R_G의 엄격성을 따지기 위해서라면 또 다른 종류의 대립 가설을 가정해야 할 터인데 대체 어떤 대립 가설이 가능하겠느냐고 묻기도 하였다.

하지만 정동욱의 말처럼 $G(e_0)$가 비록 h_0와 양립 가능할지라도, 그 양자를 서로 대립하는 가설로 설정할 수는 없다. 앞서 언급한 메커니즘의 차이 때문이다. 그러므로 문제는 단순히 그가 말하는 "시점"에만 놓여 있는 게 아니다. 나아가, 이러한 관점에서 본다면, R_G의 엄격성을 따지기 위해 필요한 또 다른 종류의 대립 가설이 무엇이겠느냐는 물음에 대해서도 그것은 $G(e_0)$와 동일한 메커니즘에 놓여 있는 예컨대 $G(e_1)$이 될 수 있다고 답해 줄 수도 있다. 이 경우 그 가설들에 대한 시험으로는 예컨대 '증거 e_k(k=0, 1, ⋯, n)를 기반으로 가장 우도가 큰 가설 G_k를 통과시킨다'라는 절차를 삼을 수 있다. 이제 이 시험을 앞서의 R_G와는 다른 새로운 시험 R_{Gnew}라고 해 보자. 그렇다면 이 R_{Gnew}에 따를 때, 만일 e_1에 의해 $G(e_0)$가 거짓이고, 그래서 $G(e_1)$이 참이라 할지라도, 해당 시험이 e_0에 의해 $G(e_0)$를 통과시킬 확률은 1이 되고, 따라서 그

시험의 엄격성은 0이 된다. 이와는 달리, $h(e_0)$, $h(e_1)$, h_0는 모두 같은 메커니즘상의 서로 다른 세 가설이다. 그리고 대칭적인 이항 분포나 정규 분포상에서는 가설 $h(e_0)$나 $h(e_1)$ 모두 h_0와 비교될 때 해당 시험의 엄격성이 가장 커진다. 그러므로 이때에는 $h(e_0)$나 $h(e_1)$ 각각 h_0와 대립되는 가설로 설정되어 아무런 문제없이 엄격한 시험에 부쳐질 수 있는 것이다.

자신의 두 번째 비판과 관련해 메이요의 편에 서서 정동욱은 다음과 같은 선택지를 고려하기도 하였다. 곧 $G(e_0)$가 h와 양립 불가능한 가설이 아니므로 그러한 겔러식 가설은 대립 가설로서 배제해야만 한다는 선택지이다. 하지만 그는 곧 이 선택지를 따른다 할지라도 다른 선택지들과 마찬가지로 결국 표준적인 통계적 시험 절차조차 $G(e_0)$와 같은 종류의 대립 가설은 배제함으로써만 엄격한 시험으로 간주될 수 있을 것이라는 회의론은 편 바 있다. 하지만 이 경우에도 우리는 이제 메이요가 그러한 선택지를 택할 필요가 전혀 없음을 분명하게 말해 줄 수 있다. 바로 앞서와 같은 이유들로, $G(e_0)$와 h는 서로 다른 메커니즘상의 두 가설로, 양자가 비록 서로 양립 가능하다 할지라도, 그것들을 엄격한 시험을 위한 대립 가설들로 볼 필요는 전혀 없고, $G(e_0)$는 어디까지나 엄격성이 최소인 가설이라는 것이다.

물론 이 과정들에서 나는 메이요 역시 전혀 잘못이 없다고는 생각지 않는다. 적어도 Mayo (1996)의 202쪽에서 그녀 역시 앞서 시험 R_G로서 '바로 증거 e를 기반으로, 가설 h_0는 탈락시키고 최대로 우도가 큰 가설 $G(e)$는 통과시킨다'는 절차를 제시함으로써, 가설 $G(e)$가 h_0와 대립될 수 있는 것으로 잘못 가정하고 있다. 지금의 나의 관점에서, 이는 매우 큰 실수라 생각한다. 나는 바로 이 대목이야말로 정동욱으로(그리고 이후에 좀더 자세히 논할 이세다로) 하여금 잘못된 비판에 이르게 한 핵심 대목일 것이라 추측한

다. 앞서 제2절에서 정동욱이 강조한 대로, 메이요가 어떤 시험의 엄격성을 이해함에 있어 자료가 아니라 '자료 생성'의 방식이 고정된 것으로 본 것은 옳다고 생각한다. 그러나 자신의 잘못된 실수로 불필요한 비판을 자아낸 것은 시정되어야 할 것이다. 이 점에서 나는 메이요가 단지 증거에 해당하는 자료의 '자료 생성 메커니즘'(data-generating mechanism)에만 주목할 것이 아니라[21] 바로 그 증거와 관련 가설 모두를 함께 고려하는 '증거-가설 생성의 메커니즘'에 주목했어야 할 것이라 본다.

