

An Empirical Study of the Multifactor Tests for Trademark Infringement

Barton Beebe[†]

The thirteen circuits' thirteen different multifactor tests for the likelihood of consumer confusion have long played a role of central importance in American trademark litigation, yet they have received little academic attention and no empirical analysis. Courts, commentators, and practitioners have all the while speculated about which factors, if any, drive the outcome of the tests, how the factors interact, and, most importantly, whether the circuits' different tests, given the same facts, would yield different outcomes. With a view to the settling of these questions and ultimately to the reform of the multifactor tests, this Article sets forth the results of an empirical study of all reported federal district court opinions for the five-year period from 2000 to 2004 in which a multifactor test for the likelihood of consumer confusion was used. In the process, it presents the multifactor test for the likelihood of consumer confusion as an ideal case study in legal multifactor decision making and develops a methodology and theoretical toolkit for the study of this form of legal analysis across the many areas of law that employ multifactor tests.

Working from an original data set of 331 opinions, this Article finds significant variation among the circuits in the application and outcome of their respective tests. Drawing upon recent social science learning on cognition and decision making, it further shows that judges employ “fast and frugal” heuristics to short-circuit the multifactor analysis. A few factors prove to be decisive; the rest are at best redundant and at worst irrelevant.

Copyright © 2006 California Law Review, Inc. California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

[†] Associate Professor of Law, Benjamin N. Cardozo School of Law, Yeshiva University. B.A., University of Chicago; Ph.D., Princeton University; J.D., Yale University. Thanks to Jake Barnes, Richard Bierschbach, Rochelle Dreyfuss, Marc Henri, Justin Hughes, Jacob Jacoby, Max Minzner, Michael Perino, Dan Simon, Amrit Singh, and Alex Stein for comments on previous drafts. This project greatly benefited from presentations to the faculties of Fordham University School of Law, Hofstra University School of Law, and St. John's University School of Law, and at the Intellectual Property Scholars Conference 2005. For excellent research assistance, I thank Carl Regelmann. I used Stata 9.2 to conduct the statistical analysis presented in this Article. The data set as well as documentation necessary to reproduce my results is available at <http://www.bartonbeebe.com>. Further comments welcome: barton@bartonbeebe.com.

Judges tend to “stampede” these remaining factors to conform to the test outcome, particularly when they find infringement.

This Article also makes a variety of factor-specific findings that contravene conventional wisdom in trademark law. For example, courts typically declare that no single factor outcome is dispositive. The data clearly contradict this assertion. A finding that the similarity of the marks factor does not favor a likelihood of confusion is, in practice, dispositive, and a finding that the proximity of the goods factor does not favor a likelihood of confusion is nearly dispositive. Relatedly, at least one circuit, the Second, has declared the defendant’s intent factor to be “irrelevant.” Meanwhile, the data show that a finding of bad faith intent is in fact dispositive, as much in the district courts of the Second Circuit as in those of any other. Furthermore, many believe survey evidence to be the best and most persuasive form of evidence of the likelihood of confusion. The data reveal, however, that surveys are rarely presented by parties or credited by courts. Additionally, this Article finds that the venerable doctrine of trademark strength—and in particular of inherent distinctiveness—has broken down. The data show a remarkable divergence between traditional circuit court (and casebook) doctrine on this and current district court practice.

Introduction

The overriding question in most federal trademark infringement litigation is a simple one: is the defendant’s trademark, because of its similarity to the plaintiff’s trademark, causing or likely to cause consumer confusion as to the true source of the defendant’s goods? In answering this question, each circuit requires that the district court conduct a multifactor analysis of the likelihood of consumer confusion according to the factors set out by that circuit. As the Seventh Circuit has recently explained, the multifactor test operates “as a heuristic device to assist in determining whether confusion exists.”¹

This heuristic device is the fulcrum of American trademark law, and yet, for all of its importance, the test is in a severe state of disrepair. Its current condition is Babelian. Each circuit has developed its own formulation of the test.² In the Eighth and Tenth Circuits, the test consists of six factors.³ Other circuits, such as the Seventh,⁴ typically use seven factors,⁵

1. Sullivan v. CBS Corp., 385 F.3d 772, 778 (7th Cir. 2004).

2. J. Thomas McCarthy, 3 McCarthy on Trademarks and Unfair Competition § 24:30 (4th ed. 2006) (“[E]ach of the thirteen federal circuit courts of appeal has developed its own version of the list and each appears to be jealous of its own formulation of factors.”).

