

Spring 2013

Standards of Proof in Civil Litigation: An Experiment from Patent Law

David L. Schwartz

Illinois Institute of Technology Chicago-Kent School of Law

Christopher B. Seaman

Washington and Lee University School of Law, seamanc@wlu.edu

Follow this and additional works at: <http://scholarlycommons.law.wlu.edu/wlufac>



Part of the [Litigation Commons](#)

Recommended Citation

David L. Schwartz & Christopher B. Seaman, *Standards of Proof in Civil Litigation: An Experiment from Patent Law*, 26 Harv. J. L. & Tech. 429 (2013).

This Article is brought to you for free and open access by Washington & Lee University School of Law Scholarly Commons. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Washington & Lee University School of Law Scholarly Commons. For more information, please contact osbornecl@wlu.edu.

**STANDARDS OF PROOF IN CIVIL LITIGATION:
AN EXPERIMENT FROM PATENT LAW**

*David L. Schwartz and Christopher B. Seaman**

TABLE OF CONTENTS

I. Introduction.....	430
II. Standards of Proof — An Overview	433
A. <i>The Burden of Proof</i>	433
B. <i>The Role and Types of Standards of Proof</i>	435
C. <i>Previous Empirical Studies</i>	437
1. Survey Evidence.....	437
2. Experimental Studies.....	440
III. <i>Microsoft v. i4i</i> and the Presumption of Validity in Patent Law	442
IV. Methodology and Study Design.....	447
A. <i>Hypotheses About i4i’s Impact</i>	447
B. <i>Why an Experiment?</i>	448
C. <i>Study Design</i>	451
V. Discussion	458
A. <i>Results</i>	459
B. <i>Implications</i>	466
C. <i>Directions for Future Research</i>	469
VI. Conclusion	473
Appendix A	474
Appendix B	479

* David L. Schwartz is an Associate Professor and Co-Director of the Center for Empirical Legal Studies of Intellectual Property at IIT Chicago-Kent College of Law. Christopher B. Seaman is an Assistant Professor at Washington and Lee University School of Law.

The authors thank Stefan Bechtold, Chris Buccafusco, Colleen Chien, Kevin Clermont, Kevin Collins, Howard Eglit, James Greiner, Ed Lee, Mark Lemley, Brian Love, Glynn Lunney, Rob MacCoun, Greg Mandel, Matthew Sag, Francis Shen, Stephanie Stern, Jay Tidmarsh, and participants of a faculty workshop at the IIT Chicago-Kent College of Law, the 2012 IP Scholars Conference (“IPSC”) at Stanford University Law School, the 2012 Works-in-Progress in Intellectual Property (“WIIP”) Colloquium at the University of Houston Law School, the 7th Annual Conference on Empirical Legal Studies (“CELS”) at Stanford University Law School, the 2012 IP Scholars Roundtable at Drake University Law School, PatCon 3 at the IIT Chicago-Kent College of Law, and the Junior Faculty Workshop series at IIT Chicago-Kent College of Law for their valuable comments and suggestions regarding this project. Finally, the authors thank Deborah Ginsburg of the IIT Downtown Campus Library and Ronald Fuller of the Washington and Lee University School of Law Library for their research assistance.

I. INTRODUCTION

Our litigation system is based upon the assumption that standards of proof matter.¹ They serve “to instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions.”² The various standards of proof reflect the legal system’s judgment about the proper allocation of risk between litigants, as well as the relative importance of the issues at stake.³ For example, in criminal cases where the defendant’s liberty may be at stake, the prosecution carries the burden of proving every element of the criminal charge “beyond a reasonable doubt.”⁴ In contrast, “preponderance of the evidence,” a much less stringent standard, is most common in civil cases.⁵ The third main standard, “clear and convincing evidence,” is an intermediate standard employed in civil litigation when “the individual interests at stake . . . are both ‘particularly important’ and ‘more substantial than mere loss of money.’”⁶

But despite the perceived importance of standards of proof, few empirical studies have tested lay jurors’ understanding and application of standards of proof, particularly in civil litigation.⁷ Specifically, to our knowledge, there has not been a large-scale study of a demographically representative population comparing jurors’ decisions when confronted with the two standards of proof used in civil litigation: (1) preponderance of the evidence, and (2) clear and convincing evidence.⁸

Patent law recently presented an opportunity to assess the impact of varying the standard of proof in civil litigation. Under Section 282

1. See, e.g., Bruce L. Hay & Kathryn E. Spier, *Burdens of Proof in Civil Litigation: An Economic Perspective*, 26 J. LEGAL STUD. 413, 413 (1997) (“Burden of proof rules . . . are central to the adversary system.”); Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738, 741 (2012) (“Systems of adjudication base outcomes on whether the strength of available evidence satisfies a designated burden of proof.”).

2. *In re Winship*, 397 U.S. 358, 370 (1970) (Harlan, J., concurring).

3. See *infra* Part II.B (explaining in more detail the role of standards of proof).

4. *In re Winship*, 397 U.S. at 361 (“[B]eyond a reasonable doubt . . . is now accepted in common law jurisdictions as the measure of persuasion by which the prosecution must convince the trier of all the essential elements of guilt.” (internal quotation marks omitted)); see also *Apprendi v. New Jersey*, 530 U.S. 466, 478 (2000).

5. See *Addington v. Texas*, 441 U.S. 418, 423–24 (1979); see also *Grogan v. Garner*, 498 U.S. 279, 286 (1991) (“Because the preponderance-of-the-evidence standard results in a roughly equal allocation of the risk of error between litigants, we presume that this standard is applicable in civil actions between private litigants unless particularly important individual interests or rights are at stake.” (internal quotation marks omitted)).

6. *Santosky v. Kramer*, 455 U.S. 745, 756 (1982) (quoting *Addington*, 441 U.S. at 424).

7. See *Addington*, 441 U.S. at 424–25 (noting that there are “no directly relevant empirical studies” regarding “what lay jurors understand concerning the differences among these three [standards] or the nuances of a judge’s instructions on the law”); Kaplow, *supra* note 1, at 809 (“The question of what probability factfinders actually associate with, say, the preponderance of the evidence rule — and how that minimum required probability varies by context — is an empirical one. Furthermore, it is one about which little is known.”).

8. See *infra* Part II.C.

of the Patent Act, every claim in a patent issued by the U.S. Patent and Trademark Office (“USPTO”) is presumed to be valid.⁹ This same statute also provides that “[t]he burden of establishing invalidity of a patent or any claim thereof shall rest on the party asserting such invalidity.”¹⁰ However, the statute fails to specify the standard of proof necessary to overcome this presumption of validity.

In a recent case, *Microsoft Corp. v. i4i Ltd. Partnership* (“*i4i*”), the U.S. Supreme Court heard competing arguments regarding the proper standard of proof for finding a patent invalid.¹¹ Microsoft argued that the preponderance of the evidence standard should apply, at least when the USPTO had not considered the prior art that allegedly invalidated the patent.¹² But the Court unanimously affirmed the U.S. Court of Appeals for the Federal Circuit’s (“Federal Circuit”) longstanding interpretation that invalidity must be proven “by clear and convincing evidence.”¹³

However, the Court also held that if the party challenging a patent’s validity could introduce new evidence in litigation that had not previously been considered by the USPTO during the patent’s examination, then “the challenger’s burden to persuade the jury of its invalidity defense by clear and convincing evidence may be easier to sustain.”¹⁴ In such situations, the jury should be instructed that “it has heard evidence that the [US]PTO had no opportunity to evaluate before granting the patent” and to “consider that fact when determining whether an invalidity defense has been proved by clear and convincing evidence.”¹⁵ But at the same time, the Court declined “to endorse any particular formulation” for such an instruction.¹⁶

Both the parties in *i4i* and the Court apparently assumed that the standard of proof would affect lay jurors’ decisions regarding invalidity, at least on the margins.¹⁷ In *i4i*’s wake, we tested this assumption by conducting an experiment involving the validity of a hypothetical

9. See 35 U.S.C. § 282 (2006 & Supp. V 2012) (“A patent shall be presumed valid. Each claim of a patent . . . shall be presumed valid independently of the validity of other claims . . .”).

10. *Id.*

11. 131 S. Ct. 2238 (2011).

12. *Id.* at 2244, 2249.

13. *Id.* at 2243, 2252; see also *Am. Hoist & Derrick Co. v. Sowa & Sons, Inc.*, 725 F.2d 1350, 1360 (Fed. Cir. 1984) (“[Section] 282 creates a presumption that a patent is valid and imposes the burden of proving invalidity on the attacker. That burden is constant and never changes and is to convince the court of invalidity by clear evidence.”).

14. *i4i*, 131 S. Ct. at 2251.

15. *Id.*

16. *Id.*

17. Cf. Brief of Apple Inc. and Intel Corp. as Amici Curiae Supporting Reversal at 15, *i4i*, 131 S. Ct. 2238 (2011) (No. 10-290) (arguing that parties in patent litigation frequently believe “[t]he evidentiary standard has a compelling effect on jurors . . . and it often is dispositive of the verdict” in patent validity challenges, particularly “in cases involving weak patents”).

patent. In our experiment, each subject received a short fact pattern simulating the information considered by a juror in a patent infringement lawsuit.¹⁸ In this fact pattern, a key piece of prior art introduced at trial by the party challenging the patent's validity had not been considered by the USPTO. After reading the fact pattern and a summary of the parties' arguments, each subject received a randomly assigned jury instruction regarding the standard of proof for establishing invalidity: (1) clear and convincing evidence, (2) clear and convincing evidence with an *i4i*-type instruction regarding the new evidence not considered by the USPTO, and (3) preponderance of the evidence.

Our results suggest that *i4i*'s impact — and the role of burdens of proof in civil litigation more generally — may be more complex than is generally assumed. Following conventional wisdom, we anticipated that a clear and convincing evidence standard with an *i4i*-type instruction would result in an intermediate burden between clear and convincing evidence (highest burden) and preponderance of the evidence (lowest burden), and that the subjects' decisions regarding the invalidity challenge would reflect these differences.¹⁹ This expectation was partially supported by our experiment. Subjects who received the preponderance of the evidence standard found the patent invalid more often than those who received the clear and convincing evidence standard.²⁰ Similarly, subjects who received a clear and convincing evidence standard with an *i4i*-type instruction (intermediate burden) found the patent invalid more often than those who received the plain clear and convincing evidence standard (i.e., without an *i4i*-type instruction).²¹ Both of these differences were statistically significant.²²

Subjects who received the clear and convincing standard with an *i4i*-type instruction (intermediate burden) unexpectedly found the patent invalid at rates statistically indistinguishable from those who received the lower preponderance of the evidence standard.²³ As discussed in more detail below, this finding may have important ramifications for invalidity challenges in patent litigation. It appears that the addition of an *i4i*-type instruction to the clear and convincing standard may effectively result in a standard of proof that, at least in jurors' views, is equivalent to the preponderance standard explicitly rejected by the Court in *i4i*.²⁴ This surprising result suggests that Microsoft may have actually achieved its desired outcome in *i4i* by mak-

18. See *infra* Part IV for an explanation of the study's methodology. The entire fact pattern and jury instructions given to subjects are reproduced *infra* Appendix A.

19. See *infra* pp. 19–20 for the specific hypotheses tested in our experiment.

20. See *infra* Part V.A.

21. See *infra* Part V.A.

22. See *infra* Part V.A.

23. See *infra* Part V.A.

24. See *infra* Part V.B.

ing it easier for juries to invalidate questionable patents, even though Microsoft lost the case.

Part II of this Article provides an overview of burdens of proof in litigation. It also analyzes the handful of existing empirical studies regarding the standard of proof. Part III summarizes the development of the presumption of validity in patent law, including the Supreme Court's decision in *i4i*. Part IV explains the methodology and experimental design employed in our experimental study. Part V recounts and analyzes the study's results, and then it concludes with some implications and potential directions for future research in this area.

II. STANDARDS OF PROOF — AN OVERVIEW

First, this Part distinguishes the standard of proof, which is the degree of certainty required for a jury or judge to find for a party on an issue, from the broader “burden of proof.” Second, it explains the purposes of standards of proof in litigation and the three main standards that are used. Third, it summarizes the literature on previous empirical studies regarding standards of proof.

A. The Burden of Proof

The burden of proof, described as one of “the slipperiest member[s] of the family of legal terms,”²⁵ has “vexed [both] courts and commentators for decades.”²⁶ This is at least partially because the burden of proof encompasses several different components, but courts and litigants have not always taken care to distinguish between them.²⁷

First, it can mean the “burden of production.” Initially described by James Bradley Thayer in his famous treatise on evidence,²⁸ the burden of production identifies which party must offer evidence in

25. *Schaffer v. Weast*, 546 U.S. 49, 56 (2005) (alteration in original) (quoting 2 J. STRONG, MCCORMICK ON EVIDENCE § 342, at 433 (5th ed. 1999)) (internal quotation marks omitted); see also CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE: PRACTICE UNDER THE RULES § 3.1, at 159 (2d ed. 1999) (stating that the burden of proof “is among the most slippery in the larger areas of procedure and evidence”).

26. *Nelson v. Hughes*, 625 P.2d 643, 645 (Or. 1981).

27. See *Dir., Office of Workers' Comp. Programs, Dep't of Labor v. Greenwich Collieries*, 512 U.S. 267, 272 (1994) (explaining that “[f]or many years the term ‘burden of proof’ was ambiguous because the term was used to describe two distinct concepts” — the “burden of persuasion” and the “burden of production”); *United States v. Hollis*, 569 F.2d 199, 204 n.6 (3d Cir. 1977) (“The meaning of the phrase, burden of proof, has been clouded by an unfortunate confusion of terminology” (internal quotation marks omitted)).

28. See JAMES BRADLEY THAYER, 1 A PRELIMINARY TREATISE ON EVIDENCE AT THE COMMON LAW 355 (1898) (describing this burden as “the duty of going forward in argument or in producing evidence[,] whether at the beginning of a case or at any later moment throughout the trial”).

order to raise an issue in litigation.²⁹ In other words, “[t]he burden of production is the obligation to make a prima facie case.”³⁰ This burden is relatively light; it is satisfied when a litigant has “produce[d] sufficient evidence so that a rational jury or other factfinder could find that each of the elements of the claim has been proven.”³¹

Second, the burden of proof can mean the “burden of persuasion” — i.e., the ultimate obligation on a party to persuade the decision maker that the party should prevail on a contested issue.³² The latter concept was famously explained by John Henry Wigmore as the “risk of nonpersuasion.”³³ In civil litigation, the burden of persuasion for all elements of a claim typically rests with the plaintiff, while the defendant has the burden of persuasion for most affirmative defenses.³⁴

Third, the burden of proof has been used to mean the “standard of proof.”³⁵ This concept “refers to the quantum of proof required to sustain the burden of persuasion.”³⁶ In *i4i*, the Supreme Court explained the “standard of proof” as:

[T]he degree of certainty by which the factfinder must be persuaded . . . to find in favor of the party bearing the burden of persuasion. In other words, the term “standard of proof” specifies how difficult it will be for the party bearing the burden of persuasion to convince the jury of the facts in its favor.³⁷

In this Article, we are concerned exclusively with the standard of proof, as opposed to the other aspects of the burden of proof.

29. See *Schaffer*, 546 U.S. at 49; see also Pamela R. Metzger, *Cheating the Constitution*, 59 VAND. L. REV. 475, 521 (2006) (“The burden of production requires a party to properly raise an issue to put that issue into play.”).

30. 31A C.J.S. EVIDENCE § 189 (2012).

31. 29 AM. JUR. 2D EVIDENCE § 181 (2012); see also *St. Mary’s Honor Ctr. v. Hicks*, 509 U.S. 502, 506 (1993) (holding the burden of production is satisfied in the employment discrimination context when the plaintiff establishes “by a preponderance of the evidence, a ‘prima facie’ case”).

32. See *Microsoft Corp. v. i4i Ltd. P’ship*, 131 S. Ct. 2238, 2245 n.4 (2011) (using the term “burden of persuasion” to “identify the party who must persuade the jury in its favor to prevail”).

33. See *Dir., Office of Workers’ Comp. Programs, Dep’t of Labor v. Greenwich Collieries*, 512 U.S. 267, 282 (citing 9 J. WIGMORE, EVIDENCE § 2486 (J. Chadbourne rev. 1981)); see also Fleming James, Jr., *Burdens of Proof*, 47 VA. L. REV. 51, 51 (1961).

34. See *Schaffer*, 546 U.S. at 57.

35. Some have classified the “standard of proof” (along with the risk of nonpersuasion) as part of the “burden of persuasion.” See, e.g., Lawrence B. Solum, *Presumptions and Transcendentalism: You Prove It! Why Should I?*, 17 HARV. J.L. & PUB. POL’Y 691, 691 (1994) (“The burden of persuasion has two components, the risk of nonpersuasion and the standard of proof.”).

36. Metzger, *supra* note 29, at 521.

37. *i4i*, 131 S. Ct. at 2245 n.4 (citation omitted).

