

유죄결정 기준에 관한 실증과학적 접근*

김 상 준 안 정 호 박 광 배†
서울고등법원 부장판사 김앤장법률사무소 변호사 충북대학교

본 논문에서는 형사재판의 유무죄 판단기준(합리적 의심의 초월)을 질적, 양적으로 이해하기 위한 연구들을 중심으로 그 기준에 관한 실증과학적 관점을 고찰하고, 그 관점을 한국의 국민참여재판에서 나타나는 평결-판결 불일치의 의미를 해석하는데 적용하였다. 유무죄 판단기준에 관한 실증연구에서 두가지 함의가 도출되었는데, 첫째는 일반인들의 판단에 관한 실증과학적 이론과 지식에 기초하여 법관이 형사재판에서 판단하는 양식을 이해하는 것이 가능하다는 것이다. 둘째는 일반인이 법관보다 유죄판단을 위한 합리적 의심의 초월 기준을 다소 높게 설정하고, 그로 인하여 장기적으로 제1종 오류를 범할 가능성이 법관보다 더 낮을 가능성이 있는 반면, 법관은 장기적으로 제2종 오류를 범할 가능성이 일반인보다 더 낮을 가능성이 있다는 것이다. 이 함의들은 한국의 국민참여재판에서 발견되는 평결-판결 불일치의 양상과 일치하였다.

주제어 : 유죄결정 기준, 합리적 의심의 초월, 법관, 배심원

* 이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2013S1A5A2A03044871).

† 교신저자 : 박광배, 충청북도 청주시 개신동 충북대학교 심리학과, E-mail : kwangbai@chungbuk.ac.kr

재판이라는 제도가 발명되고 유지되는 한가지 이유는 불확실성 때문에 발생하는 갈등을 해소하고 사회에 예측가능성을 제공하기 위해서이고, 그 목적을 위해서 재판은 무엇이 실제적 진실(ground truth)인가, 좀 더 겸손하게는, 무엇이 법적 진실인가를 알아내는 장치(시스템)이다. 그러나 민사나 형사를 막론하고 거의 모든 재판에서 진실은 증거에 기초하여 추정될 수 있을 뿐 증명되지 않는다(Brook, 1982). 어떠한 증거 혹은 증거들의 집합도 진실에 관한 완벽한 확실성을 제공하지 않기 때문이다. 완벽한 증명력을 가지는 증거가 존재하지 않는다는 것은 재판에서 알아내는 사실이 진실이 아닐 가능성도 항상 존재한다는 뜻이다(Hammond, 1996). 재판에서 이루어지는 모든 판단에는 ‘더 이상 감소되지 않는 불확실성(irreducible uncertainty)’이 항상 내재하지만, 그럼에도 불구하고 재판의 결론은 언제나 확정적이라야 한다.

진실과 비진실이 공존하는 추정에 기초한 양단간의 판단, 연속적인 가능성을 범주로 변환하는 모든 판단에는 판단오류의 가능성이 불가피하게 상존하므로 판단오류를 최소화할 수 있는, 혹은 판단오류로 인하여 발생하는 해악을 최소화할 수 있는 합리적인 기준 혹은 판단역치를 필요로 한다(Newman, 1993). 더 구체적으로는, 판단기준이 ‘합리적(rational)’이기 위해서는 그 기준을 일관되게 사용하여 판단을 반복적으로 하는 경우, 불가피하게 발생하는 판단오류로 인한 손해 및 폐해가 최소화되어야 한다(Newman, 1993). 형사재판에서 사용하는 합리적 기준은 “합리적인 의심이 없는 정도”(Proof Beyond Reasonable Doubt)이다(형사소송법 제307조 제2항). 이 기준은 검사(정부)가 주장하는 바에 합리적 의심이 없을 때만

비로소 피고인의 유죄를 판단할 수 있다는 뜻이며, 법적 판단에 사용되는 모든 기준들 중에 가장 높은 수준의 증명력을 요구하는 기준이다(Newman, 1993; United States v. Fatico, 1978).

형사재판에서는 두종류의 판단오류가 생길 수 있다. 첫째는 무고한 피고인에 대하여 유죄판단을 하는 제1종 오류고, 둘째는 범죄자인 피고인에 대하여 무죄판단을 하는 제2종 오류다. 제1종 오류가 생기면 무고한 피고인의 명예와 인권이 심각하게 침해되고, 피고인이 마땅히 자신의 가족과 사회의 행복과 발전에 기여할 수 있는 기회가 박탈되며, 국가가 정당한 개인에게 부당한 고통을 초래하는 도덕적 과실범으로 전락하고, 정의가 회복되고 위험이 제거되었다는 거짓 환상에 안도하는 사회에서 진짜 범죄자(진범)가 자유롭게 활보하여 추가범죄의 위험이 높아지는 폐해가 발생한다. 반면, 제2종 오류가 생겨서 진범이 해방되면 기 발생한 범죄를 해결하기 위해 소비된 사회적 자원이 낭비되고 사회가 새로운 범죄위험에 노출된다.

형사재판에서 유죄판단을 할 수 있는 증명의 기준(standard of proof)을 합리적 의심의 여지가 없는 정도까지 최고로 높게 설정하는 이유는 앞서 설명한 바와 같이 불가피하게 발생하는 판단오류로 인한 손해 및 폐해를 최소화하기 위해서다. 제1종 오류의 해악과 제2종 오류의 해악은 대등하지 않기 때문에, 어떤 오류가 더 나쁜 오류인가의 문제는 한국사회에서 종종 관찰되는 피고인의 인권이 더 중요한가, 피해자의 인권이 더 중요한가의 피상적 논쟁의 대상이 아니다. 현대의 형사사법을 주도하는 중요한 이념 중 하나는 제1종 오류를 범하는 경우에 발생하는 해악이 제2종 오류를

범해서 발생하는 해악보다 크다는 것이다 (Addington v. Texas, 1979; In re Winship, 1970; Santosky v. Kramer, 1982). “열명의 범죄자를 놓치더라도 한명의 억울한 사람을 만들지 말아야 한다(It is better that ten guilty persons escape than that one innocent suffer)”는 William Blackstone (1765-1769)의 유명한 경구가 그 이념을 반영한다. 이 유명한 경구에서 10:1 이라는 특정 비율 보다는, 그것이 극단적으로 비대칭이라는 것이 중요한 골자다. ‘합리적 의심’ 기준은 형사재판에서 제2종 오류의 가능성이 증가하는 것을 무릅쓰고 제1종 오류를 최소화하기 위한 기준이며, 형사재판에서 이 기준의 사용은 헌법이 규정하는 적법절차(due process) 혹은 ‘정당한 법의 절차’ 개념의 핵심이다(In re Winship, 1970; Jackson v. Virginia, 1979; Newman, 1993).¹⁾ 다시 말해서, 현대사회가 형사재판의 증명력 역치를 합리적 의심 기준에 두는 가치는 헌법적 가치다. 따라서 현대의 형사재판에서 제2종 오류를 범할 것을 염려해서 합리적 의심을 초월하지 못하는 증거에 기초하여 피고인의 유죄를 판단하거나 양형을 하는 것은 바로 국가작용의 헌법적 요건인 적법절차를 국가가 비밀스럽게 파기하는 것이다. 국가의 헌법적 가치를 퇴색시키는 것은 제2종 오류를 실제로 범하는 것보다 훨씬 심각한 사회적 손실이다(In re Winship, 1970).²⁾

문명화된 현대사회의 형사사법에서 적법절

차의 핵심을 이루는 합리적 의심 기준은 어떤 기준인가? 합리적 의심이란 어떤 의심인가? 합리적 의심을 초월하는 믿음은 어떤 심리적 상태를 말하는 것인가? 심리학은 실증과학인 까닭에 이 의문들에 대하여 재판처럼 단정적이고 결론적인 해답을 제공하지는 않지만, 그 기준을 이해하는데 많은 기여를 하였다. 여기에서는 합리적 의심 기준을 질적, 양적으로 이해하고, 그 기준의 일관된 적용을 도모하기 위한 심리학 연구들을 중심으로 그 기준에 관한 실증과학적 관점을 분석해 보고자 한다.

질적 의미

합리적 의심 기준의 질적 의미에 대한 실증과학 연구들은 1990년대에 들어 그 기준을 어떻게 배심원들에게 이해시키고 설명해야 되는지에 관한 미국 법원의 위헌논쟁에 의해 촉발되었다. 그 기준의 질적 의미가 무엇인가의 문제는 법적 정의(헌법에 함치하는 정의)의 문제이기도 하다.³⁾

3) 한국의 판례법리에서 합리적 의심 개념이 유죄 판단 기준으로 형성되어온 경과는 김상준(2013)에 일별되어 있다. 한국에서 최초로 합리적 의심 기준이 형사재판의 유무죄 판단 기준으로 언급된 것은 1965년 6월 22일 대법원 판결(65도 370) 이지만, 그 기준의 정의가 최초로 제시된 것은 1997년 7월 25일 대법원 판결(97도974)로 알려졌다(김상준, 2013; 강동우, 2013). 대법원은 97도974 판결에서 「형사재판에 있어서 공소사실에 대한 거증책임은 검사에게 있는 것이고 유죄로 인정하기 위한 증거의 증명력은 논리와 경험칙에 따른 객관적이고 합리적인 증거평가의 결과 합리적인 의심을 배제할 정도의 확신을 가져올 수 있는 것이어야 하나, 여기에서 합리적인 의심이라 함은 모든 의문, 불신을 포함하는 것

1) “누구든지 …… 법률과 적법한 절차에 의하지 아니하고는 처벌·보안처분 또는 강제노역을 받지 아니한다.”(대한민국 헌법 제12조 제1항). “누구라도 정당한 법의 절차에 의하지 아니하고는 생명, 자유 또는 재산을 박탈당하지 아니한다.”(미국 수정헌법 제5조)

2) In re Winship, 397 U.S. 358(1970).