3. See Sally Beauty Co. v. Beautyco, Inc., 304 F.3d 964, 972 (10th Cir. 2002); SquirtCo v. Seven-Up Co., 628 F.2d 1086, 1091 (8th Cir. 1980).

4. See Helene Curtis Indus., Inc., v. Church & Dwight Co., 560 F.2d 1325, 1330 (7th Cir. 1977).

5. The Fourth, Fifth, and Eleventh Circuits also typically use a seven-factor test. For the Fourth Circuit, see Pizzeria Uno Corp. v. Temple, 747 F.2d 1522, 1527 (4th Cir. 1984) (using a seven-factor

while still others, such as the Second⁶ and the Ninth,⁷ typically use eight.⁸ The Third Circuit uses ten⁹ and the Federal Circuit thirteen,¹⁰ of which the last is “any other established fact probative of effect of use.”¹¹ While there is overlap among some of the factors used, there is also great diversity—not just in which factors are employed, but in how they are employed. Some circuits claim to weigh heavily under certain factors what other circuits claim to ignore,¹² and nearly every factor or combination of factors has been called the “most important” by one court or another. To make matters worse, scattered among the circuits are factors that are clearly obsolete, redundant, or irrelevant, or, in the hands of an experienced judge or

test). *But see* *Perini Corp. v. Perini Constr., Inc.*, 915 F.2d 121, 127 (4th Cir. 1990) (using an eight-factor test); *Teaching Co. Ltd. P’shp. v. Unapix Entm’t, Inc.*, 87 F. Supp. 2d 567, 580 (E.D. Va. 2000) (citing *Pizzeria Uno* and *Perini* and using a nine-factor test); *Christian Sci. Bd. of Dirs. of the First Church of Christ v. Robinson*, No. 1:99-cv-148-T, 2000 U.S. Dist. LEXIS 21309, at *16-17 (W.D. N.C. 2000) (citing *Pizzeria Uno* and using an eleven-factor test). There is a degree of confusion in the Fourth Circuit as to what the factors are called. *See, e.g.*, *Sterling Acceptance Corp. v. Tommark, Inc.*, 227 F. Supp. 2d 454, 463 (D. Md. 2002) (referring to the “*Lone Star* test” and the “*Lone Star* Factors”); *Giant Brands, Inc. v. Giant Eagle, Inc.*, 228 F. Supp. 2d 646, 654 (D. Md. 2002) (referring to the “*Pizzeria Uno* factor[s]”); *Eurotech, Inc. v. Cosmos European Travels Aktiengesellschaft*, 213 F. Supp. 2d 612, 622 (E.D. Va. 2002) (referring to the “seven *Petro Stopping* factors”). For the Fifth Circuit, see *Roto-Rooter Corp. v. O’Neal*, 513 F.2d 44, 45 (5th Cir. 1975) (using a seven-factor test). *But see* *Sunbeam Prods., Inc. v. W. Bend Co.*, 123 F.3d 246, 257 (5th Cir. 1997) (using an eight-factor test), *overruled on other grounds* by *Traffix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23 (2001); *Urgent Gear Inc. v. Savoia*, No. 3:01-CV-2190-D, 2001 U.S. Dist. LEXIS 20459, at *10 (N.D. Tex. 2001) (citing *Sunbeam* and using an eight-factor test). For the Eleventh Circuit, see *Alliance Metals, Inc. v. Hinely Indus., Inc.*, 222 F.3d 895, 907 (11th Cir. 2000). *But see* *Coach House Rest., Inc. v. Coach & Six Rest. Inc.*, 934 F.2d 1551, 1561 (11th Cir. 1991) (using an eight-factor test).

6. *See* *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492, 495 (2d Cir. 1961). *But see* *Star Indus., Inc. v. Bacardi & Co. Ltd. Corp.*, No. 02 Civ. 4239(HB), 2003 WL 23109750, at *3 (S.D.N.Y. 2003) (referring to the “now famous seven-step analysis spelled out by Judge Friendly in *Polaroid*”).