이렇게 본다면, 나아가 (앞서 제2절에서 인용한 대로) 같은 202쪽에서 시험 R_G와 관련해 메이요가 "$G(e)$가 거짓이고 h_0가 참일지라도(즉 문제의 동전이 '공정하다'고 할지라도) [⋯]"라고 한 대목 역시 정확히는 '$G(e)$에 대한 유리 겔러 분포 가정이 거짓이고 h_0에 대한 이항 분포 또는 정규 분포 가정이 참일지라도 시험 R_{Gnew}에서 ⋯'와 같이 표현했어야만 했을 것이다. 메이요가 만일 이처럼 표현했다면, 적어도 정동욱이(그리고 이후에 논할 이세다가) 잘못된 비판에 이를 가능성은 거의 사라졌을 것이라 본다.

이로써 메이요에 대한 정동욱의 핵심적 비판들에 대한 대응을 마친 셈이다.[22] 그렇다면 이제 이러한 대응을 바탕으로 이세다의 비판에 대해서도 유사하게 대응이 가능하리라 본다. 이에 대해서는 절을 바꾸어 논해 보기로 하자.

---

21) 이 용어는 메이요가 예컨대 Mayo (2014), p. 84에서 실제 사용하고 있는 용어이기도 하다.

22) 정동욱 (2018)에서 메이요에 대한 비판은 계속 이어진다. 특히 활자와 관련된 사례는 매우 흥미롭기도 하다. 하지만 그 사례 역시 증거-가설 생성의 메커니즘 차이로 쉽사리 대응이 가능하다. 다만 그의 추가 비판과 그에 대한 나의 대응은 더 이상 필요치 않으리라 본다. 남은 지면을 위해 그러한 논의는 생략하기로 하자.

4. 이세다의 비판과 그에 대한 대응

메이요에 대한 이세다의 비판 역시 그 핵심에 있어 정동욱과 다르지 않다. 즉 어떤 시험에 대한 그녀의 엄격성 판단이 일관적이지 않다는 것이다. 그러므로 사실상 이세다의 주요 비판 중 하나 역시 앞서 정동욱의 첫 번째 비판과 동일하다. 오히려 정동욱의 그것보다 시기적으로 앞서고 있다.[23] 따라서 이에 대한 재론은 불필요할 듯하다. 하지만 이세다의 나머지 비판들은 정동욱의 그것과는 다소 차이가 있다. 이하에서는 그에 대한 논의가 짧은 순서에 따라 나아가 보기로 한다.

이세다의 나머지 비판 중 첫 번째는 다음과 같다.[24] 먼저 앞서 시험 R(X)에 따라 주어진 가설 h_0와 유사한 형태로, 이미 주어진 증거에 맞춰 구성된 모든 가설들[25]의 집합을 고려해 보기로 하자. 이 경우 그 집합 중 적어도 하나의 가설은 참이 될 것이라는 의미로 문제의 집합은 망라적이라 할 수 있다. 따라서 그 대립되는 가설들 가운데 있을 법하지 않은 것들을 제거함으로써 우리는 믿을 만한 가설로 나아갈 수 있다. 하지만 이제 앞서 메이요의 시험 R_G에 따르면, 우리는 h_0가 무엇이든 그것을 탈락시킬 수밖에 없다. 그리고 이렇게 된다면, 그 가설들의 집합은 결코 망라적이지 않고, 따라서 방금의 전략도 적용 불가능하게 된다. 이세다는 이 두 가지 경우의 차이를 과학의 실례로써 보여 주고 있다. 잘 알려진 대로, 아인슈타인은 이미 알려져 있던 현상인 수성의 근일점 이동에 맞춰

---

23) Iseda (1999), pp. S410-411. 정동욱 (2018), p. 38에서도 이 점을 지적하고 있다.

24) Ibid., pp. S411-413. 여기서도 이세다의 용어와 기호법을 우리에 맞게 수정하였고, 이하 마찬가지이다.

25) 이때의 가설을 흔히 ‘사용-구성된’(use-constructed) 가설이라 부른다. 즉 그 가설은 이미 해당 증거에 맞게 용의주도하게 구성된 가설인 셈이다.

자신의 일반 상대성 이론을 구성한 바 있다. 그럼에도 불구하고 그 현상은 해당 이론의 적절한 증거로 인정을 받았는데, 그 이유는 아인슈타인의 해당 이론 외에는 문제의 현상을 설명할 수 있는 여타의 경쟁 이론이 불가능했기 때문이다. 반면 지동설이 대두되었음에도 불구하고, 새로이 나타난 이상 현상에 대해 그에 맞춰 임시변통적으로 그를 포섭해 계속 지동설을 유지하는 일은 정당화될 수 없는 일이라고 이세다는 지적한다. 왜냐하면, 바로 그러한 포섭을 통해, 있을 수 있는 또 다른 대립 가설의 존재를 무시했기 때문이다. 이것은 좋지 않은 과학적 방법이다. 이세다는 이처럼 사용-구성된 가설이 임시변통적인 포섭을 통해 다른 대립 가설들을 회피하는 전략을 "연관된 대립 가설의 체계적 무시"(systematic neglect of relevant alternatives)라 불렀다.