루이지애나 일심법원에서 일급살인죄로 기소된 피고인에게 사형이 선고되었다(Cage v. Louisiana, 1990). 일심법원의 유무죄 판단 단계에서 판사는 배심원들에게 합리적 의심 기준을 사용할 것을 실시하면서 ‘합리적 의심’을 “심대한 불확실성(grave uncertainty)을 일으키는 의심”과 “실재하고 실체적인(actual substantial) 의심”이라고 정의하였다. 이 재판은 결국 연방대법원까지 가게 되었는데, 연방대법원은 심대한 불확실성을 일으키는 의심과 실재하고 실체적인 의심은 형사피고인의 무죄를 결정하는데 요구되는 의심보다 훨씬 강한 의심을 암시하기 때문에 루이지애나 일심법원 판사의 설명이 합리적 의심 기준을 적절하게 표현하

이 아니라 논리와 경험칙에 기하여 요증사실과 양립할 수 없는 사실의 개연성에 대한 합리성 있는 의문을 의미한다고 판시하였다. 그 후, 대법원은 2004년 6월 25일 2004도2221 판결에서 그 정의를 더 구체적으로 표현하였는데, 「증거의 증명력은 법관의 자유판단에 맡겨져 있으나 그 판단은 논리와 경험칙에 합치하여야 하고, 형사재판에 있어서 유죄로 인정하기 위한 심증형성의 정도는 합리적인 의심을 할 여지가 없을 정도여야 하나, 이는 모든 가능한 의심을 배제할 정도에 이를 것까지 요구하는 것은 아니며, 증명력이 있는 것으로 인정되는 증거를 합리적인 근거가 없는 의심을 일으켜 이를 배척하는 것은 자유심증주의의 한계를 벗어나는 것으로 허용될 수 없다 할 것인바, 여기에서 말하는 합리적 의심이라 함은 모든 의문, 불신을 포함하는 것이 아니라 논리와 경험칙에 기하여 요증사실과 양립할 수 없는 사실의 개연성에 대한 합리성 있는 의문을 의미하는 것으로서, 피고인에게 유리한 정황을 사실인정과 관련하여 파악한 이성적 추론에 그 근거를 두어야 하는 것이므로 단순히 관념적인 의심이나 추상적인 가능성에 기초한 의심은 합리적 의심에 포함된다고 할 수 없다」고 판시하였다.

지 못했다고 판시하였다. 즉, 합리적 의심은 루이지애나 일심법원 판사가 정의한 것과 같은 강력한 의심을 의미하지 않기 때문에 결국 루이지애나 일심법원은 유죄판단을 위한 증명기준을 낮추는(i.e., 유죄 판단을 하기 쉽게 만드는) 잘못을 범했다는 것이다. 이후, Sullivan v. Louisiana(1993)재판에서도 합리적 의심에 대한 동일한 설명이 사용되었는데, 이 재판에 대해서도 연방대법원은 합리적 의심 기준에 대한 일심판사의 설명이 수정헌법 제5조(적법절차)와 제6조(배심원에 의한 재판을 받을 권리)를 위반했다고 판시하였다. 반면, Victor v. Nebraska(1994) 재판에서 합리적 의심을 “실재하고 실체적인 의심”으로 설명한 것에 대해서는 연방대법원이 합헌 판단을 하였는데, “실체적(substantial)”이라는 형용사가 질을 나타내는 단어고, 양을 나타내는 단어가 아니기 때문이고, Cage(1990)에서와는 달리 Victor(1994)에서는 실재하고 실체적인 의심이 단순가능성, 상상, 공상에 의한 의심과 비교되어 설명되었으며, Victor(1994)에서는 합리적 의심에 대한 또 다른 정의인 “행동을 주저하게 만드는 의심(doubt that causes one to hesitate to act)”이 함께 설명에 포함되었기 때문이다. Sandoval v. California(1994)재판에서는 판사가 합리적 의심을 “단순가능성에 불과한 의심이 아닌(not a mere possible doubt)” 의심으로 정의한 것이 무죄 판단을 위해 필요한 의심의 정도를 초과했다는 피고인의 주장을 연방대법원이 기각하였는데, 그 이유는 배심원들이 “단순가능성에 불과한 의심”을 상상해낸(imaginary) 의심 혹은 공상적인(fanciful) 의심 등의 의미로 이해할 것이지, 이성(reason)에 기초한 의심으로 이해하지는 않을 것이기 때문이다.

Victor(1994) 재판에서 Ginsburg 대법관은 별

도로 작성한 의견을 통해서 일심법원이 합리적 의심 기준에 관해 사용한 설명이 전체적으로 합헌이라는 다수의견에는 동의하였으나, 그 설명에 사용된 “행동을 주저하게 만드는”과 “양심적 확실성(moral certainty)” 등의 구체적 표현들에 대하여 문제를 제기하였다. 그러면서 Ginsburg 대법관은 연방사법센터(Federal Judicial Center: FJC)가 1987년에 제안한 설명을 합리적 의심 기준에 관한 가장 분명하고 정확한 설명으로 적시하였는데, FJC가 제안한 설명의 핵심은 “합리적 의심을 초월하는 증거이란 피고인의 유죄를 확고하게 납득케 하는 증거를 말한다(Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt)”는 것이다. FJC가 제안한 이 설명을 흔히 “확고한 납득(firmly convinced)” 설명이라고 부른다.

합리적 의심 기준의 의미와 설명방법에 대한 미국 연방대법원의 판시들에는 몇가지 경향들이 포착된다: (1) 유죄판단을 위해 필요한 증거의 증명력에 대하여 실질적 확실성(substantial certainty) 보다 낮은 수준의 확실성을 의미하는 설명은 대법원이 위헌판단을 할 가능성이 높다; (2) 그럼에도 불구하고 대법원은 비교되는 의미맥락이 충분히 함께 제공되는 경우에는 합리적 의심에 대해 모호한 용어와 표현이 사용된 정의에 대해서도 허용적이다; (3) 대법원은 합리적 의심 기준에 대한 판사의 설명을 판단 당사자인 배심원들이 어떻게 이해하고 받아들이는가에 대해서는 속단(e.g., 배심원들이 “단순가능성에 불과한 의심”을 상상해낸(imaginary) 의심 혹은 공상적인(fanciful) 의심 등의 의미로 이해할 것으로 단정)을 한다.

합리적 의심 기준에 관한 다양한 설명들을

듣는 사람이 미국 연방대법원이 단정하는 바와 같이 그 기준을 이해하는가? 이 실증적 질문에 답하기 위하여, Horowitz & Kirkpatrick (1996)은 검사측 증거가 강한 사건 시나리오와 그것이 약한 사건 시나리오를 이용하여 합리적 의심에 대한 각기 다른 정의에 대해 일반인들이 어떻게 다르게 반응하는지 알아보기 위한 실험을 하였다. 그 실험에서는 합리적 의심을 따로 정의하지 않는 경우(무정의)를 포함하여 다섯 가지의 정의가 사용되었다: (1) “확고한 납득(firmly convinced)”; (2) “양심적으로 확실(morally certain)”; (3) “긴가, 아닌가 하지 않는(not wavering or vacillating)”; (4) “이성과 상식, 그리고 공정한 고려에서 비롯된 ‘현실적 의심’이 없는(having no real doubt based upon reason, common sense and impartial consideration)”; (5) 무정의: “검사는 합리적 의심을 초월하는 정도로 배심원을 납득시켜야 한다(the prosecution must convince jurors beyond a reasonable doubt)”. 만약 합리적 의심 기준이 적절히 사용된다면, 검사측 증거가 강한 경우에 비해서 그것이 약한 경우에 피고인에 대한 유죄판단을 하는 사람이 현저히 적어야 한다. 이 실험에서 검사측 증거가 강할 때는 합리적 의심 기준이 어떻게 정의되건 유무죄판단에서 차이가 생기지 않았다. 그러나 검사측 증거가 약한 경우에는 그 차이가 나타났는데, “확고한 납득” 설명으로 정의되었을 때는 유죄판단이 전혀 없었고(0%), 나머지 정의들로 합리적 의심 기준이 설명된 경우에는 평균 47% 정도의 유죄판단 비율이 나타났다. 특히, 합리적 의심 기준에 관해서 별도의 정의가 제공되지 않은 무정의 조건과, ‘합리적 의심의 초월’을 피고인의 유죄에 대한 긴가, 아닌가하지 않고 안정된 믿음으로 정의한 조건에서는 증거강도에

다른 유무죄 판단 차이가 나타나지 않았다. 반면, 합리적 의심의 초월을 “확고한 납득”(FJC가 제안한 정의)으로 정의한 조건과 “현실적 의심이 없는”으로 정의한 조건에서는 증거 강도에 따른 유무죄 판단 차이가 크게 나타났다. 이 실험에서 참가자들은 또한 유죄판단을 할 수 있는 최소한의 유죄심증을 0%와 100% 사이에서 “피고인이 유죄일 확률”로 평정하였다. “확고한 납득” 설명을 들은 참가자들은 유죄판단을 위한 최소한의 유죄심증을 평균 81%로 평정한 반면, 다른 네가지 정의 조건의 참가자들은 그보다 훨씬 낮은 확률(평균 50%-64%)로 평정하였다.

Koch & Devine(1999)은 Horowitz & Kirkpatrick(1996)에서 나타난 FJC 정의(확고한 납득)의 우수성을 더 구체적으로 파악하기 위하여, 그 정의가, 무정의에 비해서, 가능한 처벌이 달라지는 경우에도 합리적 의심 기준을 일관되게 사용하는데 기여하는지를 알아보기 위한 실험을 하였다. 합리적 의심 기준이 FJC 정의로 설명된 조건에서는 유죄가 결정되면 피고인이 고의적 살인(murder)으로 처벌받을 것이라고 참가자들에게 알려준 경우와, 고의적 살인 혹은 과실치사(manslaughter)로 처벌받게 된다고 알려준 경우에 참가자들의 유무죄 판단에서 차이가 나타나지 않았다. 그러나 합리적 의심 기준에 대한 별도의 정의가 설명되지 않은 무정의 조건에서는 피고인이 살인 혹은 과실치사로 처벌받는다 알려준 경우의 유죄판단 비율(64%)이 살인으로 처벌받는다 알려준 경우의 유죄판단 비율(8%)보다 무려 8배 높게 나타났다. 이 결과는 합리적 의심 기준을 적절하게 정의하여 설명하지 않으면 특정 사건에서 가능한 처벌의 강도 혹은 피고인이 받는 혐의의 심각성에 따라 그 기준의 적용이 크게

가변할 수 있다는 것을 보여준다. 즉, 유죄가 선고되는 경우에 피고인이 받게 될 처벌이 심각할수록, 판단자는 유죄오판을 범하지 않으려는 동기를 강하게 가지게 되어 합리적 의심 기준을 더 엄격히 적용해서 판단하는 경향이 있다.