7. *See* *AMF, Inc. v. Sleekcraft Boats*, 599 F.2d 341, 348-49 (9th Cir. 1979); *Cf. Dr. Seuss Enter., L.P. v. Penguin Books USA, Inc.*, 109 F.3d 1394, 1404 n.13 (9th Cir. 1997) (“The Ninth Circuit test for likelihood of confusion also has been described as a six-factor test in *J.B. Williams Co. v. Le Conte Cosmetics, Inc.*, 523 F.2d 187, 191 (9th Cir. 1975), *cert. denied*, 424 U.S. 913 (1976); *Century 21 Real Estate Corp. v. Sandlin*, 846 F.2d 1175, 1179 (9th Cir. 1988); and a five-factor test in *Rodeo Collection, Ltd. v. W. Seventh*, 812 F.2d 1215, 1217 (9th Cir. 1987). We use the eight-factor *Sleekcraft* test in this case simply to be over-inclusive.”).

8. The First, Sixth, and D.C. Circuits also typically use an eight-factor test. For the First Circuit, see *I.P. Lund Trading ApS v. Kohler Co.*, 163 F.3d 27, 43 (1st Cir. 1998); *Boston Athletic Ass’n v. Sullivan*, 867 F.2d 22, 29 (1st Cir. 1989). For the Sixth Circuit, see *Frisch’s Rests., Inc. v. Elby’s Big Boy*, 670 F.2d 642, 648 (6th Cir. 1982). For the D.C. Circuit, see *Basile, S.p.A. v. Basile*, 899 F.2d 35, 37 (D.C. Cir. 1990) (citing with approval *Polaroid Corp. v. Polaroid Elecs. Corp.*, 287 F.2d 492, 495 (2d Cir. 1961)); *see also* *Whitehead v. CBS/Viacom, Inc.*, 315 F. Supp. 2d 1, 35-36 (D.D.C. 2004) (using the *Polaroid* eight-factor test).

9. *See* *Interpace Corp. v. Lapp, Inc.*, 721 F.2d 460, 462-63 (3d Cir. 1983). As with the Fourth, the Third Circuit’s name for the multifactor test varies. *See, e.g.*, *Gideons Int’l, Inc. v. Gideon 300 Ministries, Inc.*, 94 F. Supp. 2d 566, 582 (E.D. Pa. 1999) (referring to the ten-factor test “known as the *Lapp* factors, or the *Scott Paper* factors”).

10. *See In re E. I. Du Pont de Nemours & Co.*, 476 F.2d 1357, 1361 (C.C.P.A. 1973).

11. *Id.*

12. *See infra* Part III.

litigator, notoriously pliable.¹³ This is a circuit split of considerable proportions, one which produces, as this Article will show, excessive inter-circuit variation in the application and outcome of the multifactor tests.

Despite their critical importance, the multifactor tests have received little academic analysis beyond that of the treatise writers¹⁴ and no empirical analysis.¹⁵ For their part, the circuits have not sought to harmonize their tests and the Supreme Court, despite having issued seven trademark opinions in the last twelve years,¹⁶ has so far declined to intervene. Courts, commentators, and practitioners have all the while speculated about which factors, if any, actually drive the outcome of the test, how the factors interact, and most importantly, whether the different tests, given the same facts, would yield different outcomes. No convincing answers to any of these questions have yet emerged.

We lack knowledge of the multifactor tests because these are empirical questions that can only be answered through the use of empirical methods. This Article sets forth the results of an empirical study of all reported district court opinions for the five-year period from 2000 to 2004 that made substantial use of a multifactor test for the likelihood of consumer confusion.¹⁷ In working from an original data set of 331 opinions, it pursues two related goals, the first specific to trademark law, the second broader and