B. The Role and Types of Standards of Proof

The standard of proof serves multiple purposes. First, it determines how to allocate the risk of error between litigants.³⁸ In litigation, there is always some risk the decision maker will reach an erroneous conclusion.³⁹ As Justice Harlan explained in *In re Winship*, “in a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what *probably* happened.”⁴⁰ The standard of proof thus attempts to balance the risk between so-called Type I errors (i.e., false positives), such as an erroneous finding of liability in a civil case or the conviction of an innocent person, and Type II errors (i.e., false negatives), such as the denial of a meritorious claim in a civil case or an erroneous acquittal of a criminal defendant.⁴¹

Second, the standard of proof is used to “indicate the relative importance attached to the ultimate decision.”⁴² In particular, the Supreme Court has recognized that a heightened standard of proof — i.e., something greater than a mere preponderance of the evidence — is required when fundamental rights or important liberty interests are at issue.⁴³

Because the costs of error and consequences of the outcome obviously vary in different settings, three primary standards of proof exist at common law. From lowest to highest degree of certainty required, they are: (1) preponderance of the evidence, (2) clear and convincing evidence, and (3) beyond a reasonable doubt.⁴⁴

Preponderance of the evidence is the standard of proof for most issues in civil litigation, reflecting the legal system’s assumptions that the standard “results in a roughly equal allocation of the risk of error

38. See *Herman & MacLean v. Huddleston*, 459 U.S. 375, 389 (1983) (explaining that the standard of proof “serves to allocate the risk of error between the litigants” (quoting *Addington v. Texas*, 441 U.S. 418, 423 (1979) (internal quotation marks omitted)); see also Kaplow, *supra* note 1, at 741 (“The stringency of the proof burden determines how error is allocated between mistakes of commission — improper assignment of liability — and mistakes of omission — improper exoneration.”).

39. See, e.g., *Speiser v. Randall*, 357 U.S. 513, 525 (1958) (“There is always in litigation a margin of error, representing error in factfinding, which both parties must take into account.”).

40. *In re Winship*, 397 U.S. 358, 370 (1970) (Harlan, J., concurring).

41. RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 827 (8th ed. 2011); see also *In re Winship*, 397 U.S. at 370–71 (Harlan, J., concurring).

42. *Addington*, 441 U.S. at 423.

43. See *id.* at 424 (explaining that heightened standards of proof are necessary when the interests at stake are “particularly important” and “more substantial than mere loss of money”).

44. *Id.* at 423–24; see also Kevin M. Clermont, *Standards of Proof Revisited*, 33 VT. L. REV. 469, 486 (2009).

between litigants”⁴⁵ and the costs of a “mistaken judgment for the plaintiff is no worse than a mistaken judgment for the defendant.”⁴⁶ In other words, as Chief Justice Rehnquist explained, “[t]he preponderance-of-the-evidence standard . . . is employed when the social disutility of error *in either direction* is roughly equal.”⁴⁷

In contrast, proof beyond a reasonable doubt is constitutionally required for conviction in criminal cases.⁴⁸ This standard has been described by the Court as the “bedrock . . . principle whose ‘enforcement lies at the foundation of the administration of our criminal law.’”⁴⁹ In terms of allocation of risk, the beyond a reasonable doubt standard places almost all risk upon the government rather than the criminal defendant.⁵⁰ In other words, it strongly prefers Type II errors (erroneous acquittals) instead of Type I errors (erroneous convictions). This is exemplified by Blackstone’s famous formulation that “it is better that ten guilty persons escape, than that one innocent suffer.”⁵¹ The preference for Type II errors embodied in the beyond a reasonable doubt standard is due to the severe consequences of a conviction — the loss of liberty as well as the stigma and collateral consequences that accompany a criminal record.⁵²

Clear and convincing evidence represents an “intermediate standard” between preponderance of the evidence and reasonable doubt.⁵³ Although the exact terminology used to refer to this heightened standard of proof has varied over time,⁵⁴ jury instructions often explain that

45. *Grogan v. Garner*, 498 U.S. 279, 286 (1991); *see also* Chad M. Oldfather, *Appellate Courts, Historical Facts, and the Civil-Criminal Distinction*, 57 VAND. L. REV. 437, 467 (2004) (“The preponderance of the evidence standard is thus appropriate in the civil context because it has little if any tendency to generate erroneous verdicts disproportionately in favor of plaintiffs or defendants.”).

46. 2 MCCORMICK ON EVIDENCE § 341, at 570 (Kenneth S. Broun ed., 6th ed. 2006).

47. *Santosky v. Kramer*, 455 U.S. 745, 788 n.13 (1982) (Rehnquist, J., dissenting).

48. *In re Winship*, 397 U.S. 358, 364 (1969); *see also* *Sullivan v. Louisiana*, 508 U.S. 275, 278 (1993) (holding that the Sixth Amendment guarantee of trial by jury requires a jury verdict of guilty beyond a reasonable doubt).

49. *In re Winship*, 397 U.S. at 363 (quoting *Coffin v. United States*, 156 U.S. 432, 453 (1895)).

50. *See, e.g.*, *Addington v. Texas*, 441 U.S. 418, 423–24 (1979) (“In the administration of criminal justice, our society imposes almost the entire risk of error upon itself . . . by requiring . . . that the state prove the guilt of an accused beyond a reasonable doubt.”).

51. 4 WILLIAM BLACKSTONE, COMMENTARIES *358–59. The 10:1 ratio articulated by Blackstone is contested; as Alexander Volokh has pointed out, the perceived optimal ratio between erroneous acquittals and erroneous convictions has varied dramatically in Anglo-American jurisprudence. *See generally* Alexander Volokh, n *Guilty Men*, 146 U. PA. L. REV. 173, 182–91, 198–203 (1997).

52. *In re Winship*, 397 U.S. at 363 (“The accused during a criminal prosecution has at stake interests of immense importance, both because of the possibility that he may lose his liberty upon conviction and because of the certainty that he would be stigmatized by the conviction.”).

53. *Addington*, 441 U.S. at 424.

54. *Compare id.* (explaining this standard “usually employs some combination of the words ‘clear,’ ‘cogent,’ ‘unequivocal’ and ‘convincing’”), with *Buildex Inc. v. Kason Indus., Inc.*, 849 F.2d 1461, 1463 (Fed. Cir. 1988) (“Although not susceptible to precise defi-

it is higher than a preponderance of the evidence standard, but it does not require proof beyond a reasonable doubt.⁵⁵ Less commonly employed than the preponderance standard used for most civil issues, clear and convincing evidence is often required when “[t]he interests at stake . . . are deemed to be more substantial than mere loss of money.”⁵⁶ For example, the Court has held that clear and convincing evidence is required for involuntary civil commitment due to mental illness,⁵⁷ deportation for violation of immigration law,⁵⁸ and the termination of parental rights.⁵⁹ In patent law, a variety of issues must be proven by clear and convincing evidence, including willful infringement,⁶⁰ inequitable conduct,⁶¹ correction of inventorship,⁶² and overcoming the presumption of priority in an interference proceeding with an issued patent.⁶³

C. Previous Empirical Studies

1. Survey Evidence

Several previous studies have attempted to empirically measure standards of proof by asking potential decision makers to assign percentage probabilities to the various standards.

In the early 1970s, Rita James Simon and Linda Mahan asked judges and jurors to quantify the beyond a reasonable doubt standard and the preponderance of the evidence standard as percentage proba-

tion, ‘clear and convincing’ evidence has been described as evidence which produces in the mind of the trier of fact ‘an abiding conviction that the truth of [the] factual contentions are ‘highly probable.’” (quoting *Colorado v. New Mexico*, 467 U.S. 310, 316 (1984)), and *State v. Addington*, 588 S.W.2d 569, 570 (Tex. 1979) (“Clear and convincing evidence is defined as that measure or degree of proof which will produce in the mind of the trier of fact a firm belief or conviction as to the truth of the allegations sought to be established.”).

55. *See, e.g.*, KEVIN F. O’MALLEY, JAY E. GREING & HON. WILLIAM C. LEE, 3 FEDERAL JURY PRACTICE & INSTRUCTIONS § 104:02 (6th ed. 2012) (“Clear and convincing evidence involves a greater degree of persuasion than is necessary to meet the preponderance of the evidence standard. This standard does not require proof to an absolute certainty, since proof to an absolute certainty is seldom possible in any case.”).

56. *Addington*, 441 U.S. at 424.

57. *Id.* at 431–33.

58. *Woodby v. INS*, 385 U.S. 276, 286 (1966).

59. *Santosky v. Kramer*, 455 U.S. 745, 747–48 (1982).

60. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc).

61. *Therasense, Inc. v. Becton, Dickinson & Co.*, 649 F.3d 1276, 1287 (Fed. Cir. 2011) (en banc).

62. *Hess v. Advanced Cardiovascular Sys., Inc.*, 106 F.3d 976, 980 (Fed. Cir. 1997).

63. *Price v. Symsek*, 988 F.2d 1187, 1191 (Fed. Cir. 1993). Although the Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112-29, 125 Stat. 284 (2011), has switched U.S. patent law to a first-inventor-to-file system for all patent applications filed on or after March 16, 2013, interference proceedings commenced before September 16, 2012, involving patent applications filed before the AIA’s effective date are unaffected by this change. *See* AIA § 6(f)(3)(B); 37 C.F.R. § 42.200(d) (2012).

bilities.⁶⁴ Both judges and jurors believed that the beyond a reasonable doubt standard demanded a high likelihood in order to be satisfied. Specifically, the median judge quantified this standard as requiring 88% probability (i.e., 8.8 on a 10-point scale), with almost a third of judges responding that it required 100% probability (i.e., 10.0) and another third that it required 90% or 95% probability (i.e., 9.0 or 9.5).⁶⁵ Similarly, the median juror⁶⁶ quantified the beyond a reasonable doubt standard as requiring 86% probability (i.e., 8.6 on a 10-point scale), with over a third articulating it as requiring 100% probability (i.e., 10.0) and another third between 80% and 95% probability (i.e., 8.0 to 9.5).⁶⁷

In the same study, however, judges and jurors substantially disagreed about the percentage likelihood needed to satisfy the preponderance standard. For the median judge, this standard was satisfied with a 54% probability (i.e., 5.4 on a 10-point scale).⁶⁸ In contrast, the median juror believed that 75% probability was required (i.e., 7.5 on a 10-point scale) to satisfy the preponderance standard.⁶⁹

In the early 1980s, C.M.A. McCauliff conducted a survey of all active, senior, and retired federal judges regarding the level of certainty, on a scale of 0%–100%, required by nine phrases treated as standards of proof.⁷⁰ Her results generally paralleled the judges' responses in the Simon & Mahan study. For preponderance of the evidence, the

64. See generally Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 LAW & SOC'Y REV. 319 (1971). The survey contacted a nationwide sample of federal and state judges by mail. *Id.* at 320; see also Rita James Simon, *Judges' Translations of Burdens of Proof into Statements of Probability*, in THE TRIAL LAWYER'S GUIDE 103, 109–13 (John J. Kennelly et al. eds., 1969) (explaining the results of the judges' survey).

Some academics argue that the standards of proof cannot be reduced to a "mere statistical probability" and instead are interpreted by jurors as referring to varying degrees of "belief." See, e.g., Richard W. Wright, *Proving Causation: Probability Versus Belief*, in PERSPECTIVES ON CAUSATION 195 (Richard Goldberg ed., 2011). Regardless of how one conceptualizes the standards of proof, it is undisputed that a clear and convincing burden is meant to be a higher standard of proof than a preponderance of the evidence burden.

65. Simon & Mahan, *supra* note 64, at 324 tbl.4.

66. Jurors in the Simon & Mahan study consisted of a sample of people called for jury service in state court in Champaign County, Illinois. *Id.* at 320–21.

67. *Id.* However, a quarter of jurors reported 50%–60% probability (i.e., score of 5.0 to 6.0 on a 10-point scale) for the beyond a reasonable doubt standard. *Id.*

68. *Id.* at 327 tbl.7. However, a surprising number of judges — almost 20% of respondents (62 of 334, excluding 17 not answering) — reported a score exceeding 7.0 (greater than or equal to 70% probability) for preponderance of the evidence. *Id.*

69. *Id.* at 327 tbl.7. This included more than 10% of respondents (8 of 69) who reported the highest possible score of 10.0 on a 10-point scale (i.e., 100% probability) for the preponderance of the evidence standard. *Id.*

70. C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1324–25 (1982). The nine phrases used to articulate standards of proof in McCauliff's study were "beyond a reasonable doubt," "clear and convincing evidence," "clearly erroneous," "preponderance of the evidence," "more probable than not," "substantial evidence," "probable cause to believe," "reasonable cause to believe," and "reasonable suspicion." *Id.* at 1325 n.184.

overwhelming majority of judges (154 of 175) equated this standard with a probability of 50% or 60%, with an average probability of 55.3%.⁷¹ For beyond a reasonable doubt, nearly all judges (160 of 171) rated this standard between 80% to 100% probability, with an average probability of 90.3%.⁷² Finally, for the clear and convincing evidence standard, the majority of judges (111 of 170) rated this standard as 70% to 80% probability, with an average probability of 75.0%.⁷³

More recently, Bradley Saxton surveyed actual jurors in civil and criminal cases after trial to determine their comprehension of various jury instructions, including instructions regarding the standard of proof.⁷⁴ In criminal cases, jurors were asked whether the state only needed to convince the jury “that it is more likely than not” that the defendant committed the accused crime.⁷⁵ The vast majority of jurors — almost 85% — correctly responded that they were “very sure” or “pretty sure” this statement was false, while almost 10% were “very sure” or “pretty sure” it was true and 6% did not know.⁷⁶ In civil cases, jurors were asked the opposite question — whether the jury “has to be convinced beyond a reasonable doubt that the plaintiff’s claims are correct.”⁷⁷ The results for this question suggested that a significant number of jurors misunderstood the preponderance standard: 38% incorrectly responded that they were “very sure” or “pretty sure” it was true, while an additional 3% did not know.⁷⁸

In sum, these studies suggest that judges generally understand the preponderance standard to be satisfied by any probability exceeding 50%, clear and convincing evidence to be in the range of 70% to 80% probability, and beyond a reasonable doubt to require at least 80% probability.⁷⁹ It is unclear whether lay jurors have a similar under-

71. *Id.* at 1331 tbl.7, 1332 tbl.8. Of the remaining respondents, 7 judges (out of 175) rated preponderance of the evidence at 70% probability, another 7 judges rated it 80% probability, 3 judges rated it at 90% probability, another 3 judges rated it at 100% probability, and 1 judge (bizarrely) rated it at 0% probability. *Id.* at 1331 tbl.7.

72. *Id.* at 1325 tbl.2, 1332 tbl.8.

73. *Id.* at 1328 tbl.5, 1332 tbl.8. Of the remaining respondents, 4 judges (out of 170) rated clear and convincing evidence at 50%–55% probability, 27 judges rated it at 60%–65% probability, 24 rated it at 85%–90% probability, and 4 judges rated it at 95%–100% probability. *Id.* at 1328 tbl.5.

74. Bradley Saxton, *How Well Do Jurors Understand Jury Instructions? A Field Test Using Real Juries and Real Trials in Wyoming*, 33 LAND & WATER L. REV. 59 (1998) (surveying jurors in Wyoming state courts).

75. *Id.* at 98 fig.13.

76. *Id.* (N = 183).

77. *Id.* at 100 fig.14.

78. *Id.* (N = 68). For the remaining (correct) respondents, 22% were “pretty sure” it was false, while 37% were “very sure” it was false. *Id.*

79. In addition to the studies listed above, in the late 1970s, Judge Jack Weinstein — the author of several treatises on evidence and procedure — reported the results of an informal survey of ten of his colleagues on the U.S. District Court for the Eastern District of New York regarding probabilities associated with these three standards of proof. The surveyed

standing. The Simon & Mahan and Saxton studies surveying actual jurors had sample sizes that were too small to draw definitive conclusions, failed to ask for a percentage probability for all three standards of proof, or both. However, these studies suggest that some jurors perceive the preponderance standard to be somewhat more rigorous than do judges. Furthermore, these surveys do not test whether jurors or judges actually apply these standards consistently in litigation.

2. Experimental Studies

Another area of empirical research has been the use of experiments to test the impact of standards of proof as articulated in jury instructions, primarily in criminal cases. In several experiments, “the wording used to convey the standard of proof” was found to have “a substantial impact” on mock jurors’ verdicts.⁸⁰

In a study conducted by the Jury Project at the London School of Economics in the early 1970s, panels of mock jurors in the United Kingdom heard a tape-recorded trial of a rape case.⁸¹ The only variation was the jury instruction on the standard of proof.⁸² Two groups of mock juries received differing instructions regarding the beyond a reasonable doubt standard,⁸³ while a third group received a preponderance of the evidence standard.⁸⁴ The juries’ conviction rates were highest for the preponderance instruction, slightly lower for the first reasonable doubt instruction, and much lower for the second reasona-

judges all reported that any probability greater than 50% would satisfy the preponderance of the evidence standard; most estimated the clear and convincing evidence standard as between 60% and 75% (although several judges refused to assign a discrete value); and the judges reported that the probability associated with the beyond a reasonable doubt standard was between 76% and 95% (with a median of 85%). *United States v. Fatico*, 458 F. Supp. 388, 410 (E.D.N.Y. 1978), *aff’d*, 603 F.2d 1053 (2d Cir. 1979).

80. Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 *PSYCHOL. PUB. POL’Y & L.* 622, 665 (2001).

81. London Sch. of Econ., *Juries and the Rules of Evidence*, 1973 *CRIM. L. REV.* 208 [hereinafter *L.S.E. Jury Project*], discussed in Barbara D. Underwood, *The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases*, 86 *YALE L.J.* 1299, 1309–10 (1977).

82. See Underwood, *supra* note 81.

83. See Underwood, *supra* note 81. The first beyond a reasonable doubt instruction stated: “[Before you convict] you should be sure beyond reasonable doubt and by reasonable doubt I mean not a fanciful doubt that you might use to avoid an unpleasant decision, but a doubt for which reasons can be given.” See Underwood, *supra* note 81 at 1309 (alteration in original) (quoting *L.S.E. Jury Project*, *supra* note 81, at 213) (internal quotation marks omitted). The second beyond a reasonable doubt instruction stated: “Before you convict you must feel sure and certain on the evidence you have heard that the accused is guilty.” Underwood, *supra* note 81, (quoting *L.S.E. Jury Project*, *supra* note 81, at 213) (internal quotation marks omitted).