또한 Horowitz(1997)는 참가자들에게 범주적인 유무죄 판단을 하는 대신 피고인의 ‘유죄정도’를 1점부터 6점까지의 리커트 척도에 평정도록 하였다. 이 실험에서 합리적 의심 기준에 관한 FJC의 설명(확고한 납득)을 들은 참가자들의 ‘유죄정도’ 평정치가 다른 종류의 설명을 들은 참가자들의 평정치보다 가장 작은 변산(variability)을 보였다. 즉, 합리적 의심 기준을 FJC의 정의로 설명하였을 때 모든 사람들이 그 판단기준을 가장 유사하게 이해하고 균일하게 적용한다는 것이다.

Horowitz & Kirkpatrick(1996) 연구와 Koch & Devine(1999), 그리고 Horowitz(1997)의 연구는 합리적 의심 기준의 질적 의미에 대하여 두가지 중요한 함의를 제공하였다. 첫째는 ‘합리적 의심의 초월’ 개념이 고도로 추상적이어서 그 의미에 대한 이해와 적용이 사람들마다, 그리고 사건마다 현저히 달라질 수 있으므로 그 기준을 구체적이고 정확히 정의하는 것이 필요하다는 것이다. 둘째는 합리적 의심의 초월에 대하여 FJC가 제안한 정의인 “확고한 납득”이 그 기준의 법적 의미에 가장 근접하고, 그 기준을 사용하는 법이념(적법절차)을 가장 일관되게 구현할 가능성이 높다는 것이다.

형사재판에서 피고인의 유무죄 여부를 판단하는 사실판단자가 합리적 의심 기준을 어떻게 이해하는지 파악하는 또 다른 방법은 사실판단자가 그 기준을 다른 법적 판단기준들과 어떻게 다르게 사용하는지를 파악하는 것이다.

민사재판에서 사용되는 판단기준은 ‘증거의 비교우위(preponderance of evidence)’ 기준이다. 이 기준은 민사재판에서는 형사재판에서와는 달리 판단오류로 인한 비용/손실이 원고와 피고에게 동등하다는 가정에 기초한다(Santosky v. Kramer, 1982). 따라서 비교우위 기준은 원고(혹은 피고)의 주장이 상대방의 주장에 비하여 사실일 가능성이 조금이라도 더 높으면 원고(혹은 피고)의 승소를 판단하는 기준으로 이해된다(McCauliff, 1982; Simon, 1969; United States v. Fatico, 1978). 다시 말해서, 서로 상반된 주장을 하는 당사자들의 주장 중에서 사실일 확률이 50% 이상인 주장을 옳은 주장으로 판단한다(McCauliff, 1982; Simon, 1969; United States v. Fatico, 1978). 민사재판에서 사용되는 또 하나의 판단기준은 ‘명확하고 수긍되는(clear and convincing)’ 기준이다. 이 기준은 정부가 개인을 상대로 소를 제기하는(정부가 원고인) 민사사건 혹은 준형사사건(quasi-criminal case)에서 피고의 자유가 구속될 수 있거나, 피고에게 사회적 불명예 혹은 낙인이 생길 수 있을 때 사용된다. ‘명확한 수긍’ 기준은 원고(혹은 피고)의 주장에 대하여 비교우위 기준보다는 높고 합리적 의심의 초월 기준보다는 낮은 증명력을 요하는 기준인데, 원고(혹은 피고)의 주장이 사실일 확률이 최소한 70%-80% 이상일 때 원고(혹은 피고)의 주장이 옳다는 판단을 하는 기준이다(McCauliff, 1982; United States v. Fatico, 1978).

Dorothy K. Kagehiro는 일련의 실험연구들(Kagehiro & Rosen, 1986; Kagehiro & Stanton, 1985; Kagehiro, 1990)을 통해서 사람들이 ‘합리적 의심’ 기준과 ‘명확한 수긍’ 기준을 잘 구별하지 못한다는 것을 보여주었다. 이 실험들에서는 참가자들에게 세 종류의 판단기준(합리

적 의심, 명확한 수긍, 비교우위) 중 한가지를 알려주고 그 기준에 의해서 민사재판 시나리오의 피고에 대한 판단을 하도록 하였다. 전체적으로, 참가자들이 피고 승소를 결정한 비율이 합리적 의심 기준을 사용하도록 지시한 경우와 명확한 수긍 기준을 사용하도록 지시한 경우에서 차이가 없었다. 심지어 일부 실험에서는 합리적 의심 기준과 비교우위 기준 사이에도 차이가 나타나지 않았다. 그런데, Kagehiro & Stanton(1985)은 세번째 실험에서 세 가지 판단기준 각각을 다섯개의 다른 방법으로 설명해 주었는데, 합리적 의심 기준에 대해서는 FJC 설명(확고한 납득)을, 명확한 수긍 기준에 대해서는 캔사스 일심판사협회(Kansas District Judges' Association)가 만든 모범설시 설명(1997)을, 비교우위 기준에 대해서는 연방 모범설시 설명(Devitt & Blackmar, 1977)을 사용한 경우에서는 판단기준들 사이에 유의미한 차이가 나타났다.

합리적 의심 기준의 설명/해설 방법에 관한 미국연방대법원의 규범적 논의와 그 기준이 사실판단자의 인식체계에서 피고인의 유무죄 판단에 실제로 적용되는 양상에 관한 경험적 연구들을 종합하여, Stoffelmayr & Diamond(2000)는 합리적 의심 기준에 대한 설명이 필수적으로 갖추어야 할 네가지 요건을 명시하였는데, 이 네가지 요건들은 합리적 의심 기준의 질적 의미를 구성하는 내용이기도 하다.

1. 형사재판에서 피고인의 유죄를 판단하기 위해서 증거가 모든 의심을 초월하는 완벽한 확실성(absolute certainty)을 가져야 하는 것은 아니다(Gilbert, 1756; Morano, 1975).
2. 유죄판단을 하기 위해서는 사건의 모

든 핵심적 요소(“each and every essential element”)에 대한 증거가 높은 수준의 역치를 능가하는 증명력을 지녀야 한다(In re Winship, 1970; Sullivan v. Louisiana, 1993).

3. 합리적 의심 기준이 그보다 낮은 기준들인 ‘증거의 비교우위(preponderance of evidence)’ 기준, ‘명확한 수궁(clear and convincing)’ 기준, 등과 명시적으로, 명확히 구별되어야 한다.

4. 합리적 의심의 의미가 사실판단자의 도덕관념과 개인적 경험에 따라 달리 해석/적용되어서는 안된다.(i.e., ‘양심적 확실성(moral certainty)’ 등의 표현으로 정의되어서는 안 된다).

Stoffelmayr & Diamond(2000)는 합리적 의심 기준의 네가지 의미요소를 갖춘 설명으로 미국연방사법센터(Federal Judicial Center: FJC)가 1987년에 제안한 설명을 적시하였다. 그러나 Stoffelmayr & Diamond(2000)는 FJC의 설명에 다섯번째 요건을 추가할 것을 제안하였는데, 그들이 스스로 논란이 될 수 있다는 단서를 붙이면서 제안한 다섯번째 요건은 “판단오류에 수반되는 손실/비용에 따라서 합리적 의심 기준의 수위를 적절하게 조절할 수 있는 여지(room for flexible tailoring of the standard to the costs of error)가 있어야 한다”는 것이다. Stoffelmayr & Diamond(2000)는 합리적 의심 기준이 모든 형사사건에서 균일한 의미로 적용되는 것은 불합리한데, 그 이유는 피고인이 유죄 선고를 받는 경우 피고인에게 과해지는 처벌의 심각성이 사건마다 달라서 유죄 선고가 오류인 경우에 발생하는 비용(손실)도 사건마다 다르기 때문이다. 앞서 설명한 바와 같이 형사재판에서 유죄판단을 할 수 있는 증명

력의 역치를 합리적 의심의 여지가 없는 정도까지 높게 설정하는 이유는 불가피하게 발생하는 판단오류로 인한 손해 및 폐해를 최소화하기 위해서다. 피고인이 중범죄로 기소된 사건에서 유죄오판이 이루어지는 경우에는 피고인이 부당하게 감수해야할 비용이 매우 크다. 특히 사형이 가능한 재판에서의 유죄오판은 그 부당한 비용의 규모를 가늠하기 어렵다. 법에 의해 허용가능한 처벌이 심각할수록 유죄오판이 발생해서는 더욱 안되므로, 중범죄로 기소된 피고인에게는 합리적 의심 기준이 특별히 더 엄격하게 적용될 때 그 기준의 기본적인 기능과 헌법이념이 더 잘 달성될 수 있다는 것이다. 이러한 관점에서 Stoffelmayr & Diamond(2000)는 배심재판의 경우, 해당 사건에서 피고인이 유죄선고를 받는 경우에 가능한 처벌이 무엇인지를 판사가 배심원들에게 사전에 미리 알려줄 것을 함께 제안하였다.⁴⁾

양적 의미

법이 의도하는 합리적 의심 기준의 정확한 정의가 무엇인지 규명하기 어렵고, 그 때문에 사실판단자에 따라서, 그리고 사건의 내용에 따라서 그 의미가 달리 해석/적용될 가능성이 높기 때문에, 그 기준을 숫자로 표현되는 확률개념으로 정의하는 것이 바람직하다는 의견들(e.g., Kagehiro & Stanton, 1985)이 나타났다. 일부 판사들은 합리적 의심을 초월하는 믿음

4) 이 제안은 1994년의 ‘더 효과적인 배심단 사용에 관한 아리조나주 대법원 위원회’의 보고서에서 제안된 바 있다(Arizona Supreme Court Committee on the More Effective Use of Juries(1994) Jurors: The Power of 12).

을 “사실일 확률이 90%-95% 인 믿음”으로 해석하기도 하였다(McCauliff, 1982; Simon, 1969; *United States v. Fatico*, 1978). 그러나 합리적 의심 기준을 그와 같이 양적으로 정의하는 것에 반대하는 견해도 함께 나타났는데, *McCullough v. State*(1983) 재판에서 네바다주 대법원은 일심법원의 판사가 배심원에게 0부터 10까지의 척도에서 증거의 비교우위 기준은 5에 해당하고 합리적 의심 기준은 7.5에 해당한다고 설명한 것이 잘못되었다는 이유로 일심판결을 파기하였다. 그러한 설명이 잘못된 이유는 각 기준에 해당하는 숫자의 크기 때문이 아니라, 법적 판단기준들은 본질에 있어 질적인 개념이므로 양적으로 정의할 수 없기 때문이라는 것이다.