13. See *infra* Part IV.B.

14. See *e.g.*, Jerome Gilson, Anne Gilson LaLonde, & Karin Green, Trademark Protection and Practice § 5.05 (2005); Richard L. Kirkpatrick, Likelihood of Confusion in Trademark Law (2006); McCarthy, *supra* note 2, at §§ 23:19-23:31, 24:22-24:50. For non-treatise treatments of the multifactor test, see David J. McKinley, *Proving Likelihood Of Confusion: Lanham Act vs. Restatement*, 12 J. Contemp. Legal Issues 239 (2001) (briefly reviewing the circuits' adoptions of multifactor tests for trademark infringement); Michael H. Bierman & Jeffrey D. Wexler, *Toward a Reformulation of the Test for Determining Trademark Infringement*, 80 Trademark Rep. 1 (1990); Robert G. Sugarman & Doreen G. Small, *Proving Likelihood of Confusion in Trademark Cases*, Trial, Mar. 1990, at 51 (discussing strategies of litigating the multifactor test for trademark infringement). For analysis of the actual confusion factor in particular, see Mark D. Robins, *Actual Confusion in Trademark Infringement Litigation: Restraining Subjectivity Through a Factor-Based Approach to Valuing Evidence*, 2 Nw J. Tech. & Intell. Prop. 2 (2004) (exhaustively reviewing courts' treatment of the actual confusion factor and advocating a "factor-based approach" to the analysis of the factor); Edwin S. Clark, *Finding a Likelihood of Confusion With Actual Confusion: A Critical Analysis of the Federal Courts' Approach*, 22 Golden Gate U.L. Rev. 393 (1992) (arguing that courts give the actual confusion factor more weight than other factors in the multifactor test and approving of this approach).

15. But see Graeme W. Austin, *Trademarks and the Burdened Imagination*, 69 Brook. L. Rev. 827, 866-72 (2004) (conducting an "impressionistic survey of trademark infringement cases over the last decade [from May 1993 through May 2003], which discloses that in about half the litigated cases, litigants offered no survey evidence to the court or, if they did, courts accorded it little weight").

16. See *KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc.*, 543 U.S. 111 (2004); *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23 (2003); *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418 (2003); *Traffix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23 (2001); *Wal-Mart Stores, Inc. v. Samara Bros.*, 529 U.S. 205 (2000); *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159 (1995); *Two Pesos, Inc. v. Taco Cabana, Inc.*, 505 U.S. 763 (1992).

17. For a description of the method by which the opinions were collected and coded, see Appendix A. As Appendix A explains, I defined "substantial use" liberally as any application of the multifactor test that did more than merely cite the test without analysis.

more theoretical in nature. First, this Article seeks to show how the tests actually work in practice—and to do so by using easily understood statistical methods. My hope is that the study’s findings will enhance predictability and proficiency in the application of the circuits’ current tests and serve as evidence in support of a national standard multifactor test.

Second, this Article draws upon recent social science learning on cognition and decision making to reflect more generally on the nature of legal multifactor decision making. This is an important inquiry not simply for trademark law, where so much of adjudication has taken the form of either statutorily-prescribed or judge-made multifactor tests,¹⁸ but for a wide and diverse assortment of other areas of law as well, including copyright,¹⁹ takings,²⁰ evidence,²¹ conflict of laws,²² and criminal law,²³ to name only a

18. See, e.g., 15 U.S.C. § 1125(c)(1)(A)-(H) (2001) (prescribing a nonexhaustive eight-factor test to determine whether a mark is “distinctive and famous”); 15 U.S.C. § 1125(d)(1)(B)(i) (2001) (prescribing a nonexhaustive nine-factor test to determine whether a defendant has a “bad faith intent to profit” from the use of a domain name “identical or confusingly similar” to plaintiff’s trademark); *Heartland Bank v. Heartland Home Fin., Inc.*, 335 F.3d 810, 819 (8th Cir. 2003) (setting forth a nonexhaustive seven-factor test to determine secondary meaning); *Echo Travel, Inc. v. Travel Assocs., Inc.*, 870 F.2d 1264, 1267 (7th Cir. 1989) (same); *E-Systems, Inc. v. Monitek, Inc.*, 720 F.2d 604, 607 (9th Cir. 1983) (setting forth six factors to determine whether laches should bar relief); *In re Morton-Norwich Prods., Inc.*, 671 F.2d 1332, 1340-41 (C.C.P.A. 1982) (setting forth four factors to consider in assessing the functionality of product features); Council Directive 89/104 of 21 December 1988, To Approximate the Laws of the Member States Relating to Trade Marks, 1988: O.J. (L40) 1, <http://oami.europa.eu/en/mark/aspects/direc/direc.htm> (setting forth four factors to be considered in determining the likelihood of consumer confusion). The *Trademark Dilution Revision Act of 2005*, also relies heavily on multifactor tests. See H.R. 683, 109th Cong. (2005) (revising 15 U.S.C. § 1125(c)(1) to establish a nonexhaustive four-factor test for famousness and a non-exhaustive six-factor test for blurring); See also Barton Beebe, *A Defense of the New Federal Trademark Antidilution Law*, 16 *Fordham Intell. Prop. Media & Ent. L.J.* 1143 (2006) (proposing a twelve-factor test to determine the likelihood of dilution).