그런데 앞서 소개된 겔러식 가설의 구성이야말로 바로 이와 같은 체계적 무시의 한 전형이다. 그러므로 이세다는 우리가 겔러식 가설을 배격해야 하는 이유는 메이요가 말하듯 그 최소의 엄격성 때문이 아니라 바로 그 대립 가설의 체계적 무시 때문이라고 주장한다. 게다가 겔러식 가설이 구성되어 R_G가 행해지면, 심지어 증거를 얻기 이전일지라도 우리는 미리 h_0를 탈락시킬 수 있음도 알고 있다. 이렇게 되면 어떠한 새 증거라도 그것을 포섭하는 체계적 전략으로써 아예 경험적 증거를 쓸모없는 것으로 전락시키고 만다.[26)] 이것 역시 엄격성 이전에 겔러식 가설 구성이 또 달리 문제되는 이유이다.

그런데 만일 겔러식 가설이 그 엄격성 때문이 아니라 이제 이상과 같은 이유들 때문에 무시되어야만 하는 것이라면, 그것은 대립 가설로서 아무런 연관성이 없기 때문일 것이다. 하지만 이때 그 연관성의 여부를 가리는 기준이 대체 무엇일 수 있는가? 이에 대해

26) 이러한 비판 역시 앞서 정동욱의 두 번째 비판과 크게 달라 보이지 않는다.

이세다는 어쩌면 그러한 가설에 대해서는 “상당히 낮은 확률”을 부여함으로써 처리할 수 있을 법한데, 그것은 너무나도 주관적이라 그에 대해 어떤 형식적 규칙을 만들 수 있을지 확신이 서지 않는다며 앞으로의 계속적인 탐구를 요구하고 있다.

하지만 이세다의 이상과 같은 비판과 회의는 사실 불필요한 것이다. 우선 우리의 증거-가설 생성의 메커니즘 개념에 따르면, 이미 언급한 대로, 가설 h_0와 $G(e)$는 서로 대립할 필요가 없는 관계이다. 게다가 시험 R_G가 아니라 R_{Gnew}에 따르면, $G(e)$에 대한 시험은 최소의 엄격성을 가질 수밖에 없다. 그러므로 겔러식 가설 구성에 의해 이세다가 말하는 바와 같은 문제들이 발생한다 할지라도, 그러한 가설은 또한 엄격한 시험에서도 이미 탈락되어야만 할 가설이다. 만일 사정이 이러하다면, 또한 겔러식 가설을 배제함에 있어 연관성 문제에 대한 고민도 더 이상 있을 필요가 없다. 연관성 여부를 따지기 이전에 그 가설은 엄격성 시험을 통과하지 못하기 때문이다.

다음으로 이세다의 나머지 비판 중 두 번째는 다음과 같다.[27] 그는 우선 사용-구성된 가설에 대한 기어리의 입장을 소개하며 시작한다.[28] 기어리에 따르면, 만일 어떤 가설 h가 사용-구성되었다면, “어느 경우이든”(no matter what) 그것은 확실하게 해당 증거와 성공적으로 부합된다. 그러므로 그 가설이 비록 거짓이라 할지라도 그것의 성공 확률은 거의 1로 높게 된다. 이것은 그 가설의 신뢰성에서 분명 문제가 되는 상황이다. 그러므로 기어리는 어떤 증거가 h에 대해 제대로 지지를 하려면 그것은 결코 h를 구성하는 데 사용되어서는 안 되고 어디까지나 사용-참신한(use-novel) 것이어야 한다고 주장한다. 하지만 이에 대해 메이요는 반대하며, 어떤 증거

27) Iseda (1999), pp. S405-411.

28) 기어리의 입장 자체에 관해서는 Giere (1983) 참조.

가 해당 가설에 대해 제대로 된 증거이기 위해 그것이 반드시 사용-참신할 필요는 없다고 주장하였다. 이를 위해 메이요는 기어리가 말하는 “어느 경우이든”이라는 대목은 “증거가 무엇이든”(no matter what evidence is)이라는 뜻일 뿐, 그것이 “가설 h가 참이든 거짓이든 관계없이”(no matter if h is true or false)라는 뜻과 동일한 것은 아니라며, 다음과 같은 두 가지 확률을 구별한다.[29]

A. 시험 T가 그것이 시험하는 가설을 통과시킬 확률
B. 시험 T가, 그것이 시험하는 가설이 거짓인 경우, 그것을 통과시킬 확률

여기서 말하는 시험 T는, 가설 h가 증거 e에 맞춰 사용 구성된 가설이고, 그러한 증거 e로써 h를 통과시키는 절차를 말한다.

이렇게 두 확률을 구별한다면, 물론 확률 A는 1이 분명하다. 하지만 메이요는, A가 1임에도 불구하고, 확률 B는 1보다 작을 뿐만 아니라 아주 작은 값이기를 바라는 셈이다. 그러나 이제 이세다는 이것이 불가능함을 다음과 같이 보여 준다.