법적인 유무죄 판단은 질적, 범주적인 양단간의 판단이다. 그러나 모든 디지털 신호가 아날로그 신호로 표상될 수 있고, 그 반대도 가능하듯이, 질적이고 범주적인 판단은 양적이고 연속적인 판단으로 개념화될 수 있을 것이다. 법적인 ‘유죄’ 개념을 양적으로 정의하는 한 방법은 ‘검사측 증거 혹은 기소사실에 의거해 피고인이 유죄일 사후(조건)확률 (posterior probability)’을 사용하는 것이다(Dawid, 2005). 그 사후확률은 모든 증거를 고려한 후에 피고인이 유죄일 조건확률인데, 0과 1 사이의 범위 내에서 숫자로 정의되는 $P(G|E)$ 로 표기할 수 있다(P =확률, G =유죄, E =증거).

사후확률 $P(G|E)$ 은 주관적 확률이며, 검사측 증거 혹은 증명된 사실을 전제로 피고인의 유죄에 대한 믿음의 정도(the degree of belief)로 해석된다. 사후확률 혹은 사후믿음을 수학적으로 정의한 것이 베이즈 공리(Bayes's theorem)다. 베이즈 공리에 의해 $P(G|E)$ 를 추정하기 위해서는 다음과 같은 정보가 필요하다: (1)

$P(E|G)$. 피고인이 실제로 유죄일 때 검사가 제시하는 증거가 나타날 확률(다시 말해서, 실제로 유죄인 피고인을 검사가 기소할 확률). 검사기소와 검사측 증거의 정확성을 반영하는 이 확률을 ‘민감성(sensitivity)’이라고 부른다; (2) $P(E|N)$. 피고인이 실제로 무죄일 때 검사의 증거가 나타날 확률(실제로 무죄인 피고인을 검사가 기소할 확률). 검사기소와 검사측 증거의 부정확성을 반영하는 이 확률을 ‘비특정성(non-specificity)’이라고 부른다; (3) $P(G)$. 피고인의 유죄에 대한 증거 전 믿음(검사측 증거를 보기 전 피고인의 유죄에 대한 선형적 믿음). 판단자의 주관적인 이 선형적 믿음을 ‘사전확률(prior probability)’이라고 부른다(혹은 ‘기저율(baserate)’로 지칭되기도 한다). 이 정보들을 이용하여 ‘검사측 증거 혹은 기소사실에 의거해 피고인이 유죄일 사후(조건)확률’ $P(G|E)$ 는 다음과 같이 정의된다.

$$P(G|E) = \frac{P(E|G)P(G)}{P(E|G)P(G) + P(E|N)[1 - P(G)]}$$

검사증거(기소)의 민감성과 비특정성의 비율 $P(E|G)/P(E|N)$ 을 ‘진단성(diagnosticity)’이라고 부른다. 즉, 검사의 증거 혹은 기소사실이 피고인의 유죄 여부를 정확히 진단하는 정도다. 민감성이 비특정성에 비해 클수록 검사측 증거의 유죄진단 정확성이 높다. 요약하면, 베이즈 공리는 피고인의 유죄에 대한 믿음의 정도가, ① 검사가 제시하는 증거의 진단성과, ② 사전믿음에 의해 결정된다는 것을 암시한다.

형사사법의 가장 중요한 원칙 중의 하나인 무죄추정원칙(presumption of innocence)을 확률 개념으로 설명한다면, 재판 전에 유죄 사전확률 $P(G)$ 이 무죄 사전확률 $(1-P(G))$ 에 비해서

작아야 한다는 원칙(i.e., $P(G) < .5$)이다. 만약 검사측 증거의 진단성이 100 이라면(즉, 검사증거의 민감성이 비특정정보보다 100배 크다면), $P(G)$ 가 무죄추정원칙이 허용하는 최고치에 근접한 0.49일 때, $P(G|E)$ 는 0.99가 된다. 반면, 만약 증거진단성이 2라면, $P(G)$ 가 0.49일 때, $P(G|E)$ 는 0.66에 불과하다. 위의 베이즈 공리는 판단자의 마음 속에서 재판 전에는 작았던 $P(G)$ 가 재판 중 제시되는 증거들에 의해 $P(G|E)$ 로 수정되는 즉, 유죄심증이 증거에 의해 개정되는 원리를 수학적으로 정의한 규범 이론(normative theory)이다. 실제 사실판단자들의 유죄심증 형성이 베이즈 공리에 부합하는 규범적이고 합리적인 방식으로 이루어진다는 증거는 거의 없다(Hastie, 1994; Hastie, Penrod, & Pennington, 1983). 사실판단자들이 유죄심증을 형성할 때 증거의 진단성 혹은 사전확률을 도외시하기 때문이 아니라, 무수히 많은 다른 요인들이 판단에 개입하기 때문이다(Pennington & Hastie, 1992).

베이즈 공리는 유죄확률 $P(G|E)$ 를 산출하는 규범적인 방법일 뿐, 유죄확률을 산출하는 비규범적인 방법은 무수히 많다. 사실판단자들은 비록 베이즈 공리를 사용하지는 않지만 다른 방법에 의해 유죄확률을 산출하고 그것을 확률역치와 비교하여 유무죄 판단을 할 수 있다. 따라서 사실판단자들이 추정하는 유죄확률이 베이즈 공리로 계산되는 유죄확률과 거의 언제나 다르지만, 그럼에도 불구하고 실증 연구들은 합리적 의심 기준을 유죄확률로 정의하는 것의 장점을 보여주었다. 그 장점의 한가지는 형법과 민법에서 각기 다른 판단기준들이 사용되는 법의도가 더 잘 구현된다는 것이고, 두번째는 형사재판의 유무죄 판단에서 제 1중 오류 확률(무고한 사람을 처벌하는

확률)이 더 줄어들 가능성이 있다는 것이다.

Kagehiro & Stanton(1985)은 세집단의 실험참가자들에게 ‘합리적 의심’ 기준, ‘명확한 수궁’ 기준, ‘증거의 비교우위’ 기준을 원고의 주장이 사실일 확률이 91%, 71%, 51%인 기준이라고 알려준 후, 재판시나리오에 대하여 세기준 중 하나를 사용하여 판단하게 하였다. 그 결과, 각 기준을 사용한 집단에서 예측에 부합하는 방향으로 무죄판단 비율의 차이(‘합리적 의심’ 기준 > ‘명확한 수궁’ 기준 > ‘증거의 비교우위’ 기준)가 명확히 나타났다. Kagehiro & Stanton(1985)은 판단기준이 엄격해질수록 무죄판단이 체계적으로 증가하는 것으로 나타난 연구결과에 기초하여, 판단기준들에 대한 법적(질적) 정의보다 양적 정의가 그 판단기준들을 사용하는 법의 의도(적법절차)를 더 잘 구현한다고 주장하였다.

Kerr et al.(1976)은 배심원들에게 합리적 의심 기준을 양적으로 설명하면 질적으로 설명하는 것보다 유죄판단을 하는 배심원의 수가 줄어든다는 것을 발견하였다. Nagel(1979)도 또한 합리적 의심 기준에 관한 양적인 설명이 질적인 설명에 비하여 그 기준의 엄격성을 강화한다는 것을 발견하였고, 동일한 결과가 Kagehiro & Stanton(1985)에서도 관찰되었다. 합리적 의심 기준을 양적으로 개념화하는 것의 심리적 효과에 관한 이 연구결과들은, 양적 차원에서 확률(probability) 혹은 가능성(likelihood)을 의식적으로 판단하는 것이 질적 차원에서 범주적으로 판단하는 것에 인지적 변화를 초래한다는 것을 암시한다. Simon & Mahan(1971)은 미국 일리노이 주에서 실제 재판에 참여하는 69명의 배심원들과 88명의 대학생들을 대상으로 합리적 의심 기준에 해당하는 확률 역치를 파악하기 위한 연구를 수행

하는데, 이 연구에서 유무죄를 ‘유죄확률’에 기초하여 양적으로 판단하는 것이 초래하는 심리적 효과에 관하여 매우 흥미로운 암시가 한가지 도출되었다. 배심원들과 대학생들은 피고인이 살인혐의로 기소된 재판의 녹음을 청취한 후, 피고인의 유무죄 여부를 범주적으로 판단하고, 피고인이 기소된 범행을 실제로 했을(피고인이 유죄일) 확률을 .5 단위로 나누어진 0-10 리커트 척도에 평정하였다. 그런데, 참가자들의 절반은 범주적인 유무죄 판단을 먼저 하고 유죄확률 판단을 나중에 한 반면, 다른 절반은 그 순서를 바꾸어서 유죄확률 판단을 먼저 하고 범주적인 유무죄 판단을 나중에 하였다. 범주적인 유무죄 판단을 먼저 한 경우(즉, 유죄확률에 대한 의식적 판단 없이 유무죄 여부를 범주적으로 판단한 경우)에는 대학생들의 20%가 유죄판단을 하였고 배심원들의 51%가 유죄판단을 하였다. 반면, 유죄확률 판단을 먼저 하고 범주적인 유무죄 판단을 나중에 한 경우에는 대학생들의 9%와 배심원들의 38%가 각각 유죄판단을 하였다. 즉, 피고인이 유죄일 확률에 대한 의식적 판단을 하므로써 유죄 판단의 기준 혹은 역치가 더 상승한다는 것이 발견되었다. 만약 Kerr et al. (1976), Nagel(1979), Kagehiro & Stanton(1985), Simon & Mahan(1971) 등의 연구에서 발견된 현상이 확고한 것이라면, 합리적 의심 기준을 확률개념으로 정의하여 형사피고인에 대한 유무죄 판단에 적용하면 장기적으로 제 1종 오류(무고한 사람을 처벌하는 오류)를 줄이는 효과가 생길 수 있다.