84. Underwood, *supra* note 81, at 1310 (“[B]efore you convict you must feel satisfied that it is more likely than not that the accused is guilty.” (alteration in original) (quoting *L.S.E. Jury Project*, *supra* note 81, at 214) (internal quotation marks omitted)).

ble doubt instruction.⁸⁵ This study “suggests that jurors can distinguish between” preponderance of the evidence and beyond a reasonable doubt standards, although one jury scholar noted that the specific wording of jury instructions regarding the latter standard could result in substantially different outcomes.⁸⁶

In a 1985 article, Dorothy K. Kagehiro and W. Clark Stanton reported the results of several experiments that tested all three standards of proof using jury instructions that explained the relevant standard either as a percentage probability (“quantified definition”) or with a conventional model jury instruction that did not explicitly quantify each standard (“legal definition”).⁸⁷ The subjects, undergraduate psychology students, were divided into an experimental and a control group.⁸⁸ All subjects were presented with a summary of the evidence in a civil trial.⁸⁹ Then the subjects in the experimental group, but not the subjects in the control group, were instructed regarding the standard of proof.⁹⁰ Finally, all subjects were asked to individually reach a decision regarding the plaintiff’s claim.⁹¹ The authors reported that although “verdicts favoring the plaintiffs decreased as the standard became stricter” in both experiments,⁹² these differences were only statistically significant for the quantified definitions.⁹³ In contrast, “[f]or the legal definitions, the multivariate effect [for the] standard of proof was not significant.”⁹⁴

In 1996, Irwin A. Horowitz and Laird C. Kirkpatrick tested the impact of various definitions of the reasonable doubt standard.⁹⁵ The subjects — jury-eligible adults — heard one of two versions of a murder trial, one that was calibrated to be a “weak” case likely to result in a not guilty verdict, while the other was a “strong” case with

85. *Id.*

86. *Id.* Similarly, in another study led by Norbert L. Kerr in the mid-1970s, mock jurors reached statistically significant outcomes regarding a rape trial depending on the particular jury instruction used to define reasonable doubt. Norbert L. Kerr et al., *Guilt Beyond a Reasonable Doubt: Effects of Concept Definition and Assigned Decision Rule on the Judgments of Mock Jurors*, 34 J. PERSONALITY & SOC. PSYCHOL. 282 (1976).

87. Dorothy K. Kagehiro & W. Clark Stanton, *Legal vs. Quantified Definitions of Standards of Proof*, 9 LAW & HUM. BEHAV. 159, 161–63 (1985).

88. *Id.* at 162.

89. *Id.* at 162, 165–66.

90. *Id.*

91. *Id.* For the first experiment, $N = 198$ (after excluding 54 respondents because of errors on one or more manipulation checks), and for the second experiment, $N = 220$. *Id.* at 163, 165.

92. *Id.* at 164–65, 168.

93. *Id.* at 164, 168 ($p < 0.001$ for the first experiment using the quantified definition, $p < 0.001$ for the second experiment using the quantified definition).

94. *Id.* at 164.

95. Irwin A. Horowitz & Laird C. Kirkpatrick, *A Concept In Search of a Definition: The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts*, 20 LAW & HUM. BEHAV. 655 (1996).

evidence strongly suggestive of guilt.⁹⁶ Each juror then received one of five definitions of reasonable doubt and was asked to reach a preliminary decision.⁹⁷ Subsequently, panels of mock juries were assembled and deliberated with no time limitation.⁹⁸ The authors reported statistically significant differences between mock juries that received varying definitions of reasonable doubt.⁹⁹

In general, although some of these experiments have limitations that impact their external validity — such as the use of undergraduate psychology students rather than a more broadly representative population — and the evidence is not clear cut,¹⁰⁰ they cumulatively suggest that jurors' decisions may be swayed by the standard of proof. In addition, they imply that the particular wording of jury instructions can play a significant role in jurors' decision making. However, none of the previous empirical studies directly answers the question we study here — whether altering the burden of proof in civil litigation is likely to result in different outcomes by a diverse, broadly representative group of jury-eligible citizens. This unresolved question is the subject of our experiment.

III. *MICROSOFT V. I4I* AND THE PRESUMPTION OF VALIDITY IN PATENT LAW

Section 282 of the Patent Act, first adopted in 1952, expressly provides that patents are to be presumed valid, but does not specify the standard of proof.¹⁰¹ Before the creation of the Federal Circuit in 1982, some of the regional circuits held that, in certain circumstances, the standard for overcoming the presumption of validity was prepon-

96. *Id.* at 661.

97. *Id.* at 661–62. The authors labeled the definitions (1) “firmly convinced” (based on the Federal Judicial Center’s instructions), (2) “moral certainty,” (3) “does not waver or vacillate,” (4) “real doubt,” or (5) “undefined” (for no explanation). *Id.* at 660–61.

98. *Id.* at 661. However, a so-called “Allen” charge (a jury instruction intended to prevent a hung jury by encouraging jurors in the minority to reconsider their positions) was given after 90 minutes of deliberation. All mock juries eventually reached a decision. *Id.*

99. *Id.* at 662–63 & tbl.1. Specifically, in the “weak” case, the “firmly convinced” instruction differed significantly (lower) from all other instructions ($p < 0.05$), while in the “strong” case, the “moral certainty” instruction was marginally more significant (higher) compared to the “does not waver or vacillate” and “undefined” instructions. *Id.* at 663 & tbl.1.

100. See Elisabeth Stoffelmayr & Shari Seidman Diamond, *The Conflict Between Precision and Flexibility in Explaining “Beyond a Reasonable Doubt,”* 6 PSYCHOL. PUB. POL’Y & L. 769, 774–76 (2000) (summarizing previous studies, including Kagehiro & Stanton, *supra* note 87, and concluding that “[e]mpirical research indicates that jurors may have some difficulty distinguishing” the clear and convincing standard of proof).

101. 35 U.S.C. § 282 (2006 & Supp. V 2012) (“A patent shall be presumed valid The burden of establishing invalidity of a patent or any claim thereof shall rest on the party asserting such invalidity.”).

derance of the evidence.¹⁰² However, the Federal Circuit, in a decision not long after its creation, held that the proper standard of proof was clear and convincing evidence.¹⁰³ According to the Federal Circuit, the clear and convincing standard is “constant and never changes.”¹⁰⁴

In 2011, the U.S. Supreme Court directly confronted the question of how high the standard of proof should be for proving patent invalidity in litigation.¹⁰⁵ More specifically, the Court in *i4i* examined whether Section 282 of the Patent Act required clear and convincing evidence to overcome a patent’s presumption of validity.

In the lower courts, Microsoft unsuccessfully challenged *i4i*’s patent on the basis that *i4i* had previously sold software that embodied the invention more than one year before it applied for a patent, which if true would render it invalid under Section 102(b) of the Patent Act.¹⁰⁶ Because the prior art sale of software had not been presented by *i4i* to the Patent Examiner when the USPTO was considering *i4i*’s patent application, Microsoft argued that it should be required to prove invalidity by only a preponderance of the evidence.¹⁰⁷ The district court judge rejected Microsoft’s contention, ruled that Microsoft had to prove invalidity by clear and convincing evidence, and denied Microsoft’s request for an alternative instruction.¹⁰⁸ Thereafter, the jury found Microsoft to have willfully infringed *i4i*’s patent, rejected Microsoft’s invalidity defense, and awarded over \$200 million in damages.¹⁰⁹ The Federal Circuit affirmed the finding of infringement, validity, and the damages award.¹¹⁰ The Supreme Court granted Microsoft’s petition for certiorari on the issue of the proper standard of proof for Microsoft’s invalidity defense.¹¹¹

At the Supreme Court, Microsoft advanced two arguments in support of reversal. First, it asserted that an accused infringer need only prove a patent claim invalid by a preponderance of the evidence,

102. See, e.g., *Baumstimler v. Rankin*, 677 F.2d 1061, 1066 (5th Cir. 1982) (stating that when a patent is challenged for failure to consider prior art, “the challenger of the validity of the patent need no longer bear the heavy burden of establishing invalidity either ‘beyond a reasonable doubt’ or ‘by clear and convincing evidence.’”); *Mfg. Research Corp. v. Graybar Elec. Co.*, 679 F.2d 1355, 1364 (11th Cir. 1982) (noting that the trial court “erred in instructing the jury to apply a clear and convincing evidence standard to the defense of invalidity” when the prior art evidence admitted at trial had not been considered by the USPTO).

103. *Am. Hoist & Derrick Co. v. Sowa & Sons, Inc.*, 725 F.2d 1350, 1360 (Fed. Cir. 1984).

104. *Id.*

105. *Microsoft Corp. v. i4i Ltd. P’ship*, 131 S. Ct. 2238 (2011).

106. *Id.* at 2243–44 (citing 35 U.S.C. § 102(b) (2006)).

107. *Id.* at 2244.

108. *Id.* (discussing the district court proceedings).

109. See *i4i Ltd. P’ship v. Microsoft Corp.*, 598 F.3d 831, 839 (Fed. Cir. 2010) (discussing the district court proceedings).

110. *Id.*

111. *Microsoft Corp. v. i4i Ltd. P’ship*, 131 S. Ct. 647 (2010) (mem.). Microsoft did not seek Supreme Court review of the decisions on willful infringement and damages. *i4i*, 131 S. Ct. at 2238.

rather than by clear and convincing evidence.¹¹² Because § 282 was silent regarding the standard for carrying the burden of proving invalidity, Microsoft contended that “the default standard of proof in civil cases” — preponderance of the evidence — should apply.¹¹³ In addition, it argued that the private property rights granted by a patent fell outside the “narrow category of cases implicating uniquely important individual liberty interests,” which “warrant a heightened standard of proof.”¹¹⁴

Second, Microsoft argued in the alternative that the preponderance of the evidence standard was required at least when the party challenging the patent’s validity relied on prior art references not considered by the USPTO.¹¹⁵ Even if a heightened standard of proof was warranted out of deference to the USPTO’s decision to grant a patent, this deference was appropriate only when the Patent Examiner had actually considered the relevant evidence regarding patentability.¹¹⁶ If the examiner had not considered some of the relevant prior art, then, as the Court stated in *KSR v. Teleflex*, “the rationale underlying the presumption of . . . validity — that the [US]PTO, in its expertise, has approved the claim — seems much diminished”¹¹⁷

In response, *i4i* argued that prior to the adoption of the 1952 Patent Act, the Supreme Court had “repeatedly and consistently held that the presumption of patent validity imposed a heightened standard to prove invalidity.”¹¹⁸ Section 282, *i4i* contended, codified this common-law precedent.¹¹⁹ In addition, *i4i* argued that the Federal Circuit had correctly required a heightened standard for proving invalidity because such a standard both promoted innovation by “foster[ing] strong, stable patent rights” and “properly incorporate[d] deference to the [US]PTO.”¹²⁰

The Supreme Court rejected both of Microsoft’s proposals.¹²¹ It first analyzed the statutory text of Section 282 of the Patent Act. The Court concluded that although the statute clearly provided that the party challenging a patent claim’s validity bore the burden of proof,

112. *i4i*, 131 S. Ct. at 2244.

113. Brief for Petitioner at 14–15, *i4i*, 131 S. Ct. 2238 (2011) (No. 10-290).

114. *Id.* at 8, 16–18.

115. *i4i*, 131 S. Ct. at 2244, 2249.

116. Brief for Petitioner, *supra* note 113, at 51–53.

117. *Id.* at 2 (quoting *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 426 (2007)).

118. Brief for Respondents at 12, *i4i*, 131 S. Ct. 2238 (2011) (No. 10-290) (citing, *inter alia*, *Smith v. Hall*, 301 U.S. 216, 233 (1937), and *Radio Corp. of Am. v. Radio Eng’g Labs., Inc.*, 293 U.S. 1, 2 (1934)).

119. *Id.* at 15 (“[B]y 1952 this Court had made clear that the common-law presumption of validity imposed a clear-and-convincing standard of proof. Congress’s use of these terms in § 282 means the Court ‘must infer, unless the statute otherwise dictates, that Congress mean[t] to incorporate’ that standard.” (alteration in original) (quoting *Neder v. United States*, 527 U.S. 1, 21 (1999))).

120. *Id.* at 30, 39.

121. *i4i*, 131 S. Ct. at 2244.

the statute was silent on the standard of proof.¹²² The Supreme Court was clear that the statute's use of "burden of proof" referred to the "burden of persuasion."¹²³ The statute also specified that patents are "presumed valid."¹²⁴ The Court further agreed with *i4i*'s argument that the phrase "presumed valid" had a settled meaning at the time of Section 282's adoption that incorporated a heightened standard of proof.¹²⁵ The Supreme Court relied heavily upon its 1934 decision in *Radio Corporation of America v. Radio Engineering Laboratories, Inc.* ("RCA").¹²⁶ In *RCA*, Justice Cardozo's majority opinion held that a patent challenger "bears a heavy burden of persuasion, and fails unless his evidence has more than a dubious preponderance."¹²⁷ As a result, the "dubious" preponderance standard was insufficient to find a patent invalid.¹²⁸

Older Supreme Court decisions used similarly unequivocal language to describe the standard of proof for finding a patent invalid. For instance, in 1874, the Court described the burden of proof relating to prior inventorship as "rest[ing] upon [the defendant], and every reasonable doubt should be resolved against him."¹²⁹ Relying upon these cases, the Supreme Court held in *i4i* that the term "presumption of validity" had a settled meaning when the Patent Act was adopted in 1952 that required the patent challenger to satisfy this heightened standard of proof by clear and convincing evidence.¹³⁰

The Court also addressed Microsoft's alternative argument that the preponderance standard should apply at least where evidence at trial was not provided to the USPTO during examination of the patent. The Court was reluctant to adopt a fluctuating standard that might require different burdens of proof within a single lawsuit.¹³¹ However, the Court acknowledged that numerous courts of appeals before the 1952 Act stated that the presumption of validity was "weakened" or "dissipated" when the evidence had not previously been considered by the USPTO.¹³² Despite this acknowledgement, the Court rejected the

122. *Id.* at 2245 ("[W]hile [35 U.S.C. § 282] explicitly specifies the burden of proof, it includes no express articulation of the standard of proof."); *see also id.* at 2254 (Thomas, J., concurring in the judgment) (noting that "§ 282 is silent as to the standard of proof").

123. *Id.* at 2245 n.4.

124. 35 U.S.C. § 282 (2006 & Supp. V 2012).

125. *i4i*, 131 S. Ct. at 2245 ("[B]y stating that a patent is 'presumed valid,' § 282, Congress used a term with a settled common-law meaning.")

126. 293 U.S. 1 (1934).

127. *Id.* at 8.

128. *Id.*

129. *Coffin v. Ogden*, 85 U.S. (18 Wall.) 120, 124 (1874).

130. *i4i*, 131 S. Ct. at 2246.

131. *Id.* at 2250.

132. *Id.* at 2250–51 (citing *Jacuzzi Bros., Inc. v. Berkeley Pump Co.*, 191 F.2d 632, 634 (9th Cir. 1951) ("largely dissipated"); *Gillette Safety Razor Co. v. Cliff Weil Cigar Co.*, 107 F.2d 105, 107 (4th Cir. 1939) ("greatly weakened"); *Butler Mfg. Co. v. Enterprise Cleaning*

view that these statements meant that the standard of proof should be anything but clear and convincing.¹³³ Thus, the Court held that a patent claim must be proven invalid by clear and convincing evidence in all circumstances.

However, and important for our study, the Court decided that new evidence in invalidity challenges was entitled to different treatment in jury instructions.¹³⁴ The Court noted that previously unconsidered evidence may “carry more weight” than evidence previously reviewed by the USPTO.¹³⁵ While the burden of proof always remains clear and convincing, the Court noted that a special instruction may be provided to the jury in these circumstances:

[I]f the [US]PTO did not have all material facts before it, its considered judgment may lose significant force. And, concomitantly, the challenger’s burden to persuade the jury of its invalidity defense by clear and convincing evidence may be easier to sustain. In this respect, although we have no occasion to endorse any particular formulation, we note that a jury instruction on the effect of new evidence can, and when requested, most often should be given. When warranted, the jury may be instructed to consider that it has heard evidence that the [US]PTO had no opportunity to evaluate before granting the patent.¹³⁶

Thus, the Court urged district court judges to provide a special jury instruction when new evidence regarding invalidity was first presented in litigation.¹³⁷ While the Court apparently felt that the burden of proof always remained clear and convincing, this language suggested the burden might be somewhat easier to satisfy in these circumstances.

Co., 81 F.2d 711, 716 (8th Cir. 1936) (“weakened”); *H. Schindler & Co. v. C. Saladino & Sons*, 81 F.2d 649, 651 (1st Cir. 1936) (“weakened”) (internal quotation marks omitted).

133. *i4i*, 131 S. Ct. at 2251 (“[W]e cannot read these cases to hold or even to suggest that a preponderance standard would apply in such circumstances, and we decline to impute such a reading to Congress.”).

134. We recognize that some may interpret the *i4i* decision as altering the weight given to previously unconsidered evidence. While this does not formally alter the standard of proof as a theoretical matter, it may be seen as performing an equivalent function. In other words, while the Court held that the standard remained unchanged, it raised the value of this unconsidered evidence such that the burden could be overcome more easily. Regardless of how *i4i* is interpreted, we have based our experiment upon model instructions that are used to explain the standard of proof to juries in patent litigation.

135. *i4i*, 131 S. Ct. at 2251 (quoting *Am. Hoist & Derrick Co. v. Sowa & Sons, Inc.*, 725 F.2d 1350, 1360 (Fed. Cir. 1984)) (internal quotation marks omitted).