우선 어느 증거가 나타나 전개될 상황인 표본 공간은 다음처럼 4가지로 나뉠 뿐이다.

<표 4.1>

	h가 참	h가 거짓
T가 h를 통과시킴	C1	C2
T가 h를 탈락시킴	C3	C4

이 경우 물론 확률 A는 (C1+C2)/(C1+C2+C3+C4)로 계산된다. 그

[29] Mayo (1996), pp. 263-271.

리고 확률 B는 C2/(C2+C4)로 계산될 것이다. 그런데 A가 1이 되는 것은 오직 공간 C3와 C4가 비어 있을 때뿐이다. 따라서 만일 A가 1이 된다면, B 역시 C2/C2=1이 될 수밖에 없다. 하지만 이러한 결과는 결코 메이요가 원하는 바가 아니다. 따라서 이 경우 만일 B가 1보다 작은 값이기를 바란다면, 이것이 가능한 경우란 오직 C4뿐만 아니라 C2 역시 비어 있어, 확률 0/0의 결과를 적절히 해석할 수 있을 때뿐이다. 그렇다면 이에 대해 어떤 해석이 가능한가?

이에 대해 이세다는 메이요의 다음과 같은 사례가 어쩌면 그 답의 하나가 될 수 있을지 모른다고 제안한다.[30] 이제 어느 한 반의 학생들에 대해 미국의 수학 능력 시험인 SAT 성적의 평균을 생각한다고 해 보자. 이 경우 그 평균이 얼마인지에 대한 가설을 어떻게 세울 수 있을 것인가? 단순히 추측이 아니고 참일 법한 가설을 세우는 자연스런 한 방법은 물론 그 반 학생들의 SAT 성적을 모두 더하고 그것을 학생 수로 나누는 방법이다. 이제 그런 방법을 통해 얻은 결과가 1121이라고 해 보자. 그렇다면 이렇게 해서 얻어진 가설이 문제의 증거, 즉 실제 그 반 학생들의 성적에 의해 제대로 지지된다고 볼 수 있겠는가? 지금의 경우 그 가설이 사용-구성된 가설임은 분명하다. 따라서 이때의 확률 A는 1이 된다고 할 수 있다. 그럼에도 불구하고 메이요는 이러한 식의 가설에 관한 시험은 엄격한 시험이라고 주장한다. 그 까닭은, 이 경우 만일 문제의 가설이 거짓이라면 해당 시험에서 문제의 가설을 탈락시킬 확률은 1이 되기 때문이다. 그러므로 이제 만일 우리가 이러한 사례의 설득력을 받아들인다면, 이세다는 앞서의 확률 B에 대해서도 그것이 0/0의 경우일지라도 우리는 그것을 0으로 해석하는 데 직관적으로 거부감이 들지 않을 것이라 지적한다. 그러나 이것이 과연 다른 사

---

30) 메이요의 사례 자체에 관해서는 ibid., pp. 271-272 참조.

례에서도 설득력 있는 것일 수 있는가?

이를 위해 이세다는 과학적으로 좀더 현실성 있는 사례로 신뢰 구간 가설의 구성 사례를 들고 있다. 이는 우리가 앞서 제2절에서 소개했던 바로 그것이다. 그리고 그 관련 시험 R(X)에 대해 메이요는 그것이 엄격한 시험이라 주장한 바 있다. 하지만 그것의 정확한 의미는 무엇인가? 이를 위해 이세다는 그 의미가 다음의 세 가지일 수 있다고 분석한다.

(SC*-1) $h(e_0)$가 거짓인 경우, R(X)가 $h(e_0)$를 탈락시킬 확률이 매우 높다.
(SC*-2) $h(x)$가 거짓인 경우, [그 x의] 각 경우에 R(X)가 $h(x)$를 탈락시킬 확률이 매우 높다.
(SC*-3) $h(e_0)$가 거짓인 경우[일지라도], 만일 증거 e_0가 얻어질 때라면 R(X)가 $h(e_0)$를 탈락시킬 확률이 매우 높다.

이 경우, 이세다는 메이요가 주장하는 결과에 제대로 이를 수 있는 해석은 (SC*-1)일 것이라 본다. 왜냐하면 $h(e_0)$가 거짓인 경우 현재의 증거인 e_0와 동일한 증거를 다시 얻기란 매우 힘들고,[31] 따라서 $h(e_0)$가 R(X)를 통과하기란 매우 힘들 것이기 때문이다. 그럼에도 불구하고 이세다는 이 경우는 메이요의 원래 목적에 비춰 문제가 있다고 지적한다. 왜냐하면 이 경우에는 확률 A, 즉 시험 R(X)가 $h(e_0)$를 통과시킬 확률이 1이 아니기 때문이다.[32] 그 까닭은, 이때에는 우리가 동일한 가설 구성의 절차를 따를 경우 매 경우마다 주어지는 증거가 e_0가 아니고 다른 증거가 나와, 그 다른

31) 이 경우 예컨대 e_0와는 달리 e_1, e_2, … 등등, 그 표본 평균이 처음과는 다른 증거들이 제시될 가능성이 크다. 이러한 점에 관해서는 앞서 제2절에서의 정동욱으로부터의 첫 번째 인용 대목을 참조하는 것도 좋을 것이다.