형법과 민법에서 각기 다른 판단기준들을 사용하는 법의 의도가 더 잘 구현되고, 형사피고인에 대한 유무죄 판단에서 제 1종 오류가 감소할 수 있다면, 합리적 의심 기준을 양

적인 확률개념으로 정의하는 것의 타당성이 있다. 그렇다면, 합리적 의심을 초월하는 주관적 유죄확률 $P(G|E)$ 은 얼마인가? 다시 말해서, $P(G|E)$ 가 판단자의 마음 속에서 어떻게 형성되는가는 불문하고, 그 주관적 확률이 얼마 이상이면 피고인의 유죄가 합리적 의심을 초월하는 것인가? 이 물음에 답을 구하기 위하여 McCauliff(1982)는 미국내 모든 판사들을 대상으로 ‘합리적 의심의 초월’에 해당하는 ‘확실성(certainty)’이 0%와 100%의 사이에서 어디에 해당하는지를 질문하였다. 이 질문에 대하여 195명의 판사들이 답변하였고, 그중 171명의 판사들이 답변을 확률(비율)로 제시하였는데, 최빈치(mode)와 평균(mean)이 공히 90%였다.

Simon(1969)은 판사들에게 다양한 범죄혐의(살인, 횡령, 강간, 중절도, 경절도, 사기)로 기소된 피고인들을 상상하도록 요구하고, “각 피고인에게 유죄를 선고하려면 피고인이 범행을 했을 확률이 최소한 얼마라야 하는지”를 질문하였다. 그 결과, 판사들이 유죄선고를 위한 $P(G|E)$ 의 역치를 범죄에 따라 달리 평가하는 것이 발견되었는데, 좀도둑(petty larceny)의 경우는 그 역치를 .87 정도로 평가하고, 살인(murder)의 경우는 그것을 .92 정도로 평가하였다. 즉, 유죄선고를 위한 $P(G|E)$ 의 역치에 관하여 판사들의 감각은 전체적으로 .88 정도였으나, 피고인의 혐의가 중할수록 그 역치를 다소 높게 평가하였다.

Simon & Mahan(1971)은 미국 일리노이 주에서 실제 재판에 참여하는 69명의 배심원들과 88명의 대학생들을 대상으로 유죄판단을 위한 확률 역치를 파악하기 위한 연구를 수행하였다. 배심원들과 대학생들은 피고인이 살인혐의로 기소된 재판의 녹음을 청취한 후, 피고

인의 유무죄 여부를 범주적으로 판단하고, 피고인이 기소된 범행을 실제로 했을 확률을 .5 단위로 나누어진 0-10 리커트 척도에 평정하였다. 참가자들은 유무죄 판단과 확률 판단을 한 후, 최종적으로 Simon(1969)에서 사용했던 확률 역치 질문 “각 피고인에게 유죄를 선고하려면 피고인이 범행을 했을 확률이 최소한 얼마라야 하는지”에 응답하였다. 이 연구에서 배심원들과 대학생들도 앞서 연구의 판사들(살인에 대해서 .92)과 유사하게 유죄판단을 위한 $P(G|E)$ 의 역치를 .86(배심원)과 .91(대학생)로 평가하였다.

Martin & Schum(1987)은 대학생들에게 피고인이 백인남성이고, 기소배심에 의해 기소되었으며, 유죄가 확정되는 경우 가능한 처벌(양형)이 무엇인지 알려준 후, “유죄판단을 하기 위해서는 피고인이 유죄일 가능성이 무죄일 가능성의 몇배라야 하는지”의 승산(odds) 판단을 하게 하였다. 그 유죄승산에서 참가자들은 살인을 제외한 범죄들에 대해서는 평균적으로 25:1 을, 그리고 피고인이 유죄선고를 받으면 무기징역 혹은 사형이 양형되는 살인의 경우는 100:1 을 판단하였다. 즉, 참가자들은 살인을 제외한 여타 범죄혐의로 기소된 피고인에게 유죄를 선고하기 위해서는 유죄일 가능성이 무죄일 가능성보다 최소 25배 더 커야 하고, 살인혐의를 받는 피고인에 대해서는 그것이 최소 100배 더 커야 한다고 판단하였다. 유죄판단을 위해 요구되는 승산을 베이즈 공리에 대입하여 유죄판단을 위한 $P(G|E)$ 의 역치를 산출하면, 살인을 제외한 범죄들에 대해서는 평균 .96이 산출되고, 살인의 경우는 .99가 산출된다.

최근에 Dhami(2008)는 새로운 방법으로 합리적 의심 기준의 양적 정의를 시도하였다.

배심원 자격이 있는 203명의 대학생들에게 형사재판의 무죄추정의 원칙, 입증책임, 합리적 의심의 초월 기준에 관한 판사의 설명을 읽도록 하였다. 그 후, 대학생들은 합리적 의심 기준에 해당하는 유죄증명의 정도를 측정하기 위한 문항에 응답하였다. 예를 들어, 응답자는 검사의 유죄증명 0%는 합리적 의심의 초월에 속하는 정도가 “0(전혀 아니다)” 이고, 유죄증명 10%는 “1”이며, ……………, 유죄증명 80%는 “18”이며, 유죄증명 90%는 “20(완전히 그렇다)”이라고 평가하였다. ‘소속함수방법(membership function method)’으로 명명된 이 측정방법에서는 ‘완전히 그렇다(20)’가 선택된 최소 유죄증명 비율(%)이 합리적 의심 기준에 대한 양적 평가로 간주되는데, 대학생들이 평가한 그 비율의 평균이 96% 였다. 즉, 대학생들은 합리적 의심 기준에 해당하는 유죄증명의 정도를 96%로 해석하였다.

합리적 의심 기준을 양적으로 정의하기 위한 연구들은 일반적으로 그 기준이 사후확률 $P(G|E)$ 에서 .80 ~ .99 정도에 해당하는 역치인데, 유죄가 확정되는 경우에 피고인이 받게 될 처벌이 심각할수록 그 역치가 높아지는 것으로 파악한다. 그런데, ‘가능한 처벌의 심각성’ 이외에도 그 역치의 변산을 초래하는 또 하나의 요인이 알려졌는데, 역치를 알아내는 방법에 따라서 파악된 역치가 크게 달라질 수 있다. 합리적 의심 기준에 해당하는 $P(G|E)$ 역치를 알아내는 방법에는 직접적 방법과 간접적 방법이 있다. 직접적 방법은 피고인이 기소된 범행을 실제로 했을 확률이 최소한 얼마일 때 유죄 판단을 할 수 있는지를 직접 응답자에게 질문하고, 응답자는 그 최소 확률을 0 과 1 사이(혹은 0%와 100% 사이)의 한 지점으로 반응하는 방법이다. 간접적 방법은 ‘결정이

론'(decision theory)을 활용하는 것이다(Tribe, 1971).

여러 분야의 사회과학에서 인간의 판단과 결정을 개념화할 때 광범위하게 활용되는 결정이론은 결정대안들이 초래하는 가능한 결과들의 효용가치를 따져서 이익이 최대가 되고 비용이 최소가 되는 결정대안이 선택된다는 이론이다. 형사재판의 유무죄 판단은 네 가지의 결과를 초래할 수 있다: (1) 무고한 사람을 무죄라고 옳게 판단; (2) 유죄인 사람을 무죄라고 잘못 판단; (3) 무고한 사람을 유죄라고 잘못 판단; (4) 유죄인 사람을 유죄라고 옳게 판단. 네 가지 결과들은 각기 이익과 비용을 초래한다. 합리적 의심 기준에 해당하는 P(G|E) 역치를 알아내는 간접적 방법은 네 가지 판단 결과들 각각의 효용가치(utility)를 응답자들이 제시토록 한 후, 그 효용가치들을 결정이론에 대입해서 역치를 알아내는 방법이다(Fried, Kaplan, & Klein, 1975).⁵⁾ 합리적 의심

5) 결정이론에 의하면, P(G|E)에서 유죄판단의 역치를 파악하기 위해서는 네가지 판단 결과들의 효용가치(utility)에 관한 정보가 필요하다: (1) U(AI). 무고한(innocent) 사람의 무죄(acquittal)를 옳게 판단하는 효용가치(이익); (2) U(AG). 유죄인(guilty) 사람의 무죄(acquittal)를 잘못 판단하는 효용가치(비용); (3) U(CI). 무고한(innocent) 사람의 유죄(convict)를 잘못 판단하는 효용가치(비용); (4) U(CG). 유죄인(guilty) 사람의 유죄(convict)를 옳게 판단하는 효용가치(이익). 이 효용가치들은 판단자의 주관적인 가치다. Fried et al.(1975)은 판단자의 네가지 주관적인 가치를 반영하는 판단역치(decision criterion: DC)를 다음과 같이 정의하였다.

$$DC = \frac{U(AI) - U(CI)}{U(AI) - U(CI) + U(CG) - U(AG)}$$

예를 들어, “열명의 범죄자를 놓치더라도 한명의 억울한 사람을 만들지 말아야 한다”는

기준에 해당하는 P(G|E)역치를 결정이론으로 파악하면 직접적 방법으로 파악하는 것보다 현저히 낮은 역치(.50 ~.60)가 파악된다(Nagel, Lamm, & Neef, 1981; Nagel, 1979; MacCoun, 1984; Thomson, Cowan, Ellsworth, & Harrington, 1984; Dane, 1985; MacCoun & Kerr, 1988; MacCoun & Tyler, 1988). 그 가능한 이유 중의 하나는 유무죄 판단이 초래하는 결과들(판단의 옳음과 틀림)의 효용가치(이익과 비용)에 대한 인식이 명확하지 않고, 그 효용가치들이 결정이론이 가정하는 합리적 방식으로 취합되지 않기 때문이다. 즉, 결정이론도 베이즈공리와 마찬가지로 규범적 이론이며 사람들이 실제로 그 이론에 부합하는 방식으로 합리적 결정을 한다는 증거가 거의 없다.

유죄판단을 위한 P(G|E)역치를 간접적 방법으로 파악하면 현저히 낮은 역치가 산출되는 또 다른 이유는 그 방법이 판단자들로 하여금 제1종 오류(무죄인 사람을 유죄라고 잘못 판단하는 오류)와 제2종 오류(유죄인 피고인을 무죄라고 잘못 판단하는 오류)의 효용가치를 인지적으로 평가토록 강요하므로, 그 오류들에 대한 가치평가를 판단자들이 의식적으로 균형화하는 부작용을 가지기 때문일 수 있다. 즉, 결정이론을 이용해서 P(G|E)역치를 간접적으로 파악하는 방법은 제1종 오류와 제2종 오류의 비용을 공평하게 분배하도록 응답자들을 유도하는 요구특성(demand characteristic)을 가질 수 있다. 이 가능성은 실증연구를 통해서 확인될 필요가 있다.