136. *i4i*, 131 S. Ct. at 2251 (citation omitted).

137. *Id.*

IV. METHODOLOGY AND STUDY DESIGN

A. Hypotheses About *i4i*'s Impact

The impact of changing the standard of proof on jurors' decisions in civil cases is uncertain. More specifically, in the area of patent law, intellectual property scholars and others have disagreed about the effect of changing the standard for proving a patent invalid. For example, Doug Lichtman and Mark Lemley argue that "[w]hile we can't prove that presumptions matter, we believe that they likely do, at least at the margins."¹³⁸ However, they admit that "this is an area of uncertainty," and the legal community "know[s] far less than [it] should about how presumptions affect litigation decisions."¹³⁹

Judge William Alsup, who has presided over numerous patent cases in the Northern District of California, has called the clear and convincing standard a "legal earthwork fortified by a protective moat" that creates "a huge advantage for the patent holder" in validity challenges.¹⁴⁰ Similarly, in a brief filed in the *i4i* case, Apple and Intel — which both own large patent portfolios and appear as frequent defendants in infringement litigation — argued that the clear and convincing evidence standard "has a compelling effect on jurors" and "insulate[s] weak patents from invalidity verdicts."¹⁴¹

Mark Janis has asserted that "the precise verbal formulation that we use for the standard of evidence for overcoming the presumption" of patent validity may be at least as important as the standard itself.¹⁴² As a result, he argues that "one possible outcome of . . . the preponderance standard for overcoming the presumption of patent validity is that [it] will cause little difference in the outcomes of cases."¹⁴³

Based on the assumption that standards of proof matter, we developed three hypotheses for our experiment:

138. Doug Lichtman & Mark A. Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45, 69–70 (2007); see also *id.* at 70 ("We don't know exactly how often the presumption makes a difference to a case outcome. But we are confident that it does in at least some cases . . .").

139. *Id.* at 69.

140. William Alsup, *A District Judge's Proposal for Patent Reform: Revisiting the Clear and Convincing Standard and Calibrating Deference to the Strength of the Examination*, 24 BERKELEY TECH. L.J. 1647, 1648 (2009).

141. Brief of Apple Inc. and Intel Corp., *supra* note 17, at 15, 19.

142. Mark D. Janis, *Reforming Patent Validity Litigation: The "Dubious Preponderance"*, 19 BERKELEY TECH. L.J. 923, 927 (2004); see also *id.* at 935 (asserting that "merely changing the language of the patent law standard to 'preponderance' by no means ensures that courts will converge around a uniform approach to assessing patent validity evidence").

143. *Id.* at 927; see also Paul F. Morgan, *Guest Post: Microsoft v. i4i — Is the Sky Really Falling?*, PATENTLYO (Jan. 9, 2011), <http://www.patentlyo.com/patent/2011/01/microsoft-v-i4i-is-the-sky-really-falling.html> (expressing skepticism that "the very small percentage of patents now being held invalid by juries would somehow greatly increase due to this one potential change [to a preponderance standard] in jury instructions").

1. A clear and convincing standard will result in fewer findings of invalidity than a preponderance standard.
2. A clear and convincing standard will result in fewer findings of invalidity than the same standard combined with an *i4i*-type instruction.
3. A clear and convincing standard combined with an *i4i*-type instruction will result in fewer findings of invalidity than a preponderance standard.

As explained in the following section, we used an experimental approach to test each of these hypotheses.

B. Why an Experiment?

Using an experimental methodology to measure the effect of varying the standard of proof has many important advantages. There is burgeoning literature about experiments conducted by legal scholars.¹⁴⁴ Experiments permit the manipulation of a single variable to determine if changes in that variable affect other variables.¹⁴⁵ Like a classic double-blind medical experiment, this method permits us to randomly assign the change, in our case the standard of proof, while holding everything else constant. Any differences in results between conditions thus can be attributed to the randomly assigned variable, the standard of proof.¹⁴⁶

We decided against a common approach used by empirical legal scholars: studying observational data.¹⁴⁷ Studying standards of proof

144. For examples of legal scholarship based upon experiments, see CASS R. SUNSTEIN ET AL., PUNITIVE DAMAGES: HOW JURIES DECIDE (2002); Christopher Buccafusco & Christopher J. Sprigman, *The Creativity Effect*, 78 U. CHI. L. REV. 31 (2011); Christopher Buccafusco & Christopher Sprigman, *Valuing Intellectual Property: An Experiment*, 96 CORNELL L. REV. 1 (2010); Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007); Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 OHIO ST. L.J. 1391 (2006); Andrew W. Torrance & Bill Tomlinson, *Patents and the Regress of Useful Arts*, 10 COLUM. SCI. & TECH. L. REV. 130 (2009).

145. See ROBERT M. LAWLESS, JENNIFER K. ROBBENOLT & THOMAS S. ULEN, EMPIRICAL METHODS IN LAW 93 (2010) (“Experiments are particularly useful when the goal of the research project is to explore one or more causes of a phenomenon. Indeed, the purpose of an experiment is to isolate the effect of one variable on another.”).

146. *Id.* at 94 (explaining that isolating the “specific variable of interest” in an experiment addresses the “question[] of cause and effect” by “eliminat[ing] . . . confounding variables”).

147. For examples of empirical scholarship based on observational data in patent litigation, see John R. Allison, Mark A. Lemley & Joshua Walker, *Patent Quality and Settlement Among Repeat Patent Litigants*, 99 GEO. L.J. 677 (2011); Christopher A. Cotropia, *Nonobviousness and the Federal Circuit: An Empirical Analysis of Recent Case Law*, 82 NOTRE

by relying on observational data is extremely difficult. Unlike experiments, observational data is not controlled by the researcher.¹⁴⁸ Instead, the researcher records the data after it occurs naturally.¹⁴⁹ The fact that the studied conditions are not randomly assigned makes it more difficult to draw causal inferences from observational data than from experiments.¹⁵⁰

For our particular research question, we believe that an experimental approach offers several advantages compared to observational data generated in litigation. First, the courts do not frequently change the standard of proof. Consequently, almost all cases in a given area of law are decided using the same standard. In particular, if we wanted to rely upon observational data to analyze the effect, if any, of the Supreme Court's decision regarding the standard for proving a patent invalid, we would need to wait for a sufficient number of trials to occur, which would likely take years.¹⁵¹ Furthermore, observational data from litigation would not be available for one of the possible standards of proof — preponderance of the evidence — because this standard was explicitly rejected by the Court.

Second, and perhaps more importantly, if the standard of proof is changed, then other features of the cases may change as well. For instance, different cases may settle before trial if the ultimate question for the jury is based upon a clear and convincing standard of proof as

DAME L. REV. 911 (2007); Kimberly A. Moore, *Forum Shopping in Patent Cases: Does Geographic Choice Affect Innovation?*, 79 N.C. L. REV. 889, 901 (2001); Kimberly A. Moore, *Judges, Juries, and Patent Cases — An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365, 367 (2000) [hereinafter Moore, *Black Box*]; Lee Petherbridge, Jason Rantanen & Ali Mojibi, *The Federal Circuit and Inequitable Conduct: An Empirical Assessment*, 84 S. CAL. L. REV. 1293 (2011); Lee Petherbridge, *Patent Law Uniformity?*, 22 HARV. J.L. & TECH. 421 (2008); David L. Schwartz, *Courting Specialization: An Empirical Study of Claim Construction Comparing Patent Litigation Before Federal District Courts and the International Trade Commission*, 50 WM. & MARY L. REV. 1699 (2009); David L. Schwartz, *Explaining the Demise of the Doctrine of Equivalents*, 26 BERKELEY TECH. L.J. 1157 (2011); David L. Schwartz, *Practice Makes Perfect? An Empirical Study of Claim Construction Reversal Rates in Patent Cases*, 107 MICH. L. REV. 223 (2008); Christopher B. Seaman, *Willful Patent Infringement and Enhanced Damages After In Re Seagate: An Empirical Study*, 97 IOWA L. REV. 417 (2012); R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105 (2004).

148. See Thomas E. Willging, *Past and Potential Uses of Empirical Research in Civil Rulemaking*, 77 NOTRE DAME L. REV. 1121, 1132 (2002) (“Observational research differs from experimental research in that any testing of the effect of a treatment does not take place under conditions that the researcher controls.”).

149. See LAWLESS, ROBBENOLT & ULEN, *supra* note 145, at 126 (“Archival data has the benefit of being drawn from real-world sources.”).

150. See Adam M. Samaha, *Randomization in Adjudication*, 51 WM. & MARY L. REV. 1, 23 (2009) (explaining that the “best empirical studies on causation often rely on random assignments to treatment and control groups”).

151. See, e.g., Seaman, *supra* note 147 (conducting a study of willful patent infringement and enhanced damages decisions in the district courts for a three-year period after the Federal Circuit's en banc decision in *In re Seagate Techs. LLC*, 497 F.3d 1360 (Fed. Cir. 2007)).

opposed to a mere preponderance standard.¹⁵² In other words, changing the standard of proof may also result in different cases being selected for resolution by juries. Priest and Klein famously argue that regardless of the law, cases where there is a high degree of certainty regarding the outcome will settle, so those cases that are tried are always the closest cases, the 50-50 cases.¹⁵³ Regardless of whether selection effects actually result in a 50-50 split for the resolution of individual issues like patent invalidity,¹⁵⁴ any significant change in the types of cases that reach a jury can make it more difficult to unpack causal relationships. An experimental approach overcomes these problems by permitting the researcher to control for all variables other than the one being studied.¹⁵⁵

152. See Fredrick E. Vars, *Toward a General Theory of Standards of Proof*, 60 CATH. U. L. REV. 1, 34 (2010) (suggesting that in some circumstances a “clear-and-convincing standard of proof could reduce uncertainty and thereby facilitate settlement” of some claims, although “a higher percentage of cases that are close to the heightened proof threshold . . . [may] fail to settle”).

153. See George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 4–6 (1984).

154. For an argument that the Priest-Klein hypothesis only applies to the selection of disputes, not the selection of individual issues in patent litigation, see Jason Rantanen, *Why Priest-Klein Cannot Apply to Individual Issues in Patent Cases* (Univ. of Iowa Legal Studies, Research Paper No. 12-15, 2012), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2132810.

155. A recent article about our topic, the effect of the *i4i* decision on patent litigation, exemplifies some of the difficulties of using observational data from litigation to study standards of proof. See Etan S. Chatlynnne, Stephen Kenny & Lucas Watkins, *Investigating Patent Law’s Presumption of Validity, Part II: An Empirical Analysis of How Unconsidered Evidence and Evidentiary Standards Affect Jury Verdicts*, 2011 CARDOZO L. REV. DE NOVO 46 (2011), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1799706. In that study, the researchers compared a first set of patent cases decided from 1966 until 1982 to a second set of patent cases decided from 2008 until 2010. *Id.* at 48–49. The authors assert that the first set of cases was from a time period when courts used a preponderance standard when the party challenging validity introduced evidence that had not been considered by the USPTO, and the second set of cases was when courts used a clear and convincing standard. *Id.* at 48–50. The authors failed to find a significant difference in outcomes between the two time periods and therefore concluded that the standard of proof — preponderance or clear and convincing — “may not be as significant a driver in invalidity outcomes as some have suggested.” See *id.* at 54 (finding that challengers introducing new evidence prevailed under a preponderance standard 32% of the time from 1966 to 1982 and prevailed under a clear and convincing evidence standard 34% of the time from 2008 to 2010).

However, there are several reasons to view the results of this prior study with caution, as the authors themselves note. See *id.* at 54–55 (“The comparison herein of data from periods separated by more than twenty-five years surely implicates additional factors such as changes in applicable practice, procedure, search technology and much of the surrounding body of legal rules . . .”). In particular, numerous major changes occurred in patent litigation between the two time periods studied that would result in strong selection effects. One is the Federal Circuit’s 1995 en banc decision in *Markman v. Westview Instruments, Inc.*, which held that claim construction — the process of interpreting the scope of the patent — was a matter of law for the court. 52 F.3d 967, 970–71 (Fed. Cir. 1995), *aff’d*, 517 U.S. 370 (1996). Another is the rise of jury trials in patent cases. Before 1982, nearly all patent cases were tried to judges; starting in the 1990s, however, the majority of patent cases reaching trial were decided by juries. See Moore, *Black Box*, *supra* note 147, at 366 (finding that from 1968 to 1970, juries tried only 2.8% of patent cases; in contrast, from 1997 to 1999,

C. Study Design

After deciding upon an experimental methodology, we turned to study design. We desired a patent litigation scenario that was understandable to subjects, so we shied away from describing a dispute involving complex technology, such as biotechnology, pharmaceuticals, or computer software. Thus, we created a one-and-a-third page fact pattern based upon a real case (not *i4i*) involving a relatively simple technology — golf ball design.¹⁵⁶ We also wanted a fact pattern without a clear answer in law. In other words, we wanted the subjects to have some discretion on which way they decided the case. The case selected as the basis for our fact pattern provided this, as it involved two separate jury verdicts that reached opposite conclusions regarding invalidity,¹⁵⁷ and the Federal Circuit on appeal stated that “the evidence at trial was such that the jury could have rationally reached either verdict with regard to [the invalidity of] the asserted claims.”¹⁵⁸

The same fact pattern and summary of the arguments were provided to all subjects using a between-subjects design.¹⁵⁹ We limited the fact pattern to a single issue: whether the patent was invalid because it was obvious. Under the law, a patent is obvious if the differences between the prior art and the invention would have been apparent to a person in the field at the time the invention was made.¹⁶⁰ To focus our experiment on the effect of jury instructions, we randomly assigned the subjects one of three different jury instructions, as described in more detail below. A copy of the fact pattern, summary

juries tried 59% of patent cases). Third, summary judgment practice radically changed between 1966 and 2009, most notably with the *Celotex* trilogy of opinions issued in 1986, thus affecting which cases ultimately reach trial. See *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986); *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242 (1986); *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986).

156. *Callaway Golf Co. v. Acushnet Co.*, 576 F.3d 1331 (Fed. Cir. 2009).

157. In *Callaway*, the jury in the initial trial found that one dependent claim (claim 5) of the patent-in-suit was invalid for obviousness, but the remaining eight claims — including the independent claim (claim 4) on which the invalid dependent claim was based — were not proven invalid. See *id.* at 1337. On appeal, the Federal Circuit vacated the judgment and remanded for a new trial on obviousness, finding that “the verdict form returned by the jury reflects an irreconcilable inconsistency” regarding invalidity. *Id.* at 1345. After retrial, the second jury found all asserted claims invalid as anticipated and obvious. *Callaway Golf Co. v. Acushnet Co.*, 778 F. Supp. 2d 487, 491 (D. Del. 2011). The District Court denied the patentee’s motion for judgment as a matter of law, holding that the second jury’s invalidity decision was supported by substantial evidence. *Id.* at 505.

158. *Callaway Golf Co.*, 576 F.3d at 1344.

159. See generally Robert A. Prentice, *Chicago Man, K-T Man, and the Future of Behavioral Law and Economics*, 56 VAND. L. REV. 1663, 1680 (2003) (“In between-subjects experiments, some subjects are tested under condition A and their responses are compared to those of subjects tested under condition B.”).

160. 35 U.S.C. § 103 (2006 & Supp. V 2012); *KSR Int’l Co. v. Teleflex, Inc.*, 550 U.S. 398, 399 (2007). Technically, the claims of a patent are determined obvious (or not obvious) on an individual, claim-by-claim basis. See, e.g., *Amazon.com, Inc. v. Barnesandnoble.com, Inc.*, 239 F.3d 1343, 1351 (Fed. Cir. 2001).

of the parties' arguments, and the jury instructions used in the experiment are reproduced in Appendix A. The sequence of steps in the experiment is reproduced in Figure 1 below.

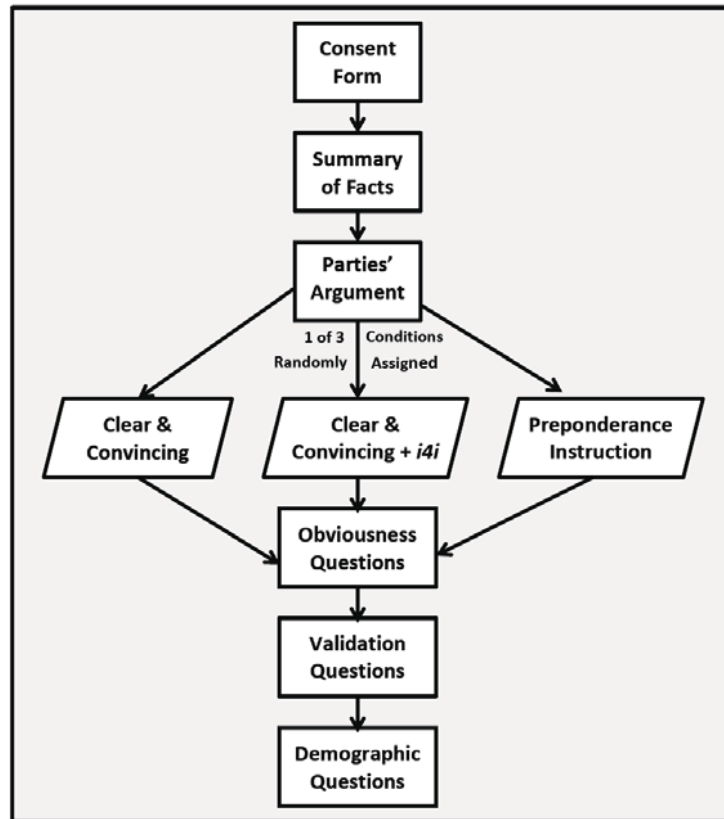


Figure 1: Steps of the Experiment

Our fact pattern involved a dispute regarding patent infringement between two fictitious competitors, Acme Golf, Inc. (“Acme”) and Bravo Sporting Equipment Co. (“Bravo”).¹⁶¹ Acme’s invention related to a new “three-piece” golf ball which combined the features of

161. We used two competitors as the hypothetical litigants in our fact pattern to avoid a potential bias against non-practicing entities (“NPEs”), which have been the subject of significant controversy in patent law. *See generally* Colleen V. Chien, *Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents*, 87 N.C. L. REV. 1571, 1573–77 (2009); Gerard N. Magliocca, *Blackberries and Barnyards: Patent Trolls and the Perils of Innovation*, 82 NOTRE DAME L. REV. 1809, 1818–19 (2007); Michael Risch, *Patent Troll Myths*, 42 SETON HALL L. REV. 457, 458 (2012) (contending that much of “the conventional wisdom about patent trolls” is inaccurate and that “patents enforced by so-called trolls — and the companies that obtained them — look a lot like other litigated patents and their owners”).

prior art “two-piece” golf balls.¹⁶² The fact pattern explained that when the USPTO examined and allowed the application for a patent, it had only been aware of several two-piece golf balls but no three-piece golf balls. However, during the course of litigation, these fictitious parties discovered a prior patent (issued to an inventor named “Charles”) that disclosed a three-piece golf ball. This prior patent was different from the Acme invention (it used different materials for some of the layers), but it was closer than any of the prior art considered by the USPTO. Thus, as in the *i4i* case, a key piece of prior art had not been considered by the USPTO during examination and was offered as evidence in litigation by the party challenging the patent’s validity.