32) 이 대목에서 이세다는 "통과"(pass)라는 말 대신 "탈락"(fail)이라는 말을 쓰고 있는데(p. S408), 이는 단순 실수나 오타로 보인다.

증거에 따라 구성되는 가설 역시 얼마든지 가능하고, 그래서 우리가 그 다른 가설을 통과시킬 가능성도 열려 있기 때문이다. 그러나 이처럼 A가 1이 아니라면, 이것은 원래 메이요가 기어리의 입장을 비판하며 의도했던 바에서 벗어나는 것 아닌가? 왜냐하면 앞서 언급한 대로 기어리가 염두에 두었던 경우란 바로 A가 1이 되는 경우였기 때문이다.

물론 이러한 난점에서 벗어날 수 있는 것은 나머지 두 경우, 즉 (SC*-2)와 (SC*-3)의 경우이다. (SC*-2)의 경우는 앞서 우리의 기호법대로 표현한다면 가설 $h(e_k)$에 대한 것으로, 각각의 증거 e_k(k =0, 1, ⋯, n)에 따라 구성된 가설을 대상으로 한 것이다. 반면 (SC*-3)의 경우는 이때의 증거를 특정하게 e_0로 한정한 경우일 뿐이다. 그런데 (SC*-2)의 경우 각 증거에 따라 구성되는 가설은 당연히 그 증거와 부합될 수밖에 없고, 이로써 그 가설이 증거와 부합되는 확률은 1이 될 수밖에 없으며, 이러한 사정은 그 특수한 경우인 (SC*-3)의 경우에도 마찬가지이다. 하지만 이때에는 앞서의 논의대로 확률 B 역시 1이 되어, 그 시험의 엄격성 정도가 최소인 0이 될 수밖에 없다고 이세다는 비판한다.

물론 이 경우일지라도 메이요는 자신이 원래 의도했던 바는 확률 B가 A와 다름을 주장해 기어리를 비판하기보다 오히려 (SC*-1)에 부합되는 나름의 사용-구성적 가설에 대해 엄격한 시험이 가능함을 주장하는 일이었다는 식으로 회피가 가능할 수도 있다고 이세다는 지적한다. 그러나 이 경우에도 메이요는 다시금 새로운 문제에 부딪친다고 이세다는 비판한다. 이 문제는 바로 우리가 제2절에서 소개했던 겔러식 가설과 관련된다.

먼저 이세다는 겔러식 가설에 대해서도 신뢰 구간 가설에서와 꼭 마찬가지로 (SC*-1), (SC*-2), (SC*-3)에 차례로 대응하는 (SC*-4), (SC*-5), (SC*-6)를 고려할 수 있다고 본다. 이들 사이의

유일한 차이란, 전자의 경우 대상 가설이 $h(e_0)$와 $h(x)$였다면, 후자의 경우 그것이 단지 $G(e_0)$와 $G(x)$로 대치되었다는 점뿐이다. 물론 이 경우 메이요는, 앞서 우리가 제2절에서 살핀 대로, $G(e_0)$는 시험 R_G에서 탈락할 수밖에 없다고 주장한 바 있다. 그러나 메이요의 이러한 주장은 앞서 신뢰 구간 가설에 대한 그녀의 입장에 비춰 매우 비일관적인 결과를 초래한다고 이세다는 비판한다.

우선 그의 생각에 신뢰 구간 가설에 대한 메이요의 입장이 가장 잘 반영될 수 있는 해석은 앞서의 (SC*-1)였다. 그렇다면 겔러식 가설과 관련해서도 메이요의 입장을 가장 잘 반영할 수 있는 해석은 그에 대응하는 해석 (SC*-4)일 법하다. 그리하여 이와 같은 해석에 따라 이제 $G(e_0)$가 거짓이고 h_0가 참이라고 해 보자. 그런데 이때에는, 앞서 정동욱의 비판에서 제시된 대로, $G(e_0)$가 R_G를 통과할 확률이 상당히 낮고(앞서 정동욱의 예에서처럼 1/16), 따라서 그 가설에 대한 해당 시험의 엄격성은 매우 높다고 할 수 있다(앞서 정동욱 예에서처럼 15/16).