합리적 의심 기준을 양적으로 정의하기 위

William Blackstone(1765)의 유명한 경구가 U(AI)=0, U(CI)=10, U(CG)=0, U(AG)=1의 효용가치를 의미한다면, Blackstone 판사의 유죄판단 역치(DC)는 .91이다.

한 직접적 방법과 간접적 방법이 공히 가지는 또 다른 문제는 그 방법들에 의해 파악된 확률 역치의 정확한 의미가 불분명하다는 것이다. 직접적 방법을 사용한 대부분의 연구들은 참가자들에게 ‘피고인의 유죄를 판단할 수 있는 증명력 수준’을 제시토록 하였다. 간접적 방법도 참가자들이 제시하는 효용가치들을 이용하여 참가자들이 피고인의 유죄를 판단할 때 사용하는 증명력 역치를 도출한다. 그런데 참가자가 생각하는 ‘피고인의 유죄를 판단할 수 있는 증명력 수준’은 ‘합리적 의심 기준에 해당하는 증명력 수준’과 동일하지 않을 수 있다. 최근에 Dhami(2008)에 의해 도입된 소속 함수방법은 참가자들이 ‘합리적 의심 기준에 속하는 검사의 유죄증명 정도’를 평가하기 때문에 직접적 방법과 간접적 방법이 가지는 의미의 모호성이 덜하다는 장점이 있다. 각 방법들이 도출하는 역치의 정확한 의미를 이해하기 위한 세심한 연구들이 요구된다.

법관과 배심원의 기준

Kalven & Zeisel(1966)은 1950 년대에 미국의 판사들을 대상으로 형사재판에 관한 대규모 조사연구를 수행하였다. 그 조사에서 555명의 판사들이 3,576개의 형사재판에 관한 정보를 제공하였는데, 그중에는 각 재판의 배심원 평결과 판사의 판단(만약 판사 본인이 유무죄 여부를 판결하였다면 어떤 판결을 했을지)이 포함되었다. 이 연구에서 배심원 평결과 판사의 판단이 78% 일치하는 것으로 나타났다. 두 판단이 불일치한 사건들 중에 86%는 배심원이 무죄를 평결하고 판사는 유죄를 판단한 사건이고, 14%는 배심원이 유죄, 판사가 무죄를

판단한 사건으로 나타났다. Kalven & Zeisel (1966)은 배심원과 판사의 판단이 불일치하는 이유를 배심원들이 증거나 사건내용을 이해하지 못하였기 때문이 아니라 합리적 의심의 초월 기준에 대한 해석, 증거평가 방식, 피고인에 대한 감정, 그리고 적용법률에 대한 수긍(인정)에서 판사들과 차이가 있기 때문인 것으로 해석하였다(Hans & Vidmar, 1991).

최근에 Eisenberg, Hannaford-Agor, Hans, Waters, Munsterman, Schwab, & Wells(2005)은 미국 국립주법원센터(National Center for State Courts: NCSC)의 데이터베이스를 이용하여 Kalven & Zeisel(1966)의 결과를 재검증하였다. NCSC 데이터베이스는 과거 재판에 대한 판사들의 응답자료뿐만 아니라 배심원과 변호인, 검사의 응답자료도 함께 수록하고 있어서 배심원-판사의 판단일치도에 관하여 40여년 전의 Kalven & Zeisel(1966)의 연구보다 더 신뢰로운 결과가 기대되었는데, 전반적으로 Kalven & Zeisel(1966)에서 나타난 배심원-판사 일치비율과 거의 유사한 75%의 일치율이 나타났다. 판사들이 평가한 재판의 ‘증거강도’를 기준으로 사건들을 구분하였을 때, 유죄증거가 약했던 사건에서는 판사들이 배심원들보다 무죄판단을 더 많이 한 반면, 증거강도가 중간으로 평가되거나 강한 것으로 평가된 사건에서는 판사들이 배심원들보다 유죄판단을 더 많이 하였다. 배심원과 판사의 판단이 서로 불일치한 사건들에서 그 불일치의 이유에 대하여, Kalven & Zeisel(1966)과 마찬가지로, Eigenberg et al.(2005)은 증거나 적용법률의 복잡성이 아닌 것으로 해석하였다.

Kalven & Zeisel(1966)과 Eigenberg et al.(2005)의 연구는 비록 재판이 종료된 후 재판참여자들의 회고적 응답에 기초한 것이지만 실제 재

판에서 배심원과 판사의 판단일치도를 확인한 가장 대규모의 대표적인 연구들인데, 배심원과 판사의 판단이 4개 중 3개가 일치한다는 공통된 결과를 제시하였다. 그외에도 배심원(일반인)과 판사가 판단능력, 판단의 정확성, 편향과 착오에 대한 취약성 등에서 차이가 없다는 것을 보여주는 많은 조사/실험 연구들이 이루어졌는데(McCoun, 2005; Robbennolt, 2005; Alexander & Sherwin, 2008), 사건내용과 증거가 복잡한 사건에서도 판사와 배심원의 판단정확성에 차이가 없고(Lempert, 1981), 의사결정의 중요한 요소들(정보의 선별, 중요도 판단, 취합 등)에서 판사와 일반인의 차이가 없으며(Robbennolt, Groscup, & Penrod, 2006), 증거능력이 없는 정보를 유무죄 판단에서 배제하거나 배제하지 못하는 정도에서도 판사와 일반인이 대동소이하고(Landsman & Rakos, 1994), 판단과 의사결정에 개입하는 것으로 널리 알려진 각종 인지적 편향과 휴리스틱에 취약한 정도에서도 차이가 없고(Guthrie, Rachlinski, & Wistrich, 2001), 통계적 개념과 수치, 그리고 과학적 증거를 이해하고 적용할 때 경험하는 어려움도 일반인과 판사가 비슷하고(Wells, 1992; Kovera & McAuliff, 2000; Kovera, McAuliff, & Herbert, 1999), 공판이 20일 이상 연속적으로 속개되는 복잡한 민사재판의 증거에 대한 이해, 기억, 판단에서도 판사와 배심원 사이에 차이가 발견되지 않는다(Cecil, Hans, & Wiggins, 1991). 특히, 배심원들도 판사들처럼 사건/수사기록을 수시로 재확인할 수 있고, 공판 중에 메모를 할 수 있으며, 증인과 피고인에게 직접 질문을 할 수 있게 하여 배심원 판단과 판사 판단의 절차적 불평등 요소들을 제거하면, 그들 사이의 정확성 차이는 거의 모두 사라진다(FosterLee, Horowitz, & Bourgeois, 1993; Hastie,

1983; Penrod & Heuer, 1998).

한국의 판사와 일반인의 판단도 대동소이할까? 한국에서 2008년 1월부터 형사재판에 배심원이 참여하여 피고인의 유무죄 여부와 양형을 재판부(판사)에 권고하는 국민참여재판이 도입되었다. 국민참여재판에서는 5내지 9인의 배심원이 피고인의 유무죄를 독립적으로 판단해서 재판부에 권고한다. 재판부는 배심원의 권고를 참고하지만, 그것에 구속받지 않고 판결하므로 한국의 국민참여재판은 실제 재판에서 일반인의 판단과 전문가의 판단을 비교할 수 있는 최상의 기회를 제공한다. Kim, Park, Park, & Eom(2013)은 한국에서 2010년 10월까지 실시된 모든 국민참여재판 274건에서 이루어진 323개의 유무죄 판단(한개의 재판에서 복수의 죄명에 대한 판단이 있을 수 있다)에서 배심원과 재판부의 판단일치도를 분석하였다. 이 연구에서 도출된 결과들은 전반적으로 일반인과 판사의 판단일치에 관한 외국연구의 결과들과 매우 유사하지만, 동일한 실제 재판사건에서 재판 당시에 이루어져서 공식적으로 재판기록에 편철된 배심원과 판사의 판단 일치도를 보여주는 즉, 소위 ‘측정오차(measurement error)’가 전무한 세계에서 유일한 자료라는 점에서 법관과 일반인 판단의 동일성과 차이에 대한 매우 귀중한 자료다.

Kim et al.(2013)의 국민참여재판 분석에서, 검사가 제시하는 증거의 강도가 비교적 강한 경우(155개의 판단)⁶⁾에는 배심원과 재판부의

6) Kim et al.(2013)의 연구에서 각 재판의 ‘증거강도(strength of evidence)’와 ‘증거복잡성(complexity of evidence)’은 최종 판결을 모르는 두명의 판사들이 서로 독립적으로 사건/재판기록을 열람하고 5점 척도에 평가하였다. 평가자간 신뢰도는 0.90이었다.

유무죄 판단이 98% 수준에서 일치하였고, 증거의 강도가 중간 정도인 경우(54개의 판단)에는 87%, 그리고 증거의 강도가 비교적 약한 경우(114개의 판단)에는 84%에서 배심원과 재판부의 유무죄 판단이 일치하였다. 한국의 국민참여재판에서 나타난 판단 일치율은 미국에서 Kalven & Zeisel(1966)와 Eigenberg et al.(2005)이 실제 재판들에서 발견한 일치율(78%와 75%)보다 높다.

한국의 국민참여재판에서 배심원과 재판부의 판단이 불일치하는 주원인은 무엇일까? 전체 323개의 유무죄 판단 중 총 28개(8.7%)의 판단에서 배심원과 재판부가 서로 불일치하였는데, Kim et al.(2013)은 그 불일치한 경우들이 특정 기소죄명(절도, 살인, 성범죄, 횡령, 강도, 등)에 몰려있거나, 다른 경우들에 비하여 증거나 적용법률이 더 복잡했다는 단서를 찾지 못하였다. 즉, 배심원이 특정 유형의 범죄 피고인에 대해 편견, 편향을 가지거나, 복잡한 증거 혹은 어려운 적용법률을 이해하지 못하여 잘못된 판단을 하였을 가능성을 찾지 못하였다. 반면, 배심원과 재판부의 판단이 불일치한 경우들 중 24개(86%)는 배심원이 무죄를 판단하고, 재판부가 유죄를 판단하였고, 4개(14%) 판단에서는 배심원이 유죄, 재판부가 무죄를 판단하였다. 흥미롭게도, 배심원과 재판부 판단이 불일치한 경우 중 배심원이 무죄를 판단하여 불일치한 비율과 유죄를 판단하여 불일치한 비율은 50여년 전 미국에서 Kalven & Zeisel(1966)이 발견한 비율들(86%, 14%)과 정확히 일치하였다. 배심원-재판부의 판단 불일치의 구체적 내용이 국가와 시대를 초월하여 동일한 정도로 크게 불균형을 이루는 이 현상은 그 불일치가 사건들에 포함된 우연적 요인이 아닌 체계적인 요인에 의해 초래된다는 것을

의미한다. 그 체계적인 요인 중 한가지는 일반인인 배심원과 전문가인 법관이 유무죄 판단 기준(합리적 의심의 초월)을 사건에 적용하는 방식이다. 배심원과 법관이 유무죄 판단 기준을 사건에 적용하는 방식에서 다르다는 것을 암시하는 구체적인 사례가 Kim et al.(2013)의 연구에서 나타났다.