The scenario concluded with a paragraph describing each of Acme’s and Bravo’s arguments regarding the issue of patent obviousness.¹⁶³ These arguments are ones commonly made by litigants in patent trials. For the patent owner, the summary stated that no individual piece of prior art taught all of the elements of the claimed invention, that the invention had benefits over the prior art, that the Charles patent was directed to a very different problem than the Acme patent, and that the jury should defer to the “trained Patent Examiner, who was in the best position to determine whether Acme’s claimed invention was obvious.”¹⁶⁴ For the patent challenger, the summary argued that the Acme invention was merely a combination of preexisting items, that the invention’s modification of the prior art was trivial, and that the jury should not defer to the Patent Examiner because the patent would not have been granted if the Examiner had known of the Charles patent.¹⁶⁵

After the fact pattern and summary of the parties’ arguments, each subject received one of three randomly assigned jury instructions regarding the legal standard for determining whether the patent was invalid. The one-page jury instructions varied only regarding the standard of proof; they contained identical descriptions regarding obviousness in patent law.¹⁶⁶ For the standard of proof, three different

162. This is modeled on the statement of the District Court in *Callaway*:

[G]olf balls are typically identified as two-piece or three-piece balls. Two-piece balls have a core, which is either solid or “wound,” and an outer layer . . . Three-piece balls have an additional layer covering the core, so that the ball is characterized as having a core, an inner cover layer and an outer cover layer.

Callaway Golf Co. v. Acushnet Co., 778 F. Supp. 2d 487, 488–89 (D. Del. 2011).

163. All subjects received the summary of the arguments in the same order: first Bravo (the patent challenger), then Acme (the patent holder). This was done to mimic the order of closing arguments at a trial limited to validity, where the party with the burden of persuasion — here, Bravo — typically goes first.

164. See *infra* Appendix A.

165. See *infra* Appendix A.

166. Our jury instruction was adapted from several model jury patent instructions that were selected due to their perceived ease of comprehension as well as their conciseness. The

instructions were given: (1) clear and convincing evidence; (2) clear and convincing evidence with an additional instruction based upon the Court's decision in *i4i*; and (3) preponderance of the evidence. We adapted language from model patent jury instructions promulgated by the American Intellectual Property Law Association ("AIPLA")¹⁶⁷ and the National Jury Instruction Project to explain these standards.¹⁶⁸ Specifically, the instructions regarding these conditions stated:

Clear and Convincing Evidence:

Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases.

Clear and Convincing Evidence with *i4i*-type Instruction:

Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a pre-

first paragraph of the instruction, *see infra* Appendix A ("The fact that the USPTO grants a patent . . ."), was adapted from the Federal Circuit Bar Association's model patent jury instructions. FED. CIR. BAR ASS'N, MODEL PATENT JURY INSTRUCTION § A.1 (2012), *available at* <http://memberconnections.com/olc/filelib/LVFC/cpages/9004/Library/2012%20Updated%20FCBA%20Model%20Patent%20Jury%20Instructions.pdf> (preliminary instructions). The second paragraph of the instruction, *see infra* Appendix A ("Under the law, Acme's patent is presumed to be not obvious . . ."), was adapted from the U.S. Court of Appeals for the Seventh Circuit's model jury instructions for patent cases. *See* COMM. ON PATTERN CIVIL JURY INSTRUCTIONS OF THE SEVENTH CIR., FEDERAL CIVIL JURY INSTRUCTIONS OF THE SEVENTH CIR. §§ 11.1.14, 11.3.1 (2009), *available at* http://www.ca7.uscourts.gov/Pattern_Jury_Instr/7th_civ_instruc_2009.pdf (invalidity due to obviousness and validity generally). The third paragraph of the instruction, which contained the randomly selected condition regarding the standard of proof, *see infra* Appendix A ("To succeed on its claim . . ."), was adapted from the American Intellectual Property Law Association's model patent jury instructions and the National Jury Instruction Project's model patent jury instructions. *See infra* notes 167–168 and accompanying text. The final paragraph of the instruction, *see infra* Appendix A ("In deciding obviousness . . ."), was also adapted from the Seventh Circuit's model jury instructions for patent cases. COMM. ON PATTERN CIVIL JURY INSTRUCTIONS OF THE SEVENTH CIR., FEDERAL CIVIL JURY INSTRUCTIONS OF THE SEVENTH CIR. § 11.3.6 (2009), *available at* http://www.ca7.uscourts.gov/Pattern_Jury_Instr/7th_civ_instruc_2009.pdf (obviousness).

167. AM. INTELLECTUAL PROP. LAW ASS'N, AIPLA'S MODEL PATENT JURY INSTRUCTIONS § 7 (2008), *available at* http://www.aipla.org/learningcenter/library/books/other-pubs/Documents/2008_03_27_AIPLA_Model_Jury_Instructions.doc (obviousness).

168. NAT'L JURY INSTRUCTION PROJECT, MODEL PATENT JURY INSTRUCTIONS §§ 1.4, 1.5 (2009), *available at* <http://www.nationaljuryinstructions.org/documents/NationalPatentJuryInstructions.pdf> (burden of proof — clear and convincing evidence and preponderance of the evidence).

ponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases.

The burden of proving obviousness is more easily satisfied when, as in this case, the prior art on which the claim of obviousness is based was not considered by the Examiner.¹⁶⁹

Preponderance of the Evidence:

A preponderance of the evidence is evidence that shows it is more probable than not that the patent was obvious. This is a lower standard of proof than either clear or convincing evidence or the beyond a reasonable doubt standard used in criminal cases.

After reviewing the jury instructions, the subjects were presented with several questions.¹⁷⁰ First, we asked subjects for their ultimate decision on invalidity.¹⁷¹ For subjects who received the first two conditions, we asked whether the patent challenger had proven by clear and convincing evidence that the patent was obvious. For subjects

169. Specifically, for this condition, the following sentence was added to the clear and convincing instruction: "The burden of proving obviousness is more easily satisfied when, as in this case, the prior art on which the claim of obviousness is based was not considered by the Examiner." This sentence was adapted from Microsoft's proposed instruction in another case, which would have instructed the jury that "the burden to prove invalidity 'is more easily carried when the references on which the assertion [of invalidity] is based were not directly considered by the examiner during prosecution.'" *Petition for a Writ of Certiorari at 12, Microsoft Corp. v. z4 Techs., Inc.*, 507 F.3d 1340 (Fed. Cir. 2008) (No. 07-1243) (alteration in original).

170. In the survey program (Qualtrics), respondents were not permitted to return to the fact pattern, summary of parties' arguments, or jury instructions after they began answering questions. In real trials, jurors are frequently not provided written copies of the jury instructions for use in deliberations. *Cf. United States v. Russo*, 110 F.3d 948, 953 (2d Cir. 1997) (holding that "the decision whether to submit written instructions to the jury properly lies within the discretion of the trial court").

171. "The ultimate judgment of obviousness is a legal determination" based on the resolution of underlying factual issues. *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S. 398, 427 (2007) (citing *Graham v. John Deere Co.*, 383 U.S. 1, 17 (1965)). In the authors' experience, the ultimate decision on obviousness is routinely submitted to juries, although the court is empowered to evaluate "the [ultimate] legal conclusion [of obviousness] de novo" during post-trial motions "to see whether it is correct in light of the presumed jury fact findings." *Kinetic Concepts, Inc. v. Smith & Nephew, Inc.*, 688 F.3d 1342, 1356-57 (Fed. Cir. 2012) (alteration in original) (quoting *Jurgens v. McKasy*, 927 F.2d 1552, 1557 (Fed. Cir. 1991)) (internal quotation marks omitted). *See also* John Guo, Note, *Special Verdicts: An Obvious Trial Procedure for Deciding Obviousness in Patent Litigation*, 40 SW. L. REV. 513, 515 (2011) ("[T]he Court of Appeals of the Federal Circuit reviews the question of obviousness 'de novo', and it reviews the underlying jury's factual findings to determine whether there is 'substantial evidence' to support the jury's findings." (footnote omitted)).

who received the preponderance condition, we asked the same question but substituted “preponderance of the evidence” for “clear and convincing.” Second, subjects were asked about their level of confidence in their answer regarding invalidity on a 1–7 scale (1 = least confident, 7 = most confident). A third question asked the subjects to indicate “how likely do you think it is that Acme’s patent is obvious” on a 0–100% scale (in 10% increments).

On the next page of the experiment, we included three validation questions designed to exclude subjects who had not read and/or adequately understood the fact pattern.¹⁷² On the final page, the study concluded with several demographic questions about the subject’s age, race, gender, education, U.S. citizenship, patent experience, and golf experience.

We ran a preliminary study to validate the language used. Approximately ten people read and completed the experiment. We discussed the experiment with these testers and made minor linguistic adjustments to improve comprehension. Subsequently, we conducted a pilot study on incoming first-year law students at Chicago-Kent College of Law during orientation in August 2011.¹⁷³ After the pilot study, we believed that the study was understandable to subjects.

We conducted the experiment upon members of the general public eligible for U.S. jury duty in November 2011 ($N = 500$).¹⁷⁴ Subjects were recruited through Amazon’s Mechanical Turk,¹⁷⁵ a website that allows participants to complete Human Intelligence Tasks for compensation. In the initial iteration, we compensated subjects \$2 for successfully completing the experiment.¹⁷⁶ Recent studies have used

172. Approximately 7% of subjects (35 of 535) were excluded from our results presented in Part V for incorrectly answering one or more validation questions. To avoid priming subjects to pay close attention to the standard of proof, the validation questions were directed to basic information from the fact pattern. The questions did not relate to the standard of proof.

During a workshop, a colleague suggested that real patent jurors sometimes do not pay attention at trial, and including subjects who failed the validation questions was appropriate to mirror the real juror population. While we believe that our approach is more methodologically sound, we note that including the responses of such subjects does not materially affect any of our results.

173. In the pilot study, there was no statistically significant difference in outcomes based on the burden of proof ($p = 0.826$). The reason for the differences in results between the pilot and this study is unclear. It may be due to the relatively small sample size of the pilot study’s population ($N = 150$), which made it difficult to obtain a statistically significant result. It also could be due to demographic differences between first-year law students and members of the general public, although we do not have a well-formed hypothesis as to why this would be the case.

174. Approval from the Illinois Institute of Technology’s (“IIT”) Internal Review Board (“IRB”), where both authors worked at the time of the experiments, was obtained prior to conducting all experiments.

175. AMAZON MECHANICAL TURK, <https://www.mturk.com> (last visited May 9, 2013).

176. We decided to compensate participants \$2 for completing the survey in the first iteration because we anticipated that it would take approximately 20–30 minutes to complete. We received all 500 responses for the first iteration in less than 24 hours. For the second

Mechanical Turk to successfully replicate the results of well-known social science experiments, including the prisoner's dilemma, providing evidence that subjects recruited through Mechanical Turk "behave similarly to subjects in physical laboratories."¹⁷⁷ Research using Mechanical Turk-based experiments has been published and presented in a number of academic fields, including peer-reviewed journals, which suggests that it meets scholarly expectations for quality web-based experiments.¹⁷⁸

In an attempt to mirror jury-eligible persons, we limited the subjects to those within the United States who were U.S. citizens,¹⁷⁹ who understood English, and who were at least eighteen years old.¹⁸⁰ Subjects who wished to participate were redirected to the experiment, which was hosted by Qualtrics, a widely used online survey program.¹⁸¹ On average, subjects spent seventeen minutes completing the

iteration, we reduced the compensation to \$1 and received 600 responses in less than 48 hours.

177. John J. Horton, David G. Rand & Richard J. Zeckhauser, *The Online Laboratory: Conducting Experiments in a Real Labor Market*, 14 J. EXPERIMENTAL ECON. 399, 406 (2011); see also John Bohannon, *Social Science for Pennies*, 334 SCI. 307 (2011).

178. See Adam J. Berinsky, Gregory A. Huber & Gabriel S. Lenz, *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351, 354–61, 365–66 (2012) (assessing the internal and external validity of experimental research conducted using the Mechanical Turk platform, finding the "MTurk subject pool is no worse than convenience samples used by other researchers in political science"). Studies relying on Mechanical Turk to conduct experiments include Janice Nadler & Mary-Hunter McDonnell, *Moral Character, Motive, and the Psychology of Blame*, 97 CORNELL L. REV. 255, 274, 280, 285 (2012), and Paul H. Robinson, Geoffrey P. Goodwin & Michael D. Reisig, *The Disutility of Injustice*, 85 N.Y.U. L. REV. 1940, 1998, 2004 (2010). Studies using other online recruiting methods include Yany Grégoire, Thomas M. Tripp & Renaud Legoux, *When Customer Love Turns into Lasting Hate: The Effects of Relationship Strength and Time on Customer Revenge and Avoidance*, 73 J. MARKETING 18, 21 (2009) (recruiting participants from consumer complaint forums) and Francis X. Shen et al., *Sorting Guilty Minds*, 86 N.Y.U. L. REV. 1306, 1334 (2011) (using Qualtrics to recruit participants).

179. Mechanical Turk allows requestors to screen subjects by country of origin. We are grateful to Matthew Sag, our discussant at the 7th Annual Conference on Empirical Legal Studies ("CELS"), for checking the IP addresses for all 500 of our subjects to determine their countries of origin. We used Plot IP to further investigate the country of origin of questionable IP addresses that Sag identified. PLOT IP, <http://www.plotip.com> (last visited May 9, 2013). After analysis, we believe that 3.6% of subjects in our study (18 of 500) apparently accessed the survey from IP addresses from outside the United States, with one additional subject from each of the U.S. Virgin Islands and Puerto Rico. This is consistent with the literature regarding other experiments using Mechanical Turk. See David G. Rand, *The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments*, 299 J. THEORETICAL BIOLOGY 172, 176 (2012) (finding that Mechanical Turk users' self-reported country of residence was over 97% accurate based on IP addresses). Excluding these non-U.S. subjects does not affect our results.

180. Age and citizenship status were self-reported.

181. QUALTRICS, <http://www.qualtrics.com> (last visited May 9, 2013); see also Victoria Barret, *Qualtrics: Tech's Hidden Gem in Utah*, FORBES (June 4, 2012), <http://www.forbes.com/sites/victoriabarret/2012/05/15/qualtrics-techs-hidden-gem-in-utah> (stating that Qualtrics users include "half the largest U.S. companies (Prudential, Geico and Microsoft among them), research arms of several federal agencies and 600 universities").

experiment. The program prohibited subjects from submitting multiple responses.

Based on self-reported demographic information, the subjects of our online experiment were more female (61%), younger (61% between 18–34 years old, 38% between 35–64 years old), somewhat less racially and ethnically diverse (82% white, 7% African-American, 5% Hispanic, 4% Asian-American, 3% other), and better educated (48% had undergraduate degree or higher) than the general population.¹⁸² When compared to a sample of jurors in long-term federal trials (20+ days) in a study by the Federal Judicial Center,¹⁸³ our experiment's subjects were of similar diversity with respect to race and sex, but were better educated and younger.¹⁸⁴ Notably, the subjects' demographics were more representative of the overall U.S. population than the law students used in our pilot study, as well as the undergraduate students used in previous experimental studies regarding jury behavior that have been published in peer-reviewed psychology journals.¹⁸⁵

V. DISCUSSION

This Part first discusses the results of our experiment. It then sets forth various implications from the experiment's findings. Finally, it discusses several potential directions for future research regarding standards of proof.

182. The general population of the United States is 50.8% female, 13.1% African-American, 16.7% Hispanic, 5.0% Asian-American, and 28.2% college graduates. U.S. CENSUS BUREAU, STATE AND COUNTY QUICKFACTS (2013), available at <http://quickfacts.census.gov/qfd/states/00000.html> (last updated Mar. 14, 2013).

183. JOE S. CECIL, E. ALLAN LIND & GORDON BERMANT, JURY SERVICE IN LENGTHY CIVIL TRIALS (1987), available at [http://www.fjc.gov/public/pdf.nsf/lookup/jurylnth.pdf/\\$file/jurylnth.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/jurylnth.pdf/$file/jurylnth.pdf). This study examined the demographic characteristics of jurors in twenty-nine complex civil jury trials that lasted at least twenty days. *Id.* at 11. These trials were conducted in six different federal districts — the Central, Northern, and Southern Districts of California; the Northern District of Illinois; the Southern District of New York; and the Eastern District of Pennsylvania. *Id.* at 12 tbl.1.