19. See, e.g., 17 U.S.C. § 107 (2001) (setting forth a four-factor test for the fair use of copyrighted material); *Cnty. for Creative Non-Violence v. Reid*, 490 U.S. 730, 751 (1989) (setting forth a nonexhaustive multifactor test to determine whether a hired party is an employee under the general common law of agency). See also Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Cases, 1978-2005* (working paper on file with author and the California Law Review).

20. See *Penn Cent. Transp. Co. v. New York City*, 438 U.S. 104, 124 (1978) (setting forth factors to be considered in resolving certain regulatory takings claims); See also Eduardo Moisés Peñalver, *Regulatory Takings*, 104 *Colum. L. Rev.* 2182, 2193-98 (2004) (discussing *Penn Central*’s progeny and the establishment of various per se rules in takings jurisprudence); Cf. *Tahoe-Sierra Pres. Council, Inc. v. Tahoe Reg’l Planning Agency*, 535 U.S. 302, 342 (2002) (holding that “the interest in ‘fairness and justice’ will be best served by relying on the familiar *Penn Central* approach when deciding [regulatory takings cases involving temporary regulations], rather than by attempting to craft a new categorical rule”).

21. Cf. *Crawford v. Washington* 541 U.S. 36, 63 (2004) (discussing “unpredictability” of multifactor tests for reliability of hearsay evidence).

22. See, e.g., *Reinsurance Co. of Am. v. Administratia Asigurarilor de Stat*, 902 F.2d 1275, 1279-83 (7th Cir. 1990) (applying the multifactor balancing test set forth in Restatement (Second) of Foreign Relations Law of the United States § 40); see *id.* at 1283 (Easterbrook, J., concurring) (expressing a “reluctan[ce] to accept an approach that calls on the district judge to throw a heap of factors on a table and then slice and dice to taste. Although it is easy to identify many relevant considerations, as the ALI’s Restatement does, a court’s job is to reach judgments on the basis of rules of law rather than to use a different recipe for each meal.”).

few.²⁴ For reasons described below, the multifactor test for the likelihood of consumer confusion provides an ideal case study in legal multifactor decision making, one that conduces to the development of a methodology and theoretical toolkit for the study of this form of legal analysis in the many other areas of law in which it is used.

Part I provides background. It briefly reviews the common origin of the multifactor tests for trademark infringement and their current diversity, and explains why judicial application of the tests is particularly amenable to empirical analysis. Part II sets out summary statistics relating to the venue, posture, and multifactor test win rates of the opinions sampled. It reveals significant intercircuit and interdistrict variation in plaintiff multifactor test win rates (the percentage of cases in which the plaintiff prevailed in the multifactor test).

Part III shows that judges employ *fast and frugal*²⁵ heuristics to short-circuit the multifactor test. Perhaps as an expression of their cognitive limitations, but more likely as an expression of their cognitive ingenuity, judges rely upon a few factors or combinations of factors to make their decisions. The rest of the factors are at best redundant and at worst irrelevant. Part III.A uses a series of classification trees to demonstrate that a limited number of core factors drive the test outcomes across the circuits.

23. See, e.g., *Barker v. Wingo*, 407 U.S. 514, 530 (1972) (identifying four factors that courts should weigh in determining whether a defendant has been deprived of his right to a speedy trial); *Doggett v. United States*, 505 U.S. 647, 657 (1992) (applying the Barker factors); see also *id.* at 670 (O'Connor, J., dissenting) (lamenting that "Barker's factors now appear to have taken on a life of their own").

24. See also *Roth Steel Tube Co. v. Comm'r of Internal Revenue*, 800 F.2d 625, 630 (6th Cir. 1986) (setting forth factors to be considered in recharacterizing debt as equity); *In re Bendectin Prods. Liab. Litig.*, 749 F.2d 300, 304 (6th Cir. 1984) (setting forth factors to be considered in issuing writ of mandamus in class action context).