이러한 결과에 직면해, 그렇다면 이제 메이요가 선택할 수 있는 해석들은 나머지 (SC*-5)와 (SC*-6)일 뿐일 것이라고 이세다는 추측한다. 왜냐하면 이 두 경우에는 해당되는 겔러식 가설에 대한 시험의 엄격성이 모두 0이 되기 때문이다. 따라서 만약 메이요가 겔러식 가설이 엄격한 시험을 통과하지 못하기 때문에 배제되어야만 하는 것이라 주장하길 원한다면, 그녀는 지금의 두 해석을 취해야만 할지 모른다. 하지만 왜 앞서의 신뢰 구간 가설에 대해서는 해석 (SC*-1)을 취한 반면, 겔러식 가설에 대해서는 해석 (SC*-5)와 (SC*-6)를 취해야 하는 것인가? 게다가 두 해석 (SC*-5)와 (SC*-6)를 취할 경우 확률 A는 여전히 1로서, 이는 메이요가 바라는 바가 결코 아닐 것이다.

이상의 이세다의 비판 역시 언뜻 메이요에게는 매우 치명적인

것으로 보인다. 그러나 정동욱의 그것에 대해 그러했듯 이세다의 비판에 대해서도 나는 '증거-가설 생성의 메커니즘'이란 개념이 매우 유효적절한 대응 수단이 될 수 있다고 생각한다.

무엇보다 <표 4.1>에 근거해 이세다가 주장하는 확률 A와 B의 연동 관계, 즉 A가 1이면 B 역시 반드시 1이 된다는 관계는 메이요에 대해 잘못 적용된 것이다. 먼저 우리는 해석 (SC^*-2)와 (SC^*-3)에 등장하는 가설들은 그 기호의 차이에도 불구하고 사실 (SC^*-4), (SC^*-5), (SC^*-6)에서 고려하고 있는 겔러식 가설과 본질적으로 다르지 않음에 주의할 필요가 있다. 왜냐하면 그때의 가설들 모두 증거가 아무리 변화한다 할지라도 모두 그 증거에 부합되게 구성된 가설들이기 때문이다. 그런데 앞서 우리는 제3절에서의 논의를 통해 겔러식 가설이 생성되는 메커니즘과 신뢰 구간 가설이 구성되는 그것이 서로 상이함을 보인 바 있다. 따라서 이러한 관점에서 우선 이세다가 제시한 <표 4.1>은 주의 깊게 두 가지 경우로 분리될 필요가 있다. 첫째는 겔러식 가설 G에 대해 표본 공간을 보여 주는 경우이다. 둘째는 신뢰 구간 가설 h에 대해 그것을 보여 주는 경우이다. 이 가운데 첫 번째 경우의 표는 사실상 <표 4.1>이 아닌 다음과 같은 <표 4.2>의 형태를 취해야만 옳다.

<표 4.2>

	G가 참	G가 거짓
T가 G를 통과시킴	C1	C2

이 새로운 표에 따르면, 확률 A가 (C1+C2)/(C1+C2)=1임에도 불구하고, 그와 연동해 확률 B를 고려할 필요는 전혀 없다. 양자는 서로 무관한 것이다.

위의 두 번째 경우에는 물론 앞서 이세다가 제시한 <표 4.1>가

그대로 적용될 수 있다. 다만 이때에는, 이세다도 잘 지적한 대로, A가 1보다 작을 뿐만 아니라, 그리하여 B 역시 반드시 1일 필요가 없다. 따라서 A와 B의 연동 관계는 성립하지 않는다. 오히려 이 경우에는 A에 비해 B가 훨씬 작아, 시험 T는 가설 h를 엄격하게 탈락시킬 수 있다. 그러므로 역시 이세다도 인정했듯, 이 경우에 잘 맞는 해석은 물론 (SC*-1)이다. 하지만 겔러식 가설에 대해 메이요가 (SC*-1)에 대응하는 (SC*-4) 대신 (SC*-5)와 (SC*-6)을 취할 수밖에 없게 되는, 엄격성 판정의 비일관성 문제는 어떻게 되는 것일까?

이에 대해서는 먼저 해석 (SC*-4)와 관련해 이세다가 $G(e_0)$와 h_0를 상호 대립적인 가설로 설정한 것 자체가 잘못임을 지적할 필요가 있다. 물론 이러한 잘못의 단초를 메이요가 제공한 점에 관해서는 앞서 제3절에서 지적한 바 있다. 하지만 그녀의 이러한 잘못을 차치한다면, 해석 (SC*-4)와 관련해 우리는 이제 (이세다의 표에서 시험 T에 대응하는) 시험 R_G 대신 바로 시험 R_{Gnew}를 고려할 필요가 있다. 그리고 이러한 새로운 시험 하에서라면, 앞서 논한 대로 그 엄격성은 0이다. 따라서 이 경우 이세다가 생각했듯 시험 R_G의 엄격성이 상당히 높다고 우려할 필요는 전혀 없다.

이렇게 본다면, 해석 (SC*-1)을 제외한 나머지 모든 해석에서 확률 B와는 무관하게 확률 A가 1이 됨은 분명하다. 우선 해석 (SC*-2)과 (SC*-3), 그리고 (SC*-5)와 (SC*-6)의 경우에 그러함은 이미 이세다가 지적한 대로이다. 하지만 이제 시험 R_G 대신 R_{Gnew}를 취하면, 해석 (SC*-4)의 경우에도 확률 B와는 무관하게 확률 A가 1이 됨을 알 수 있다. 그리고 이 모든 경우, 즉 (SC*-2)~(SC*-6)의 경우에 시험 R_{Gnew}의 엄격성은 0이다.