One apparent limitation of the Federal Judicial Center study is that it is now rather dated; the trials studied in it were conducted between 1976 and 1979, *id.* at 11, and thus do not reflect the increased racial and ethnic diversity of the United States and the increased education level of the jury-eligible population since the 1970s. However, more recent data on the demographics of federal jurors is not publicly available. We understand that while the federal district courts collect and maintain demographic information about their respective jury pools, they have a general policy against releasing this information.

184. The jurors in the long-term trials studied by the Federal Judicial Center were 57% female, 19% nonwhite, 22% college graduates, and had a mean age of 46.4 years. *Id.* at 19 tbl.4, 20.

185. *Cf.* Kagehiro & Stanton, *supra* note 87, at 162, 165 (using undergraduate psychology students as research subjects); Kerr et al., *supra* note 86, at 285 (same).

A. Results

In the experiment, the primary issue was whether the subjects' invalidity determinations varied based on the standard of proof and the presence or absence of an *i4i*-type jury instruction. First, we hypothesized that the clear and convincing standard would result in fewer findings of invalidity than the preponderance of the evidence standard. Second, we hypothesized that the clear and convincing standard would result in fewer findings of invalidity than the same standard given with an *i4i*-type instruction. Finally, we predicted that the clear and convincing standard with an *i4i*-type instruction would have fewer findings of invalidity than a plain preponderance standard.

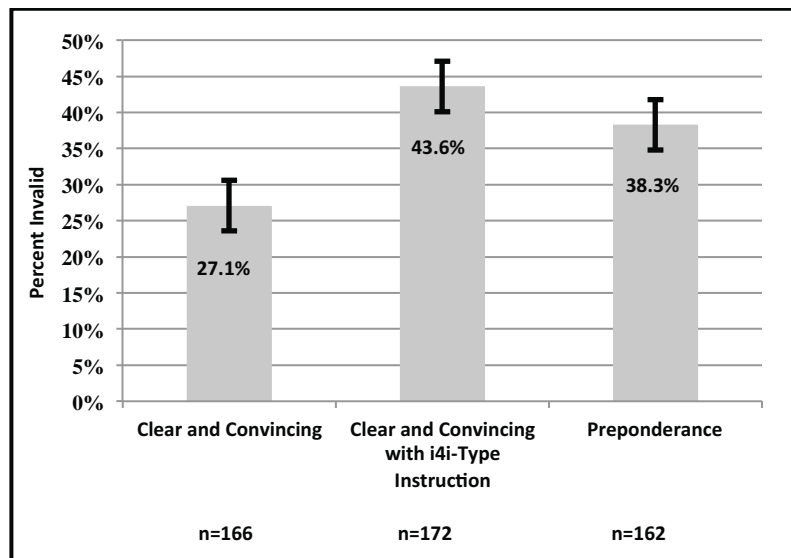


Figure 2: Invalidity Decisions by Standard of Proof

As shown in Figure 2 above, there were significant differences in mock jurors' decisions between the clear and convincing condition and the preponderance and *i4i*-type conditions.¹⁸⁶ As expected, subjects who received the clear and convincing standard found the patent invalid less often (27.1%) than those who received the preponderance standard (38.3%). Similarly, subjects who received the clear and convincing standard found the patent invalid less often than those who received the clear and convincing standard with an *i4i*-type instruction (43.6%). In a counterintuitive result, however, subjects who received

¹⁸⁶ In Figure 2, the lines above and below the grey bars represent the standard error for each condition.

the clear and convincing evidence standard with an *i4i*-type instruction found the patent invalid *more* often than those who received the preponderance standard (43.6% vs. 38.3%).

Using Pearson's chi-square (χ^2) statistic, we compared each condition against the other two to determine if the differences were statistically significant.¹⁸⁷ There was a highly statistically significant difference between the clear and convincing standard and the same standard given with an *i4i*-type instruction.¹⁸⁸ There was also a statistically significant difference in results between the clear and convincing standard and the preponderance standard.¹⁸⁹ Thus, the invalidity decisions from a plain clear and convincing standard are noticeably different from invalidity decisions using either of the other two standards. Both of these results are consistent with our first two hypotheses.

However, the difference between the clear and convincing standard with an *i4i*-type instruction and the preponderance standard was not found to be statistically significant.¹⁹⁰ In other words, given the sample size, our experiment could not discern a difference between these two instructions. In its *i4i* decision, the Court apparently assumed that the *i4i*-type instruction was distinguishable from a preponderance of the evidence standard.¹⁹¹ But our experiment could not detect any difference in results between these two conditions. Thus, while the Supreme Court explicitly rejected Microsoft's request to drop the standard to a preponderance of the evidence standard, the ruling by the Court may have effectively provided the same relief.

We then used a series of fixed-effects, multiple logistic regression models to test if this result held after controlling for other independent variables that might potentially impact subjects' decisions, including demographic characteristics (age, sex, and race), educational background, science degree, previous patent-related experience, and experience with golf, which is the subject matter of the patent in the

187. Statistical significance is the probability that an observed relationship is not due to chance. See *Getting Started with Statistics Concepts*, STATSOFT, [http://www.statsoft.com/textbook/elementary-statistics-concepts/_#What_is_\"statistical_significance\"_\(p-level\)](http://www.statsoft.com/textbook/elementary-statistics-concepts/_#What_is_\) (last visited May 9, 2013). A *p*-value of less than 0.05 is usually considered statistically significant. LAWLESS, ROBBENOLT & ULEN, *supra* note 145, at 233–34 (“When a result has less than a 5 percent chance of having been observed but is observed anyway, it is said to be *statistically significant*.”). A 5% probability is equal to a *p*-value of 0.05 or less. Results with a *p*-value of less than 0.01 are considered highly statistically significant. See *id.* at 234 n.4 (explaining that a 1% chance “represents a ‘higher’ level of significance because it indicates a less probable outcome and hence a more rigorous statistical test”). All data analysis was conducted using Stata/IC.

188. *p* = 0.002.

189. *p* = 0.031.

190. *p* = 0.322.

191. See *Microsoft Corp. v. i4i Ltd. P'ship*, 131 S. Ct. 2238, 2251 (2011) (holding that a clear and convincing standard exists, even when the party challenging invalidity offers new evidence not considered by the USPTO).

experiment.¹⁹² The short answer is yes: the differences between the clear and convincing condition on the one hand, and both the preponderance standard and the clear and convincing standard with an *i4i*-type instruction were significant in the expected direction. More specifically, even after holding all these other variables constant, the preponderance standard correlated with an increase in the odds ratio.¹⁹³ The best estimate is that there is a 90% increase in the odds ratio.¹⁹⁴ The clear and convincing standard with an *i4i*-type instruction correlated with an increase in the odds ratio. The best estimate is that the odds ratio increased by 128% for finding the patent invalid,¹⁹⁵ compared to the plain clear and convincing standard.¹⁹⁶ Furthermore, even with all of these statistical controls, we were unable to discern a statistically significant difference when comparing the preponderance standard to the clear and convincing standard with an *i4i*-type instruction. In other words, the difference in the odds ratio for invalidity between these conditions falls between -8% and 48%, which is within the standard error.¹⁹⁷ The full results for these regressions are reported in Appendix B.

In a separate question, we also asked subjects to separately indicate the percent likelihood that the patent was obvious (from 0% likely to be obvious to 100% likely to be obvious, in 10% increments). This variable served in part as a “check” on subjects’ answers regarding the patent’s invalidity. For example, if a subject found the patent was 20% likely to be obvious, but separately answered that the patent

192. See *infra* Appendix B for a more detailed description of the coding for these variables.

193. The odds ratio is defined as “the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.” Magdalena Szumilas, *Explaining Odds Ratios*, 19 J. CAN. ACAD. CHILD ADOLESCENT PSYCHIATRY 227, 227 (2010).

194. It is 95% likely that the increase in the odds ratio is in the interval between 0.43 (43% increase in the odds ratio) and 1.38 (138% increase in the odds ratio). Any range toward the middle of this interval is more likely to cover the true increase than is any range toward the interval’s extremes. In other words, a range of (0.89, 0.91) is more likely to cover the true increase than an interval of (0.43, 0.45) or an interval of (1.36, 1.38).

195. It is 95% likely that the increase in the odds ratio is in the interval between 0.72 (72% increase in the odds ratio) and 1.83 (183% increase in the odds ratio).

196. As reported in Appendix B, we used three separate specifications for the regression, each with successively more control variables. All of the models supported the same overall findings. Above we report the results from model 3, the most complete of the models and the one with the best goodness of fit. Furthermore, if one were to correct using the most conservative multiple testing penalty (which requires a *p*-value of less than 0.167), the findings from models 2 and 3 remain statistically significant. However, the difference in the base regression model 1 is not statistically significant if one were to make a multiple testing penalty. For a discussion of multiple testing errors, see generally John D. Storey, *The Positive False Discovery Rate: A Bayesian Interpretation and the *q*-Value*, 31 ANNALS OF STAT. 2013 (2003).

197. The only other statistically significant variable was female subjects, which had a negative relationship with invalidity (i.e., men were more likely to find the patent invalid than women).

was in fact invalid (for any condition), then the subject likely misunderstood the burden of proof and/or made an error in one of the responses. The reason is that under all three standards of proof we tested, the patent should only be declared invalid if the percentage obviousness is greater than 50%. We found that only 5% of responses (25 of 500) had conflicting answers regarding the percentage likelihood of obviousness and their ultimate decision on validity,¹⁹⁸ suggesting that the vast majority of subjects understood and answered these two questions consistently.¹⁹⁹

Next, using a logistic regression analysis, we tested whether there was a relationship between the percentage likelihood that the patent was obvious and the subject's ultimate answer regarding the patent's invalidity (for all three standards). Not surprisingly, the relationship between these two variables was found to be highly statistically significant with a positive coefficient that was large in magnitude.²⁰⁰ We then analyzed the subjects' level of confidence in their invalidity decisions, which were recorded on a 1–7 scale (with 1 representing “not at all confident” and 7 representing “very confident”). More specifically, we investigated whether the subjects' confidence in their decision regarding invalidity varied based on the standard of proof received. We thought it possible that subjects who believed the patent was invalid might be more confident in their decision if confronted with a lower standard of proof (i.e., preponderance instead of clear and convincing evidence), while conversely, subjects who believed the patent was valid might be more confident in their decision if faced with a higher standard of proof for overcoming the presumption of validity.

This result, however, was more ambiguous. A simple regression analysis between confidence and condition based on subjects' decision regarding invalidity revealed no statistically significant relationship between these variables for either subjects who found the patent valid²⁰¹ or subjects who found the patent invalid.²⁰² The mean confi-

198. We classified responses as “conflicting” in three situations: (1) when the percent obvious response was 100%, but the respondent found the patent valid (for any condition); (2) when the percent obvious response was 40% or less, but the respondent found the patent invalid (for any condition); and (3) for the preponderance condition only, when the percent obvious response was 60% or greater, but the respondent found the patent valid.

199. For graphical illustrations of the relationship between subjects' responses for percentage likelihood of obviousness and invalidity decisions for each condition, see Matthew Sag, *My CELS Comments on “Standards of Proof in Civil Litigation: An Experiment from Patent Law,”* COPYRIGHT LAW, FAIR USE & TECH. (Nov. 10, 2012), <http://matthewsag.com/my-cels-comments-on-standards-of-proof-in-civil-litigation-an-experiment-from-patent-law>.

200. $p < 0.001$; $\beta = 0.8206972$. The magnitude of the Beta coefficient (β) represents the relative contribution of the independent variable in the prediction of the dependent variable; it is represented as a value between -1 (negative correlation) and 1 (positive correlation). See, e.g., *Statistics Glossary*, STATSOFT, [http://www.statsoft.com/textbook/statistics-glossary/b/?button=0#Beta Coefficients](http://www.statsoft.com/textbook/statistics-glossary/b/?button=0#Beta%20Coefficients) (last visited May 9, 2013).

201. $p = 0.1873$.

dence levels for each condition, sorted by ultimate decision on invalidity, are reported in Table 1 below.

Condition	Patent Found Valid	Patent Found Invalid
Clear and Convincing	5.28	5.51
Clear and Convincing w/ <i>i4i</i> -type instruction	5.44	5.37
Preponderance	5.46	5.36

Table 1: Mean Confidence (1–7 Scale),
by Condition and by Invalidity Decision

However, subjects' mean confidence levels were found to vary depending on their assessment of the likelihood that the patent would be found obvious. For all three conditions, this resulted in a U-shaped distribution, as subjects who thought the patent was either highly likely or highly unlikely to be obvious were more confident in their responses than subjects who concluded the obviousness decision was a closer call. The latter group of subjects — i.e., those who recognized the obviousness decision reasonably might go either way — is arguably more likely to be influenced by a change in the standard of proof. The mean confidence level by percentage of likelihood obviousness for all three conditions is depicted in Figure 3 below.²⁰³

202. $p = 0.4385$.

203. At a more granular level, this U-shaped distribution exists for each of the three conditions studied: clear and convincing evidence, clear and convincing evidence with an *i4i*-type instruction, and preponderance of the evidence.

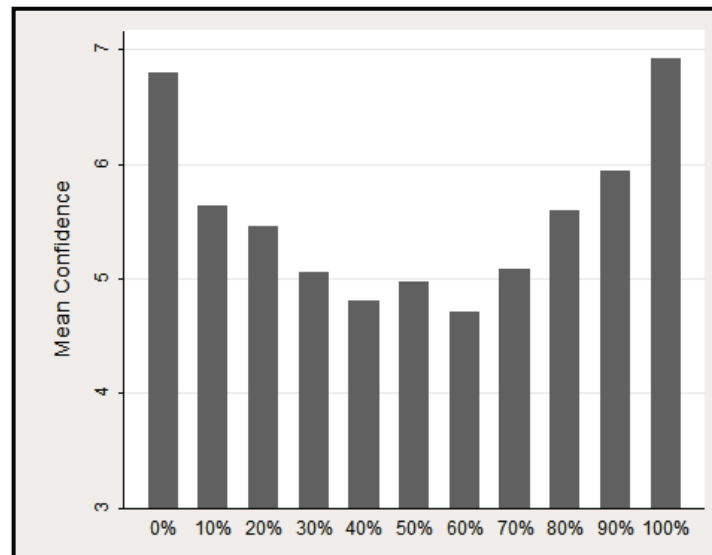


Figure 3: Mean Confidence (1–7 Scale),
by Percentage Likelihood of Obviousness (All Conditions)

When we presented our preliminary results at a recent academic conference, several colleagues suggested that our results might have been driven by the precise wording used for the *i4i*-type instruction. To consider this objection, we conducted a second iteration of the experiment in March 2012 using members of the general public eligible for U.S. jury duty (N=603).²⁰⁴ The second iteration included the identical fact pattern and arguments section. However, the three randomly assigned jury instructions were altered. We included verbatim the clear and convincing standard with the original *i4i*-type instruction designated above for one of the three randomly assigned jury instructions. To evaluate how much the precise wording of the *i4i*-type instruction mattered, the other two instructions were linguistic variations on the *i4i*-type instruction. In particular, the first variation removed the phrase “as in this case” from the final sentence of the original instruction,²⁰⁵ while the second variation²⁰⁶ replaced the final

204. For this iteration, we excluded subjects who had participated in the previous version of the experiment. This was accomplished by means of self-verification by subjects, as well as double-checking subjects’ user IDs in Mechanical Turk to determine if they had completed the first experiment. We compensated subjects \$1 for successfully completing the second iteration of the experiment.

205. This variation stated:

Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases. The

sentence of original instruction with the following two sentences: “In this case, the prior art on which the claim of obviousness is based was not considered by the Examiner. This may make it easier to satisfy the burden of proving obviousness.” The demographics of the subjects participating in the second iteration of the experiment were very similar to the original.²⁰⁷

The second iteration of the experiment confirmed our original findings. We found no statistically significant differences among the three variations of the clear and convincing standard with an *i4i*-type jury instruction.²⁰⁸ We contend that these results (or more accurately the lack of a statistically significant result) blunt the objection that the specific wording of the instruction alone drove our results.

burden of proving obviousness is more easily satisfied when the prior art on which the claim of obviousness is based was not considered by the Examiner.

206. This variation stated:

Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases. In this case, the prior art on which the claim of obviousness is based was not considered by the Examiner. This may make it easier to satisfy the burden of proving obviousness.

207. Based on self-reported demographic information, subjects for the second online experiment were more likely to be female (52%), younger (65% between 18–34 years old, 34% between 35–64 years old), were somewhat more racially and ethnically diverse than the first online experiment (80% white, 6% African-American, 4% Hispanic, 6% Asian-American, 3% other), and had roughly the same level of education as the first online experiment (48% had an undergraduate degree or higher).

208. Using Pearson’s chi-square (χ^2) statistic, $p = 0.647$ for the clear and convincing standard with the original *i4i*-type instruction compared to the first tested variation, and $p = 0.195$ for the clear and convincing standard with the original *i4i*-type instruction compared to the second tested variation. In this iteration, subjects who received the clear and convincing standard with the original *i4i*-type instruction found the patent invalid 32.8% of the time; subjects who received the first tested variation found the patent invalid 35.0% of the time; and subjects who received the second tested variation found the patent invalid 39.0% of the time.