배심원이 무죄, 재판부가 유죄를 판단한 24개의 국민참여재판 중에는 피고인이 ‘준강도’ 혐의로 기소된 사건이 3개 포함되었다. ‘준강도’는 폭행/협박과 재물절취의 시간적 순서가 ‘강도’와 뒤바뀐 경우인데, 예를 들면 절도행위 중 발각되어 도주하는 과정에서 추격하는 피해자와 몸싸움을 하여 피해자를 다치게 하거나, 피해자의 추격을 포기시킬 목적으로 위협한 경우가 포함된다. 형법 제335조는 준강도 혐의로 기소된 피고인이 유죄 판결을 받으면 사람을 폭행하거나 협박하여 재물을 절취하는 ‘강도’와 동일한 처벌(최소 징역 3년)을 받도록 규정하고 있다. 국민참여재판에 포함된 3개의 준강도 사건에서 배심원들은 무죄를 평결하고, 재판부는 유죄를 판결하였는데, 배심원들은 피고인의 범죄의도(절도)에 비해 처벌(강도)이 과한 경우에는 합리적 의심의 초월 기준을 엄격하게 적용하는 것으로 해석된다.

배심원과 법관이 합리적 의심의 초월 기준을 적용하는 방식에서 다르다는 것을 더 직접적으로 보여주는 결과가 Kim et al.(2013)의 “판단정확성” 분석에서도 나타났다. Kim et al.(2013)의 연구에서는 ‘배심원 판단’, ‘일심재판부 판단’, ‘항소심재판부 판단’, 그리고 ‘증거강도’를 관찰변인(observed variable)으로 설정하고, ‘실제 유무죄 여부’를 이론적인 잠재변인(latent variable)으로 설정한 잠재계층모형(latent class model)을 이용하여 실제 국민참여

재판에서 배심원과 재판부 판단의 정확성을 추정하였는데, 그 분석에서 매우 중요한 결과들이 도출되었다. 피고인이 실제 무죄로 가정되는 경우에는 배심원이 정확히 무죄를 판단할 확률(0.94)이 재판부가 무죄를 판단할 확률(0.79~0.81)보다 더 높게 추정된 반면, 피고인이 실제 유죄로 가정되는 경우에는 재판부가 정확히 유죄를 판단할 확률(1.00)이 배심원이 유죄를 판단할 확률(0.88~0.97)보다 더 높은 것으로 추정되었다. 배심원과 재판부의 오판 확률(오류율)을 더 세밀히 분석한 결과, 국민참여재판에서 일심재판부에 의해 이루어진 실제 유죄 판결 중 7.4~8.9%는 오판(제1종 오류)일 것으로 추정된 반면, 배심원에 의해 이루어진 실제 유죄 판단 중에는 2.2~3.0%가 오판일 것으로 추정되었다. 즉, 한국에서 최초 3년 동안 이루어진 국민참여재판에서 배심원은 재판부에 비해서 제1종 오류(무죄인 피고인을 유죄로 잘못 판단하는 오류)를 더 적게 범했을 것으로 추정되는 반면, 재판부는 배심원에 비해서 제2종 오류(유죄인 피고인을 무죄로 잘못 판단하는 오류)를 더 적게 범했을 것으로 추정되었다. 이 분석결과들은 배심원이 재판부보다 유죄판단을 위한 합리적 의심의 초월 기준을 더 높게(유죄증거의 증명력을 더 많이 요구하는 수준으로) 설정한다는 것을 의미한다.

합 의)

대한민국은 피고인의 무죄가 선고되는 형사

7) 이 합의 부분은 2014년 1월 27일자 대한변협신문에 칼럼형식(“판사의 합리적 의심 기준과 검사의 기소 정확성”)으로 게재될 예정이다.

재판의 비율이 미국과 영국 등의 보통법 국가들에 비해서 상대적으로 낮은 것으로 알려졌다. 무죄선고율 산출의 분모와 분자가 되는 형사재판 수에 어떤 사건유형들이 포함되는가에 따라 그 비율은 크게 달라질 수 있다. 예를 들어, 분모에 피고인이 유죄를 인정하는 사건을 포함할 것인지, 아니면 다투는 사건만 포함할 것인지를 여부, 분자에 전부무죄가 선고된 사건만 포함할 것인지를 여부, 아니면 일부무죄가 선고된 사건도 포함할 것인지를 여부에 따라 무죄선고율의 숫자는 크게 달라진다. 여러 나라의 무죄선고율을 표준화된 방식으로 집계하는 통계자료가 존재하지 않기 때문에, 대한민국의 무죄선고율이 다른 나라들에 비해서 얼마나 높은지, 혹은 낮은지는 정확히 파악하기 어렵다. 그러나 대한민국 대검찰청의 통계자료에 의하면, 선고가 이루어지는 제1심 재판에서 무죄가 선고되는 사건의 비율은 1%가 못 된다(2011년과 2012년의 경우 각각 0.63%). 반면, 미국의 경우는 주법원에서 무죄선고율이 16-41%에 이르고(Coughlan, 2000), 영국의 경우는 2009년에 크라운 법정에서 그 비율이 20%였다(BBC, 2012년 1월 6일. “Football association defends disciplinary system”). 비록 이 비율들이 서로 직접비교가 될 수는 없더라도, 수십배가 넘는 큰 차이는 대한민국의 무죄선고율이 영미의 그것보다 현저히 낮을 것으로 의심하기에 족하다. 무죄선고율에 영향을 주는 요인은 매우 많을 것이지만, 합리적 의심 기준과 더불어 검사의 기소 정확성도 중요한 요인이 될 수 있다.

합리적 의심 기준이 낮으면 무죄선고율이 낮아진다. 법원에서 사용하는 합리적 의심 기준이 현저하게 낮으면 거의 모든 형사재판에서 무죄가 선고될 확률이 거의 없기 때문에

무죄를 주장하려는 형사피고인의 관점에서는 재판이 실질적으로 무의미하여 차라리 허위라도 자백을 해서 양형을 줄이는 것이 현실적으로 더 현명하다는 판단을 할 수 있다. 이때에는 형사피고인들이 ‘학습된 무기력(learned helplessness)’에 침몰하여 유죄선고에 저항하지 않기 때문에 잠복된 실제적 유죄오판의 확률이 현저히 증가한다. 따라서 대한민국의 낮은 무죄선고율이 낮은 합리적 의심 기준에 기인한다면, 대한민국에서는 실제적 유죄오판이 광범위하게 발생하고 있을 것으로 추정할 수 있다.

반면, 대한민국의 무죄선고율이 낮은 이유가 판사들이 사용하는 합리적 의심 기준이 낮아서가 아니라, 검사들이 소위 ‘정밀사법’을 실천하고 있어서 기소가 정확하기 때문일 수도 있다. 실제적 무죄인 피의자는 검사에 의해 기소되지 않고, 실제적 유죄인 피의자는 기소되는 경향성(즉, 검사기소의 진단성)이 높을수록 재판에서 무죄가 선고되는 피고인의 비율은 낮아질 것이다. 대한민국의 검사들이 정밀사법을 실천하고 있다면, 낮은 무죄선고율에도 불구하고 대한민국에서는 유죄오판이 거의 없거나 매우 적을 것으로 추정할 수 있다.

그러나 대한민국의 검사가 미국이나 영국의 검사보다 피의자의 실제적 유무죄를 더 정확하게 파악하여 기소여부를 결정한다는 증거나 정황, 혹은 그렇게 믿을만한 합리적 이유는 거의 없다. 다만, 검사와 판사는 지식구조와 교육적 배경 등에서 거의 동일하거나 유사하고, 비슷한 논리법칙과 경험법칙을 사용하여 사고할 것이므로, 대한민국의 검사들이 재판에서 판사가 유죄결정을 할지, 아니면 무죄결정을 할지를 배심원이 유무죄를 판단하는 외

국의 검사들보다 더 정확히 예측할 수 있을 것으로 추정된다. 검사가 판사의 선고에 대한 정확한 예측에 기초해서 기소 여부를 결정하면, 무죄선고율이 현저히 낮아질 수 있다. 그 때문에 배심원이 참여하는 국민참여재판에서는 검사가 배심원들의 유무죄 결정을 판사의 그것만큼 정확히 예측하지 못하여 무죄선고 비율이 법관재판에서의 그것보다 더 높게 나타나는 것일지도 모른다.

만약, 대한민국 검사들의 기소 정확성과 정밀사법이 진실발견의 정확/정밀성이 아니라 판사행동에 대한 예측의 정확/정밀성이라면, 검사의 기소와 판사의 선고는 삼권분립제도가 가정하는 상호독립성이 전혀 없으므로 검사 기소의 정확성은 무죄선고율에 대해 전혀 영향을 주지 않는다. 이때에는 대한민국의 낮은 무죄선고율은 법원에서 사용하는 합리적 의심 기준이 엄격하지 않거나 낮은 것에 기인하고, 따라서 부지불식간에 실제적 유죄오판이 상당히 발생하고 있다는 것을 암시한다. 무죄추정 원칙과 적법절차에 헌법적 가치를 부여하는 대한민국에서 합리적 의심의 기준을 국민참여재판의 배심원들에게 이해시키는 노력과 더불어, 그 기준에 대한 법관의 이해와 인식을 일반인들의 그것에 조화시켜서 형사재판에서 제 1종 오류(무죄인 피고인에게 유죄를 선고하는 오류)의 가능성을 낮추는 것이 더 시급한지도 모른다. 판사의 선고가 정확해지면, 그것에 대한 예측을 무시할 수 없는 검사의 기소도 더 정확해질 것이다.