따라서 메이요가 처음에 기어리의 사용-참신성 요구에 반대하며 사용-구성된 가설의 경우에도 해당 시험이 엄격할 수 있음을 주장

하는 데 있어 해석 (SC*-1) 하에서는 그 의도가 빗나가게 된다는 이세다의 지적 역시 이제는 힘을 잃게 된다. 왜냐하면 기어리가 반대하는 사용-구성된 가설들은 ‘우리의 제대로 된 관점에 따른’ 메이요의 경우 모두 겔러식 가설에 해당하고, 이러한 가설들은 모두 엄격성이 0이 되어, 메이요가 그 엄격성 판정에서 일관성을 잃을 필요가 전혀 없기 때문이다. 이제 메이요에게 있어 사용-구성된 가설이지만 의미 있는 가설은 신뢰 구간 가설과 같은 것으로, 이것은 아무 문제없이 해석 (SC*-1) 하에서 엄격성이 높은 시험을 통과할 수 있다.

이렇게 본다면, 또한 앞서 SAT 성적 평균에 관한 메이요의 사례에 대한 이세다의 시각에 관해서도 수정이 가능하다. 그 사례를 두고 이세다는 메이요가 확률 B에 대해 0/0을 0으로 해석할 필요가 있다고 지적하고 있으나, 이제 그러한 시각은 전혀 불필요하다. 무엇보다, 앞서 제1절에서 강조한 대로, 어떤 시험 T의 엄격성 정도를 계산하기 위해 우리가 알아야 할 확률 P(시험 T가 가설 h를 통과시킴; h가 거짓)에서 “;”를 조건부 확률 계산을 위한 “/”으로 오해해서는 안 된다. 그러므로 이때 “; h가 거짓”이라는 말의 의미는 정확히는 ‘h가 생성되는 증거-가설 생성의 메커니즘과 동일한 메커니즘상에서 h 아닌 또 다른 가설 h'이 참이라고 가정할 때’라는 의미일 따름이다. 따라서 이러한 관점에서 본다면, 만일 해당 반의 SAT 성적 평균이 1121이 아니고 또 다른 값일 경우, 그 다른 값이 나올 만한 학생들의 성적 분포는 불가능하고, 확률 P(시험 T가 가설 h를 통과시킴; h가 거짓)=0이 된다. 이로써 시험 T가 엄격성이 최대인 1이 됨은 물론이다.[33)]

이러한 상황을 반영하면 위의 <표 4.1>은 또한 다음과 같이 수

33) SAT 성적 평균 사례에 관한 한, Worrall (2010), p. 148에서도 이세다와 비슷한 비판을 하고 있다. 그리고 그에 관해 Mayo (2010), p. 164에서도 지금의 나와 유사한 방식으로 답변이 이루어지고 있음을 볼 수 있다.

정 가능하다. 우선 지금의 상황에서는 'h가 거짓일 때 T가 h를 탈락시키는' 경우는 'T가 h를 통과시킬 때 h가 참'인 경우와 완전히 동일하다. 즉 <표 4.1>에서 C1과 C4는 서로 동일하다. 지금의 상황에서는 전자가 참이면 후자도 반드시 참이고, 이 역도 성립하기 때문이다. 그리고 'h가 거짓일 때 T가 h를 통과시키는' 경우나 'h가 참일 때 T가 h를 탈락시키는' 경우는 각각 가능하긴 하나 그에 해당되는 결과는 실제 나타나지 않는 경우들이다. 즉 공간 C2와 C3는 비어 있는 셈이다. 따라서 이제 <표 4.1>를 다음과 같이 바꿔 제시할 수 있다. 이때 주의할 점은 아래의 두 표는 형태적으로는 두 가지로 제시되어 있긴 하나 실은 동일한 하나의 표일 따름이라는 점이다. 이에 따라 어두운 색 영역은 그것이 따로 존재하지 않음을 보여 주며(그와 동일한 또 다른 영역으로써 이미 표현되었으므로), '×'표는 편의상 해당 공간이 비어 있음을 뜻한다.