We do not believe it is methodologically sound to directly compare results between the two iterations of the study for several reasons. They were conducted slightly over four months apart. By definition, they were taken upon different populations of Mechanical Turk users. Because we understood a direct comparison to be problematic, we re-ran the clear and convincing standard with the original *i4i*-type instruction against the two different variations so that an apples-to-apples comparison could be made. However, should such a comparison be made, it would not alter the results in a meaningful way. Specifically, if the data for the original clear and convincing standard with the *i4i*-type instruction are combined from both experiments, subjects found the patent invalid 37.8% of the time (140 of 370), which is statistically indistinguishable from the subjects’ rate of invalidity findings under the preponderance standard in the first experiment (38.3%, 62 of 162).

B. Implications

Our experiment's results suggest that standards of proof, and the jury instructions used to explain them, matter in civil litigation, although sometimes in unpredictable ways.

In patent challenges, our results suggest that if the Court in *i4i* had switched from a clear and convincing standard of proof to a preponderance standard for invalidity challenges, as Microsoft advocated, it may have resulted, *ceteris paribus*, in more patents being found invalid by juries. A switch to a preponderance standard also might have significantly affected patent holders' behavior, including bringing to trial fewer lawsuits involving patents of questionable validity.²⁰⁹ Previous empirical studies have found that a sizeable fraction — in some studies over half — of patent cases that reach a merits decision regarding validity ultimately find one or more claims to be invalid.²¹⁰ If the Court had adopted a preponderance standard in *i4i*, patentees may have been more reluctant to incur the time and expense of litigation in the face of higher odds that the patent(s)-in-suit would be found invalid if litigated to completion.²¹¹

Another notable implication is that the inclusion of an *i4i*-type instruction regarding evidence not previously considered by the USPTO may result in more invalidity findings. As noted above, we found a statistically significant difference between a plain clear and convincing standard and the same standard when given with an instruction of the type approved by the Court in *i4i*. Indeed, in our experiment, subjects found patents invalid at rates indistinguishable from the preponderance standard when such an *i4i*-type instruction was given. If this result extrapolates to patent litigation, the importance of *i4i*-type instructions may have been underappreciated by both courts and parties, as such an instruction may effectively reduce the standard of proof, even if it formally remains clear and convincing evidence. One possible explanation is that this may be partially due to a “framing effect,” where the additional sentence in the *i4i*-type instruction draws the jurors' attention to the perceived lower standard for invalidity (i.e.,

209. See Alsop, *supra* note 140, at 1650 (arguing that a shift to a preponderance standard would “deter at least some infringement actions based on weak patents”).

210. See, e.g., John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q.J. 185, 196–205 (1998) (finding that in 46% of final patent validity decisions reported in the U.S. Patent Quarterly between 1989 and 1996, the patent was determined to be invalid); Allison, Lemley & Walker, *supra* note 147, at 706 (finding that the most litigated patents since 2000 were invalidated about 70% of the time); Risch, *supra* note 161, at 483 (finding that in a study of patents asserted by nonpracticing entities that 65% of cases with a decision on the merits “invalidat[ed] at least part of a patent”).

211. However, the rise of contingency fee agreements in patent cases may partially counterbalance this by reducing the risks associated with the large attorneys' fees that are often incurred in patent litigation. See David L. Schwartz, *The Rise of Contingent Fee Representation in Patent Litigation*, 64 ALA. L. REV. 335 (2012).

“easier to satisfy”).²¹² As a result, patent infringement defendants should push for the inclusion of such an instruction at trial, as it may materially increase their odds of prevailing on this issue.

Relatedly, our findings give patentees an additional incentive to submit all potentially relevant prior art to the USPTO, during initial prosecution, reexamination, or supplemental examination,²¹³ to avoid the impact of an *i4i*-type instruction. If the patent examiner considers the submitted prior art — for instance, by citing to an issued patent, a published patent application, or a printed publication²¹⁴ — then an *i4i*-type instruction would be inappropriate. As noted in an amicus brief filed by past USPTO Commissioners and Directors in the *i4i* case, this would “encourage applicants to flood the USPTO with prior art,” with the potential effect of “strain[ing] the resources of the USPTO.”²¹⁵

In contrast, in situations where it is clear the prior art was in fact considered by the examiner during prosecution, parties challenging a patent’s validity may be better served by attempting to invalidate the patent through administrative proceedings at the USPTO, which have a lower burden of proof compared to litigation. Our experiment found a large and statistically significant difference between the clear and convincing standard and the preponderance standard, with the latter generating an approximately 90% increase in invalidity findings.²¹⁶ As a result, parties seeking to invalidate a patent in this situation likely should elect to take advantage of the preponderance standard applied in reexamination proceedings²¹⁷ and the new post-grant review proceeding provided for in the recently passed America Invents Act.²¹⁸

212. For an explanation of framing effects, see Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, 39 AM. PSYCHOLOGIST 341 (1984), and Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCI. 453 (1981).

213. Section 12 of the recently passed Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112-29, § 257, 125 Stat. 284, 325 (2011), provides that the patentee may file a “Request for Supplemental Examination” asking the USPTO “to consider, reconsider, or correct information believed to be relevant to the patent.” See 35 U.S.C. § 257(a) (2006 & Supp. V 2012).

214. 37 C.F.R. § 1.104(d) (2013) (explaining the citation of references).

215. *Amici Curiae* Brief of Former USPTO Commissioners and Directors in Support of Respondents at 29–30, *Microsoft Corp. v. i4i Ltd. P’ship*, 131 S. Ct. 2238 (2011) (No. 10-290).

216. See *supra* note 194 and accompanying text.

217. See *In re Swanson*, 540 F.3d 1368, 1377 (Fed. Cir. 2008) (explaining that “[i]n [US]PTO examinations and reexaminations, the standard of proof — a preponderance of evidence — is substantially lower than in a civil case” and that “there is no presumption of validity”). Similarly, for the new inter partes review procedure in the AIA, a preponderance standard will apply. See AIA, § 6(a) (codified at 35 U.S.C. § 316(e) (2006 & Supp. V 2012)) (“In an inter partes review instituted under this chapter, the petitioner shall have the burden of proving a proposition of unpatentability by a preponderance of the evidence.”).

218. See AIA, Pub. L. No. 112-29, § 6(d) (codified at 35 U.S.C. § 326(e) (2006 & Supp. V 2012)) (“In a post-grant review instituted under this chapter, the petitioner shall have the burden of proving a proposition of unpatentability by a preponderance of the evidence.”).

Furthermore, because most patent cases do not reach trial, the *i4i*-type instruction may be most strongly felt in summary judgment rulings. In theory, the standard of proof is included within the summary judgment standard.²¹⁹ We suspect that judges, using their intuition, may view the *i4i*-type instruction much like the Supreme Court apparently did — as creating a standard of proof somewhere between a preponderance and pure clear and convincing evidence standard. Our data suggests, however, that jurors do not view it in the same way. Instead, jurors treat the *i4i*-type standard as indistinguishable from the preponderance standard. Thus, there may be a disconnect between judges' and jurors' views on the relevant standard of proof. This disconnect may result in judges denying patent challengers' motion for summary judgment of invalidity in cases where reasonable jurors would find the patent claims invalid. If these cases proceeded to trial, the jurors would likely still find the patent invalid, and the effects of the disconnect would be small — primarily the additional cost of litigating the invalidity defense. However, many cases settle before trial if summary judgment is denied. These cases likely would settle for higher amounts based upon the disconnect between judges' and jurors' views on the relevant standard of proof.

In addition, our results may have broader implications for standards of proof in civil litigation more generally. As previously discussed, some of the previous literature found that standards of proof matter in jurors' decision making, although there were no representative studies that directly compared the clear and convincing standard with the preponderance standard.²²⁰ Our experiment suggests that jurors are sensitive to these two standards of proof and may reach different decisions accordingly.²²¹ This sensitivity to the standard of proof also tends to undermine previous criticisms that jurors in patent cases are unsophisticated and too easily swayed by tangential issues,²²² or overly impressed with “the fact that the patent was reviewed by an ‘expert agency’ with technically trained examiners.”²²³

219. See *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 252 (1986) (holding that “a ruling on a motion for summary judgment . . . necessarily implicates the substantive evidentiary standard of proof that would apply at the trial on the merits”).

220. See *supra* Part II.C.

221. Our results may also suggest that jurors have difficulty distinguishing between increasingly fine variations of the standard of proof, as evidenced by the fact that the “intermediate” clear and convincing with an *i4i*-type instruction resulted in invalidity findings comparable to the preponderance standard. For example, Kevin Clermont has argued that due to cognitive limitations, jurors measure uncertainty by referring to a “coarsely graded scale of probabilities” based on their own experiences, and thus legislators and other policymakers should avoid adopting “unrealistically finer degrees” of proof. Kevin M. Clermont, *Procedure's Magical Number Three: Psychological Bases for Standards of Decision*, 72 CORNELL L. REV. 1115, 1144–47 (1987).

222. See Gregory D. Leibold, *In Juries We Do Not Trust: Appellate Review of Patent-Infringement Litigation*, 67 U. COLO. L. REV. 623, 624 (1996).

223. See Moore, *Black Box*, *supra* note 147, at 373.

Furthermore, our experiment's results regarding the *i4i*-type instruction suggest that instructions that tell the jury to consider or give weight to particular facts may have a substantial impact on jurors' decisions. For example, some pattern jury instructions inform jurors that they may consider whether a witness (including an expert witness) has a bias, motive, or interest, including a financial interest, which might lead it to favor one of the parties, in determining whether to accept and what weight to give to that witness' testimony.²²⁴ Another example is a jury instruction regarding spoliation, which is the destruction of, or failure to preserve, evidence that is relevant to actual or threatened litigation.²²⁵ As a remedy for spoliation, numerous cases have authorized an instruction informing the jury that it may infer that the destroyed evidence would have been unfavorable or harmful to the destroying party.²²⁶ Our results suggest that individual jurors may be particularly sensitive to these types of instructions.

C. Directions for Future Research

We believe that our study has presented interesting results — that both the standard of proof and the existence of an *i4i*-type instruction apparently matter to jurors in patent challenges — and that our results are worthy of further empirical investigation. Our research could be extended in additional directions that we outline below.²²⁷

224. See, e.g., COMM. ON PATTERN JURY INSTRUCTIONS, FIFTH CIRCUIT DISTRICT JUDGES ASS'N, PATTERN JURY INSTRUCTIONS: FIFTH CIRCUIT, CIVIL CASES § 2.19 (2009) (permitting the following instruction for expert witnesses: "In deciding whether to accept or rely upon the opinion of an expert witness, you may consider any bias of the witness, including any bias you may infer from evidence that the expert witness has been or will be paid for reviewing the case and testifying . . .").

225. See, e.g., A. Benjamin Spencer, *The Preservation Obligation: Regulating and Sanctioning Pre-Litigation Spoliation in Federal Court*, 79 *FORDHAM L. REV.* 2005, 2018 (2011).

226. See, e.g., *Nucor Corp. v. Bell*, 251 F.R.D. 191, 204 (D.S.C. 2008) (authorizing the following spoliation instruction: "If you find that defendants engaged in the spoliation of evidence, you are permitted, but not required, to infer that the altered or destroyed evidence would have been unfavorable to defendants. Any inference you decide to draw should be based on all the facts and circumstances in this case.").

227. Extensions to our research can aid in overcoming certain known objections to our experiment. There is an extensive literature about the limitations on mock jury research. See generally Robert M. Bray & Norbert L. Kerr, *Methodological Considerations in the Study of the Psychology of the Courtroom*, in *THE PSYCHOLOGY OF THE COURTROOM* 287 (Robert M. Bray & Norbert L. Kerr eds., 1982); James H. Davis, Robert M. Bray & Robert W. Holt, *The Empirical Study of Decision Processes in Juries: A Critical Review*, in *LAW, JUSTICE AND THE INDIVIDUAL IN SOCIETY: PSYCHOLOGICAL AND LEGAL ISSUES* 326 (June Louin Tapp & Felice J. Levine eds., 1977); Ronald C. Dillehay & Michael T. Nietzel, *Constructing a Science of Jury Behavior*, in 1 *REVIEW OF PERSONALITY AND SOCIAL PSYCHOLOGY* 246 (Ladd Wheeler ed., 1980); Vladimir J. Konecni & Ebbe B. Ebbesen, *Methodological Issues in Research on Legal Decision-Making, with Special Reference to Experimental Simulations*, in *PSYCHOLOGY AND LAW: INTERNATIONAL PERSPECTIVES* 413 (Friedrich Losel, Doris Bender & Thomas Bliesener eds., 1992); Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 *LAW & HUM. BEHAV.* 75 (1999);

First, other researchers could attempt to replicate the results of our experiment using a different fact pattern. No individual study is dispositive of a research question, and additional experiments could help clarify if the results in our experiment reflect the variable being manipulated (here, the standard of proof), random chance, or some other unaccounted for factor.²²⁸ For instance, another experiment could use a fact pattern that is less likely to result in a 50-50 split in invalidity findings to determine if the standard of proof matters in such cases. And if replicated, one could also attempt to tease out how subjects understood the *i4i*-type instruction through, for example, an open-ended question about the standard of proof.

One concern of all mock jury experiments — ours included — is ecological validity. Ecological validity refers to the degree our experiment approximates the look, feel, and procedures of real trials, the situation we are studying.²²⁹ External validity is closely related to ecological validity, and refers to the degree the experimental result can be generalized to real trials.²³⁰ While the external validity of mock jury research is widely accepted,²³¹ there are obvious issues with ecological validity in such research, including ours.²³² For instance, in con-

Brian H. Bornstein & Sean G. McCabe, *Jurors of the Absurd? The Role of Consequentiality in Jury Simulation Research*, 32 FLA. ST. U. L. REV. 443 (2005); Robert M. Bray & Norbert L. Kerr, *Use of the Simulation Method in the Study of Jury Behavior: Some Methodological Considerations*, 3 LAW & HUM. BEHAV. 107 (1979); Shari Seidman Diamond, *Illuminations and Shadows from Jury Simulations*, 21 LAW & HUM. BEHAV. 561 (1997); Kathleen Carrese Gerbasi, Miron Zuckerman & Harry T. Reis, *Justice Needs a New Blindfold: A Review of Mock Jury Research*, 84 PSYCHOL. BULL. 323 (1977); Robert J. MacCoun, *Experimental Research on Jury Decision-Making*, 244 SCI. 1046 (1989), reprinted in 30 JURIMETRICS 223 (1990); Michael J. Saks, *What Do Jury Experiments Tell Us About How Juries (Should) Make Decisions?*, 6 S. CAL. INTERDISC. L.J. 1 (1997); Wayne Weiten & Shari Seidman Diamond, *A Critical Review of the Jury Simulation Paradigm: The Case of Defendant Characteristics*, 3 LAW & HUM. BEHAV. 71 (1979).

228. See LAWLESS, ROBBENNOLT & ULEN, *supra* note 145, at 47 (“[N]o individual study will be dispositive of a given research question. It is important that research results be replicated.”); David S. Goldman, *Legal Construct Validation: Expanding Empirical Legal Scholarship to Unobservable Concepts*, 36 CAP. U. L. REV. 79, 121–22 (2007) (“[I]t cannot be determined from the single experiment alone if the results are a reflection of the variables being manipulated, random happenstance, or a third unaccounted for factor.”).

229. See David L. Breau & Brian Brook, *“Mock” Mock Juries: A Field Experiment on the Ecological Validity of Jury Simulations*, 31 LAW & PSYCHOL. REV. 77, 78 (2007) (“The question of ecological validity asks whether a simulation is conducted under conditions that are similar to those in the real world such that the results from the simulation are generalizable to the world . . .”).

230. See LAWLESS, ROBBENNOLT & ULEN, *supra* note 145, at 39 (“The *ecological validity* . . . of a study . . . is sometimes used as a proxy for external validity. However, a study can have high external validity and be useful in legal or policy contexts even when it is not perfectly reflective of the conditions of that legal or policy setting.”).

231. See *supra* note 227.

232. *But see* Norbert L. Kerr & Robert J. MacCoun, *Is the Leniency Asymmetry Really Dead? Misinterpreting Asymmetry Effects in Criminal Jury Deliberation*, 15 GROUP PROCESSES & INTERGROUP REL. 585, 598 (2012) (arguing that mock jury experiments are not inferior to actual field experiments on jury behavior because “the latter are scarce and

trast to a full-blown trial, our experiment utilized a short, written description to convey the relevant facts and the parties' arguments. This was necessary to entice a large number of subjects to participate in the study. But real patent trials are obviously more complicated. The jurors hear live testimony from witnesses and review exhibits over an extended period of time, frequently one or two weeks. Opening and closing arguments at trials are delivered live, usually by experienced advocates. And most patent trials involve disputes over infringement, invalidity (often on several grounds), and damages, among other issues. Finally, jury instructions in patent cases can include many pages of instructions,²³³ which may dilute the importance of any single instruction. To address these concerns, future experiments could replace the written fact pattern with recorded video clips of closing arguments by patent litigation attorneys and with jury instructions by a judge in order to make the experiment seem more realistic. The addition of a judge reading the instructions may also encourage subjects to consider more carefully the jury instructions.

Another, perhaps more significant, external validity concern is the absence of deliberation. In real trials, jurors meet behind closed doors and discuss the charges, evidence, parties' arguments, and instructions. Their eventual decision is informed not only by the evidence and other aspects of the trial, but by discussions with other jurors.²³⁴ Deliberation may help diffuse jury confusion that may exist beforehand.²³⁵ As a result, our study is more reflective of the initial views of jurors before they begin the jury deliberation process, rather than a post-deliberation verdict. We cannot quantify what effect, if any, deliberation would have upon our results. However, negotiation theory suggests that higher opening offers in a negotiation should result in a

difficult to conduct, they rarely permit unconfounded experimental manipulations, and they usually lack the statistical power of mock jury experiments").