결 어

Kim et al.(2013)의 분석은 국민참여재판 뿐

만 아니라 한국에서 이루어지는 형사재판 전체에 대하여 중요한 함의를 가진다. 첫째는 일반인(배심원)이 재판에서 제시되는 증거들을 보고 피고인의 유무죄 여부를 판단하는 양식과 법관의 판단양식이 크게 다르지 않다는 것(즉, 판단 일치율이 매우 높다는 것)이다. 둘째는 판단기준(합리적 의심의 초월)의 적용에서는 일반인과 법관이 미묘한, 그러나 중요한 차이를 가질 수 있다는 것이다. 일반인이 법관보다 유죄판단을 위한 합리적 의심의 초월 기준을 높게 설정하고, 그로 인하여 장기적으로 제1종 오류를 범할 가능성이 법관보다 더 낮은 반면, 법관은 장기적으로 제2종 오류를 범할 가능성이 일반인보다 더 낮다는 것을 의미한다. 합리적 의심의 초월 기준에 대한 법관의 이해와 인식을 일반인들의 그것에 조화시켜서 형사재판에서 제1종 오류의 가능성을 낮추는 것이 필요할지도 모른다.

참고문헌

- 강동우 (2013). 형사법상 합리적 의심에 관한 연구. *홍익법학*, 14, 419~443.
- 김상준 (2013). 무죄판결과 법관의 사실인정에 관한 연구: 항소심의 파기자판 사례들을 중심으로. 박사학위 논문. 서울대학교 법학전문대학원.
- Alexander, L. & Sherwin, E. (2008). *Demystifying Legal Reasoning*. Cambridge, UK: Cambridge University Press.
- Blackstone, W. (1765-1769). *Commentaries on the Laws of England*. Oxford: Clarendon Press.
- Brook, J. (1982). *Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation*. *Tulsa Law Review*, 18, 79-109.
- Cecil, J. S., Hans, V. P., & Wiggins, E. C. (1991). Citizen Comprehension of Difficult Issues: Lessons from Civil Jury Trials. *American University Law Review*, 40, 727-774.
- Coughlan, P. J. (2000). In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting. *The American Political Science Review*, 94, 375-393.
- Dane, F. C. (1985). In search of reasonable doubt. *Law and Human Behavior*, 9, 141-158.
- Dawid, A. P. (2005). Bayes's theorem and weighing evidence by juries. In R. Swinburne (Eds.), *Bayes's Theorem*. London: British Academy Publications.
- Devitt, E. J., & Blackmar, C. B. (Eds.). (1977). *Federal jury practice and instructions* (Vol.2). St.Paul, MN: West.
- Dhami, M. K. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied*, 14, 353-363.
- Eisenberg, T., P. L. Hannaford-Agor, V. P. Hans, N. L. Waters, G. T. Munsterman, S. J. Schwab, & M. T. Wells (2005). Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's *The American Jury*. *Journal of Empirical Legal Studies*, 2, 171.
- FosterLee, L., Horowitz, I. A. and Bourgeois, M. J. (1993). Juror competence in civil trials: the effects of preinstruction and evidence technicality. *Journal of Applied Psychology*,

- 78, 14-21.
- Fried, M., Kaplan, K. J., & Klein, K. W. (1975). Juror selection: An analysis of voir dire. In R. J. Simon (Ed.), *The juror system in America: A critical overview* (pp. 58-64). Beverly Hills, CA: Sage.
- Gilbert, G. (1756). *The law of evidence*. London: H. Lintot for W. Owen; Morano, A. A. (1975). A reexamination of the development of the reasonable doubt rule. *Boston University Law Review*, 55, 507-528.
- Guthrie, C., Rachlinsky, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *The Cornell Law Review*, 86, 777-830.
- Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error and Unavoidable Injustice*. Oxford: Oxford University Press.
- Hans, V. P., & Vidmar, N. (1991) "The Twenty-Fifth Anniversary of The American Juries," 16 *Law & Social Inquiry* 323.
- Hastie, R., (1993). (Ed.). *Inside the juror: The psychology of juror decision making* (pp.84-115). Cambridge: Cambridge University Press.
- Hastie, R., Penrod, S. and Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
- Horowitz, I. A. (1997). Reasonable doubt instructions. *Commonsense justice and standard of proof. Psychology, Public Policy, and Law*, 3, 285-302.
- Horowitz, I. A. & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effects of reasonable doubt instructions, certainty of guilt standards, and jury verdicts. *Law and Human Behavior*, 20, 655-670.
- Kagehiro, D. K. (1990). Defining the standard of proof in jury instructions. *Psychological Science*, 1, 194-200.
- Kagehiro, D. K., & Rosen, C. J. (1986, October). Defining the standard of proof in jury instructions: II. Paper presented at the meeting of the American Society of Criminology, Atlanta, GA.
- Kagehiro, D. K., & Stanton, W. C. (1985). Legal vs. quantified definitions of standards of proof. *Law and Human Behavior*, 9, 159-178.
- Kalven, H., & Zeisel, H. (1966). *The American Jury*. Boston, MA: Little, Brown.
- Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. (1976). Guilt reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology*, 34, 282-294.
- Kim, S., Park, J., Park, K., & Eom, J. (2013). Judge-Jury Agreement in Criminal Cases: The First Three Years of the Korean Jury System. *Journal of Empirical Legal Studies*, 10, 1, 35-53.
- Koch, C. M., & Devine, D. J. (1999). Effects of reasonable doubt definition and inclusion of a lesser charge on jury verdicts. *Law and Human Behavior*, 23, 653-674.
- Kovera, M. B. & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? *Journal of Applied Psychology*, 85, 574-586.
- Kovera, M. B., McAuliff, B. D., Hebert, K. S.

- (1999). Reasoning about scientific evidence: Effects of juror gender and evidence quality on juror decisions in a hostile work environment case. *Journal of Applied Psychology*, 84, 362-375.
- Landsman, S., & Rakos, R. F. (1994). A preliminary inquiry into the effect of potentially biasing information on judges and jurors in civil litigation. *Behavioral Sciences and the Law*, 12, 113-126.
- Lempert, R. O. (1981). Civil Juries and Complex Cases: Let's Not Rush to Judgment." *Michigan Law Review*, 80, 68-132.
- MacCoun, R. J. (2005). Comparing legal factfinders: Real and mock, amateur and professional. Symposium on "The behavioral analysis of legal institutions," *Florida State University Law Review*, 32, 511-518.
- MacCoun, R. J. (1985). Evaluating juror performance (review of A. D. Austin's Complex litigation confronts the jury system). *Judicature*, June-July issue.
- MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology*, 54, 21-33.
- MacCoun, R. J., & Tyler, T. R. (1988). The basis of citizens' perceptions of the criminal jury: Procedural fairness, accuracy and efficiency. *Law and Human Behavior*, 12, 333-352.
- Martin, A. W., & Schum, D. A. (1987). Quantifying burdens of proof: A likelihood ratio approach. *Jurimetrics Journal*, 27, 383-402.
- McCauliff, C. M. A. (1982). Burdens of proof: Degrees of belief, quanta of evidence, or constitutional guarantees? *Vanderbilt Law Review*, 35, 1260-1335.
- Morano, A. A. (1975). A reexamination of the development of the reasonable doubt rule. *Boston University Law Review*, 55, 507-528.
- Nagel, S. S. (1979). Bringing the values of jurors in line with the law. *Judicature*, 63, 189-195.
- Nagel, S., Lamm, D., & Neef, M. (1981). Decision Theory and Juror Decision-Making. In B. D. Sales (Ed.), *The trial process*. New York: Plenum.
- Newman, J. (1993). Beyond Reasonable Doubt. *New York University Law Review*, 68, 979-1002.
- Pennington, N. & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, 62, 189-206.
- Penrod, S. & Heuer, L. B. (1998). Improving group performance: The case of the jury. In R. S. Tindale, L. Heath, J. Edwards, E. J. Posavac, F. B. Bryant, Y. Suarez-Balacazar, E. Henderson-King & J. Meyers, *Theory and Research on Small Groups*. Plenum: New York.
- Robbenolt, J. K. (2005). Evaluating juries by comparison to judges: A benchmark for judging? *Florida State University Law Review*, 32, 469-509.
- Robbenolt, J., Groscup, J. & Penrod, S. (2006). Evaluating and assisting jury competence in civil and criminal cases. In I. Weiner & A. Hess (ed.), *Handbook of Forensic Psychology* (3rd Ed.). New York: Wiley.
- Simon, R. J. (1969). Judges' translations of

- burdens of proof into statements of probability. *Trial Lawyer's Guide*, 13, 103-114.
- Simon, R. J. & Mahan L. (1971). Quantifying burdens of proof. A view from the bench, the jury, and the classroom. *Law and Society Review*, 319-330.
- Stoffelmayr, E., & Diamond, S. S. (2000). The conflict between precision and flexibility in explaining 'beyond a reasonable doubt.' *Psychology, Public Policy and Law*, 6, 769-787.
- Thompson, W. C., Cowan, C. L., Ellsworth, P. C. & Harrington, J. C. (1984). Death Penalty Attitudes and Conviction Proneness: The Translation of Attitudes into Verdicts. *Law and Human Behavior*, 8, 95-113.
- Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84, 1329-1393.
- Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, 62, 739-752.

1 차원고접수 : 2013. 6. 10.

심사통과접수 : 2013. 6. 21.

최종원고접수 : 2013. 7. 01.

Empirical Approach Toward the Standard of Proof for Criminal Trials

Sangjoon Kim

Seoul High Court
Appellate Chief Judge

Jeong-Ho Ahn

Kim & Jang
Attorney at Law

Kwangbai Park

Chungbuk
National University

The present study analyzed the perspectives of empirical science toward the standard of proof for criminal trials (“Beyond Reasonable Doubt”) as reflected on the studies that have been conducted to understand the qualitative as well as quantitative meaning of the standard. From this analysis, two implications emerged: (1) The theories and knowledge acquired through psychological studies of lay individuals can be used to understand judicial decision-making; (2) laypeople’s standard of proof for criminal trials may be higher than that of judges so that jurors may commit type-1 error less than would the judges in the long run. Based on the two implications, how to approach to the judge-jury difference, albeit small, in criminal fact-finding was discussed.

Key words : *Beyond Reasonable Doubt, standard of proof, judge, jury*