<표 4.3>

	h가 참	h가 거짓
T가 h를 통과시킴	C1(=C4)	×
T가 h를 탈락시킴	×	

	h가 참	h가 거짓
T가 h를 통과시킴		×
T가 h를 탈락시킴	×	C4(=C1)

이렇게 되면, 먼저 위의 첫 번째 표가 보여 주듯, 확률 A는 (C1+0)/(C1+0+0)=1이 됨을 확인할 수 있다. 반면 위의 두 번째 표에 따라 P(시험 T가 가설 h를 통과시킴; h가 거짓)=0/(0+C4)=0이 되고, 이는 확률 B에 해당한다. 그러므로 또한 P(시험 T가 가설 h를 탈락시킴; h가 거짓)=C4/(0+C4)=1이 됨을 알 수 있다. 이로써 A가 1임에도 불구하고 B는 작으며(극단적으로 최소인 0), 따라서

시험 T의 엄격성은 최대가 됨을 다시 확인하게 된다. 이는, 앞서 언급한 대로 이세다가 구별한 해석들 가운데 (SC*-1)의 경우가 메이요에게 가장 적합함을 다시금 의미한다. 앞서 가설 $h(e_0)$에 대한 경우와 지금의 경우 사이의 유일한 차이란, 전자의 경우 그 가설과 증거 사이의 부합 관계가 통계적인 반면, 후자의 경우 그것은 논리적 함축 관계로서의 부합일 뿐이라는 점이다. 하지만 후자는 단지 전자의 특수한 한 경우일 뿐이다. 그러므로 이는 최종적으로 해석 (SC*-1) 하에서 메이요의 엄격성 판정에 아무런 비일관성도 없음을 말해 준다.

맺는말

이상으로, 엄격성 판정의 비일관성을 중심으로 한, 정동욱과 이세다의 메이요 비판에 대해 ‘증거-가설 생성의 메커니즘’ 개념을 통해 대응해 보았다. 그 일단의 책임이 메이요의 중요한 실수에 있음도 지적하였다. 그러한 실수 때나 그 이후 어쩌면 메이요 자신 단지 실수를 넘어 나와 같은 대응 방식을 생각하지 못했을 수도 있다. 만일 그렇다면, 이 논문에서의 나의 시도는 더욱 더 의의가 있을 것이다. 하지만 나는 메이요의 그것이 단지 실수였을 것이라 믿는다.

참고 문헌

전영삼 (2018), “우도와 증거-가설 생성의 메커니즘”,『과학철학』 21권 2호, pp. 1-45.

정동욱 (2018), “반사실적 베이즈주의 증거 이론”, 서울대학교 대학원 협동과정 과학사 및 과학철학 전공 이학박사학위논문.

Fitelson, B. (2006), "Logical Foundations of Evidential Support", *Philosophy of Science* 73, pp. 500-512.

Giere, R. N. (1983), "Testing Theoretical Hypotheses", in J. Earman (ed.), *Testing Scientific Theories*, *Minnesota Studies in the Philosophy of Science*, Vol. X, Minneapolis: Univ. of Minnesota Press, pp. 269-298.

Iseda, T. (1999), "Use-Novelty, Severity, and a Systematic Neglect of Relevant Alternatives", *Philosophy of Science* 66 (Proceedings), pp. S403-413.

Mayo, D. G. (1996), *Error and the Growth of Experimental Knowledge*, Chicago: The Univ. of Chicago Press.

Mayo, D. G. (2005), "Evidence as Passing Severe Tests: Highly Probable versus Highly Probed Hypotheses", in P. Achinstein (ed.), *Scientific Evidence: Philosophical Theories & Applications*, Baltimore: The Johns Hopkins Univ. Press, pp. 95-127.

Mayo, D. G. (2010), "An Ad Hoc Save of a Theory of Adhocness?: Exchanges with John Worrall", in D. G. Mayo & A. Spanos (eds.), *Error and Inference: Recent's Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, Cambridge:

Cambridge Univ. Press. pp. 155-169.

Mayo, D. G. (2014), "Some surprising facts about (the problem of) surprising facts (from the Dusseldorf Conference, February 2011)", *Studies in History and Philosophy of Science* 45, pp. 79-86.

Mayo, D. G. (2018), *Statistical Inference as Severe Testing: How to Get Beyond the Statistical Wars*, Cambridge: Cambridge Univ. Press.

Steel, D. (2007), "Bayesian Confirmation Theory and the Likelihood Principle", *Synthese* 156, pp. 53-77.

Worrall, J. (2010), "Error, Tests, and Theory Confirmation", in D. G. Mayo & A. Spanos (eds.), *Error and Inference: Recent's Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, op. cit., pp. 125-154.

고려대학교 문과대학 강사
College of Liberal Arts, Korea University
ysamchun@hanmail.net

Severe Tests and Mechanisms Generating an Evidence-Hypothesis

Young-Sam Chun

It seems certain that even if the same evidence is in itself given for any hypotheses, the way how it is obtained makes some differences in its support degree of them. In this respect, it is worth paying our attention to Mayo's conception of "severe test" and her technical development of it, which are just concerned with the procedures of getting evidence. Nonetheless, there have been criticisms against her theory from various respects. Among them, here this paper focuses on those especially raised by Jung (2018) and Iseda (1999). And it attempts to defend Mayo's theory on behalf of her against their critiques. For this purpose, the paper also proposes particularly a new concept of what is called the "mechanism generating an evidence-hypothesis". On the way, Mayo's own faults are revealed as well.

Key Words: Severe tests, Mechanisms Generating an evidence-hypothesis, Mayo, Jung, Iseda