233. For example, the jury instructions given by the trial court judge in *i4i Ltd. Partnership v. Microsoft Corp.* were 35 pages long and addressed numerous issues, including infringement, willfulness, anticipation, obviousness, and damages, as well as the court's claim construction. Court's Charge, *i4i Ltd. P'ship v. Microsoft Corp.*, No. 6:07CV113 (E.D. Tex. May 19, 2009), available at <http://www.i4ilp.com/court/Jury%20Instructions.pdf>.

234. See, e.g., Robert J. MacCoun, *The Burden of Social Proof: Shared Thresholds and Social Influence*, 119 PSYCHOL. REV. 345, 353 (2012) (explaining "the *deliberation paradigm*, in which proponents of multiple viewpoints interact and attempt to influence each other").

235. See Shari Seidman Diamond, Beth Murphy & Mary R. Rose, *The "Kettleful of Law" in Real Jury Deliberations: Successes, Failures, and Next Steps*, 106 NW. U. L. REV. 1537, 1595 (2012) ("One possibility is that jurors faced with applying instructions during deliberations are able to assist one another in ways not captured . . . in studies of individual respondents."). But see Phoebe C. Ellsworth, *Are Twelve Heads Better Than One?*, 52 L. & CONTEMP. PROBS. 205, 218–23 (1989) (finding that 49% of juror responses to researchers' legal questions were unclear or wrong, and that deliberation did not cure individual jurors' misunderstanding).

higher final deal.²³⁶ Using the same logic, if the average juror is initially more inclined to find the patent invalid, then we suspect that the final verdict will likely be similarly more inclined. Future research could address this — albeit at the cost of significant time and expense — by adding group deliberations into the experiment.

For those who believe that ecological validity issues limit drawing inferences about real patent trials, we offer a more modest, alternative implication of our study. Sophisticated lawyers and clients frequently conduct mock jury studies as patent cases approach trial. These experienced players pay tens of thousands of dollars for such research,²³⁷ and we believe this supports external validity of the studies. However, even if mock jury studies such as ours are not representative of real jury trials, litigants can use the results of these studies for many purposes, including formulating settlement positions. Thus, even if the studies themselves are not indicative of how actual juries would decide cases, the results of the flawed studies are incorporated into the parties' settlement positions. Consequently, even if our results are not viewed as simulating actual trial results, they reveal valuable information about the likely results of settlements based upon mock jury studies.

Finally, there is a possibility that the *i4i*-type instruction may have “demand effects” on subjects. A “demand effect” exists when subjects in an experiment change their behavior based on what they perceive to be the desired or anticipated outcome by the experiment's organizers.²³⁸ It is conceivable that the *i4i*-type instruction being tested could create a demand effect by suggesting to some subjects that they should find the patent invalid. However, we do not think a substantial demand effect is likely here. Based on discussions with people who participated in the preliminary study, it was not apparent to them that we were testing the standard of proof, so they could not alter their behavior to conform to any expected outcome. In addition, demand effects are more likely to occur in a within-subjects design — that is, when the same subject is exposed to multiple conditions in the exper-

236. See Donald G. Gifford, *A Context-Based Theory of Strategy Selection in Legal Negotiation*, 46 OHIO ST. L.J. 41, 49 (1985) (“A negotiator who utilizes the competitive strategy begins with a high initial demand. Empirical research repeatedly demonstrates a significant positive relationship between a negotiator's original demand and his payoff.” (footnotes omitted)).

237. See, e.g., Tom Duffy, *Online Mock Juries Coming of Age*, LAWYERS WEEKLY USA, Aug. 30, 2004 (noting that traditional mock jury studies can cost \$20,000 to \$30,000), <http://www.ejury.com/pr/lawyersweekly.html>.

238. See Rachel Croson, *Why and How to Experiment: Methodologies from Experimental Economics*, 2002 U. ILL. L. REV. 921, 940 (2002) (“A demand effect is when a subject acts in a particular way to please the experimenter.”); see also Daniel J. Zizzo, *Experimenter Demand Effects in Economic Experiments*, 13 J. EXPERIMENTAL ECON. 75 (2010).

iment — rather than the between-subjects design used here.²³⁹ Finally, demand effects may also occur in real patent trials, with jurors perceiving an *i4i*-type instruction as a suggestion from the presiding judge that the patent is invalid.

VI. CONCLUSION

Our experiment has shown that standards of proof may play a significant role in jurors' decision making in civil litigation. Specifically, our results suggest that jurors apply standards of proof in litigated patent challenges in the manner expected by theory — i.e., that a patent's presumption of validity is more easily overcome under a preponderance of the evidence standard as compared to a plain clear and convincing evidence standard. Our experiments' results also suggest that jury instructions that highlight particular facts or arguments — here, that evidence not previously considered by the USPTO can make it “easier to satisfy” the standard for proving a patent invalid — can substantially influence jurors' decisions. Indeed, by giving the type of jury instruction authorized in *i4i*, jurors may be effectively applying a standard of proof that is more similar to the preponderance standard rejected by the Court in that case, rather than the clear and convincing evidence standard.

239. See Daniel Kahneman & Shane Frederick, *Representativeness Revisited: Attribute Substitution in Intuitive Judgment*, in *HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT* 49, 70 (Thomas Gilovich, Dale Griffin & Daniel Kahneman eds., 2002) (“The message that the [within-subjects] design conveys to the participants is that the experimenter expects to find effects of every factor that is manipulated . . .”); Prentice, *supra* note 159, at 1680 (“Within-subjects tests, for example, often suffer from demand effects, where features of the experiment itself allow the subjects to surmise the goals of the experimenter, an occurrence that results in the skewing of the subjects' responses.”).

APPENDIX A

SUMMARY OF THE DISPUTE

This dispute is between Acme Golf, Inc. (“Acme”) and Bravo Sporting Equipment Co. (“Bravo”). Acme and Bravo are the two largest manufacturers of golf balls in the United States.

Historically, golf balls consisted of two parts, a solid core and a cover layer covered with dimples (indents). Some golf balls were designed to travel long distances when struck by a club. These balls had a relatively hard plastic cover layer. However, this hard cover was relatively inflexible and created an undesirable “feel” when struck with a golf club. In addition, the hard cover made it difficult for some golfers to control the ball’s direction and/or spin.

In contrast, other golf balls were designed to exhibit other desirable characteristics, such as control for shorter shots and the proper “feel” when struck. These golf balls also have two parts (a solid core and a cover layer), but use a softer plastic material for the cover layer. One well-known and widely-used material for soft-cover golf balls was a type of plastic called polyurethane. But soft-cover balls had the disadvantage of travelling less distance than their hard-cover counterparts. In addition, soft-cover balls were less durable and tended to damage more easily. Both hard-cover and soft-cover balls were well known in the field since at least the 1950s.

In 2005, a golf ball designer at Acme designed a three-piece golf ball consisting of the following parts: (1) a solid core, (2) a hard inner layer, and (3) a softer outer (cover) layer of polyurethane covered with dimples. This three-piece design resulted in a “dual personality” ball capable of traveling long distances due to the hard inner layer, but also had the desirable control and “feel” characteristics of soft-cover balls due to the polyurethane cover layer. Acme timely applied for a patent on this three-piece golf ball.

In the United States, patents are granted by the USPTO, which is an agency of the federal government. To obtain a patent, one must first file an application with the USPTO. The USPTO employs trained Patent Examiners who review patent applications. After an application is filed, an Examiner reviews it to determine whether or not the claimed invention is patentable. During this process, the Examiner searches for and reviews certain information about the state of technology in the relevant field(s) at the time the application was filed. This information is called “prior art.” In general, prior art includes all publicly-available information about the state of technology when the patent application was filed. The Examiner reviews this prior art to

determine whether the claimed invention is truly an advance over existing technology.

One requirement for obtaining a patent is that the invention would not have been obvious in light of the prior art. A claimed invention is obvious if an ordinary skilled person in the relevant field of technology who knew about all of the prior art would have been able to come up with the invention at that time.

In this case, the Examiner reviewed the prior art regarding both hard-cover and soft-cover golf balls and determined that Acme's patent application for its three-piece golf ball was not obvious and therefore should receive a patent (the "Acme patent"). However, during examination, the Examiner did not locate and consequently did not review a patent granted to an inventor named Charles in 2000. The Charles patent disclosed a three-piece golf ball with a solid core, a hard inner layer, and an outer (cover) layer consisting of a very hard resin covered with dimples. This hard resin surface had the advantage of making the golf ball extremely durable. The Charles patent does not mention polyurethane, nor does it suggest trying to use a softer material for the outer (cover) layer of the ball. The parties agree the Charles patent is prior art to the Acme patent.

Earlier this year, Bravo started selling "Flight X" golf balls that have a three-piece design with a solid core, a hard inner layer, and a softer outer (cover) layer of polyurethane. Acme believes that Bravo's "Flight X" balls infringe Acme's patent and has sued Bravo in court. In response, Bravo has asserted that Acme's patent was obvious in light of the prior art (existing technology) at the time of Acme's claimed invention. Under the patent law, if the invention in a patent was obvious, there is no liability for infringement.

BRAVO'S ARGUMENT REGARDING OBVIOUSNESS

Bravo argues Acme's patent is obvious because it merely combines preexisting items that were already known in the prior art. Specifically, Bravo claims that the Charles patent disclosed a three-piece golf ball with inner and outer layers of different hardness. It would have been obvious to an ordinary golf ball manufacturer, Bravo contends, to modify the Charles three-piece ball to have a soft outer (cover) layer of polyurethane, which has been widely used in traditional two-piece soft-cover balls. Because of this polyurethane cover, a golf ball maker would expect such a ball to have the desirable control and "feel" characteristics of soft-cover balls. In addition, Bravo argues that the jury should not defer to the Patent Examiner's conclusion that the Acme three-piece ball was patentable because the Examiner never considered the Charles patent, which is a key piece of prior art. The Examiner's failure to consider the Charles patent satisfies Bravo's

burden of proving that the claimed invention in Acme's patent is obvious because if the Examiner had known about the highly relevant Charles patent, then Acme would not have been granted a patent.

ACME'S ARGUMENT REGARDING OBVIOUSNESS

Acme argues that its patent is not obvious for several reasons. First, none of the prior art disclosed the combination of items that resulted in the patented invention. This combination is worthy of a patent, Acme contends, because it created a golf ball with the unique benefits of the control and "feel" of a two-piece soft-cover ball, combined with the long distance of a hard-cover ball. Nothing in the prior art suggested this combination would create a ball with these favorable characteristics. Second, the Charles patent does not make Acme's patented invention obvious because the golf balls in the Charles patent were designed to solve a very different problem — the lack of durability. Third, Acme claims that the Charles patent does not make Acme's patented invention obvious because there is nothing in the Charles patent that would suggest or hint to an ordinary golf ball manufacturer that trying a softer cover like polyurethane on a three-piece ball might be a good idea. Finally, Acme contends that the jury should defer to the decision of the trained Patent Examiner, who was in the best position to determine whether Acme's claimed invention was obvious. As a result, Bravo cannot satisfy its burden of establishing that Acme's patent was obvious.

INSTRUCTIONS[‡]

Below are the rules that you must follow in deciding whether Acme's patent was obvious.

The fact that the USPTO grants a patent on a claimed invention does not necessarily mean that it in fact deserves protection under the patent laws. A party can argue in court that it is not liable for infringement because the patented invention was obvious. Here, Bravo is arguing that the Patent Examiner made an error in determining that the Acme patent was not obvious.

Under the law, Acme's patent is presumed to be not obvious. As a result, Bravo has the obligation to persuade you, which is called the burden of proof, that the claimed invention in [the]^{‡‡} Acme patent is obvious. An invention is obvious if a person of ordinary skill in the relevant field who knew about all of the prior art that existed at the time of the claimed invention would have come up with the invention at that time.

[‡] Bracketed language in the instructions was varied by condition.

^{‡‡} This word was inadvertently omitted from the original survey.

ⁱ[To succeed on its claim, Bravo must prove obviousness by clear and convincing evidence. Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases.]

ⁱⁱ[To succeed on its claim, Bravo must prove obviousness by clear and convincing evidence. Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases. The burden of proving obviousness is more easily satisfied when, as in this case, the prior art on which the claim of obviousness is based was not considered by the Examiner.]

ⁱⁱⁱ[To succeed on its claim, Bravo must prove obviousness by a preponderance of the evidence. A preponderance of the evidence is evidence that shows it is more probable than not that the patent was obvious. This is a lower standard of proof than either clear or convincing evidence or the beyond a reasonable doubt standard used in criminal cases.]

^{iv}[To succeed on its claim, Bravo must prove obviousness by clear and convincing evidence. Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases. The burden of proving obviousness is more easily satisfied when the prior art on which the claim of obviousness is based was not considered by the Examiner.]

^v[To succeed on its claim, Bravo must prove obviousness by clear and convincing evidence. Clear and convincing evidence is evidence that shows it is highly probable that the patent was obvious. This is a

i. This language was only given to subjects who received the clear and convincing evidence condition in the first iteration of the experiment.

ii. This language was given to subjects who received the clear and convincing evidence with the *i4i-type* instruction in both the first and second iterations of the experiment. In the first iteration of this experiment, there was a blank line before the last sentence of this paragraph.

iii. This language was only given to subjects who received the preponderance of the evidence condition in the first iteration of the experiment.

iv. This language was only given to subjects who received the first variation of the clear and convincing condition with the *i4i-type* instruction condition in the second iteration of the experiment.

v. This language was only given to subjects who received the second variation of the clear and convincing condition with the *i4i-type* instruction condition in the second iteration of the experiment.

higher standard of proof than a preponderance of the evidence, which means more probable than not. However, clear and convincing evidence is lower than the beyond a reasonable doubt standard used in criminal cases. This may make it easier to satisfy the burden of proving obviousness.]

In deciding obviousness, you should put yourself in the position of a person with ordinary skill in the relevant field of technology at the time of the claimed invention. You must not use hindsight; in other words, you may not consider what is known now or what was learned from Acme's patent to determine whether it is obvious. In addition, you may not use Acme's patent as a roadmap for selecting and combining items of prior art to create the claimed invention.

QUESTIONS

Based on the information provided above, please answer the following questions.

1. Did Bravo prove by [clear and convincing evidence / a preponderance of the evidence][‡] that Acme's patent was obvious?

Yes
(Obvious)

No
(Not Obvious)

2. On a scale of 1 to 7, how confident are you in your answer to Question #1?

1	2	3	4	5	6	7
Not at all confident			Moderately confident			Extremely confident

3. On a scale of 0% to 100%, how likely do you think it is that Acme's patent was obvious?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Certainly not obvious					Equally likely to be obvious or not obvious					Certainly obvious

[‡] Bracketed language in the question was varied based on condition.

APPENDIX B

Variable	Model 1	Model 2	Model 3
Preponderance jury instruction	1.7007* (.4083)	1.8925** (.4708)	1.9019** (.4761)
<i>i4i</i> jury instruction	2.1568*** (.5073)	2.1871*** (.5256)	2.2767*** (.5526)
Female	.5777** (.1110)	.5602** (.1111)	.5821** (.1213)
Age indicator		X	X
Education indicator		X	X
Race indicator		X	X
Golf experience			1.2966 (.2962)
Patent experience			1.9371 (1.3700)
Science background			1.4280 (.3928)
Constant	-.1365 (.3417)	.0091 (1.2796)	-1.9483 (1.8899)
R²	.0285	.0433	.0496

Table 2: Logistic Regression Model for Clear and Convincing Evidence with *i4i*-type Instruction and Preponderance of the Evidence[‡]

[‡] The tables in Appendix B report three separate logistic regression models that predict *obvious*, a variable that is positive when a respondent indicated that the patent is invalid as obvious. *Preponderance jury instruction* refers to the preponderance of the evidence instruction; *i4i jury instruction* refers to the clear and convincing evidence with *i4i*-type instruction; *female* is a dummy variable for whether the respondent was female. In addition, models 2 and 3 include dummy variables for *age* (including separate dummy variables for ages between 18–24, 25–34, 35–44, 45–54, 55–64, and 65 years or older), *education* (including separate dummy variables for high school degree or less, some college, 4-year college degree, and advanced degree), and *race* (separate dummy variables for White, Black, Hispanic, Asian-American/Pacific Islander, Native American, and Other). Model 3 includes dummy variables *golf* (which is positive if the respondent indicated they were somewhat knowledgeable or highly knowledgeable about golf), *patent* (personal experience with the patent system), and *science* (college degree or advanced degree in a scientific field). The values reported are odds ratios and (standard errors). Superscripts report the relevant level of statistical significance: variables which do not significantly predict the response variables (no superscript), while others are predictors at a marginal level of significance ([†]), $p \leq 0.1$; at the conventional definition of significance (*), $p \leq 0.05$; at higher level of significance (**), $p \leq 0.01$; and at an even higher level of significance (***), $p \leq 0.001$. The models were created using Stata.

Variable	Model 1	Model 2	Model 3
Clear and Convincing jury instruction	.58800* (.1412)	.5284** (.1315)	.5258** (.1316)
<i>i4i</i> jury instruction	1.2682 (.2856)	1.1557 (.2693)	1.1971 (.2811)
Female	.5777** (.1110)	.5602** (.1111)	.5821** (.1213)
Age indicator		X	X
Education indicator		X	X
Race indicator		X	X
Golf experience			1.2966 (.2962)
Patent experience			1.9371 (1.3700)
Science background			1.4280 (.3928)
Constant	.3946 (.3442)	.6470 (1.2724)	-1.3054 (1.8845)
R²	.0285	.0433	.0496

Table 3: Logistic Regression Model for Clear and Convincing Evidence and Clear and Convincing Evidence with *i4i*-Type Instruction