25. See generally *Bounded Rationality: The Adaptive Toolbox* (Gerd Gigerenzer & Reinhard Selten eds., 2001); *Simple Heuristics That Make Us Smart* (Gerd Gigerenzer et al. eds., 1999); Gerd Gigerenzer & Daniel G. Goldstein, *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*, 103 *Psych. Rev.* 650 (1996); cf. A.G.R. McClelland & F. Bolger, *The Calibration of Subjective Probability: Theories and Models 1980-94*, in *Subjective Probability* 453 (George Wright & Peter Ayton eds. 1994). Gigerenzer and his collaborators use the phrase fast and frugal to describe:

satisficing algorithms [which] operate with simple psychological principles that satisfy the constraints of limited time, knowledge, and computational might, rather than those of classical rationality. At the same time, they are designed to be fast and frugal without a significant loss of inferential accuracy, because the algorithms can exploit the structure of environments.

Gigerenzer & Goldstein, *supra* at 651. In contrast to the heuristics and biases tradition, the fast and frugal view insists that fast and frugal heuristics are *not* a source of flawed decision making. On the contrary, "whereas the heuristics-and-biases program portrays heuristics as a frequent hindrance to sound reasoning, rendering *Homo sapiens* not so sapient, we see fast and frugal heuristics as enabling us to make reasonable decisions and behave adaptively in our environment—*Homo sapiens* would be lost without them." Gerd Gigerenzer & Peter M. Todd, *Fast and Frugal Heuristics: The Adaptive Toolbox*, in *Simple Heuristics That Make Us Smart* 3, 29 (1999). Indeed, Gigerenzer and his collaborators assert that fast and frugal heuristics in many cases "outperform" rational inference in that they achieve satisfactory results while requiring less time, information, and computation than more rational, integrative decision strategies. See, e.g., Gigerenzer & Goldstein, *supra* at 660.

Courts have long asserted that no single factor in the multifactor test is dispositive.²⁶ The data contradict this, and do so in dramatic fashion. Part III.B reveals the degree to which judges stampede specific factor outcomes to conform to or support the overall test outcome. The data suggest that judges determine the test outcome based on a limited number of core factors and then adjust the rest of the factor outcomes to accord with that result. This represents strong evidence of *coherence-based reasoning*²⁷ in the courts.

Part IV addresses specific characteristics of each of the core factors and pays particular attention to intercircuit variations in the application of these factors. The data on the intent, strength, and actual confusion factors are of special interest and run counter to conventional wisdom. In light of this study's findings, Part V proposes principles for the formulation of a national standard multifactor test for trademark infringement and suggests specific language. Part VI concludes.

I

Background

A. *The Multifactor Tests Across the Circuits*

Of the multifactor tests for trademark infringement, one commentator has observed that “[a]fter a [brief] period of disparity, the lists developed by the various federal circuits have converged; differences from one list to another have become fairly minimal.”²⁸ This is not accurate. The “period of disparity” was not brief and we are still in it.

26. See, e.g., *Team Tires Plus, Ltd. v. Tires Plus, Inc.*, 394 F.3d 831, 833 (10th Cir. 2005) (“[N]o single factor is dispositive.”); *Gateway, Inc. v. Companion Prods., Inc.*, 384 F.3d 503, 509 (8th Cir. 2004) (“No single factor is dispositive.”); *AHP Subsidiary Holding Co. v. Stuart Hale Co.*, 1 F.3d 611, 616 (7th Cir. 1993) (“None of the seven confusion factors alone is dispositive in a likelihood of confusion analysis.”); *Plus Prods. v. Plus Disc. Foods, Inc.*, 722 F.2d 999, 1004 (2d Cir. 1983) (“No single *Polaroid* factor is determinative.”); *Lever Bros. v. Am. Bakeries Co.*, 693 F.2d 251, 253 (2d Cir. 1982) (“No single . . . factor is preeminent, nor can the presence or absence of one determine, without analysis of the others, the outcome of an infringement suit.”).

27. Coherence-based reasoning is a model of human decision making recently proposed and tested empirically by Professor Dan Simon and others. This model hypothesizes that decision making proceeds “bidirectionally:” premises and facts determine conclusions but those conclusions also inform and alter the premises and facts on which they are based. The decision maker’s mental model of the decision task cycles towards a state of maximal possible coherence among premises, facts, and conclusion, each modifying the others. This results in a skewing of facts and premises towards inflated support for the ultimate decision. See generally Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. Chi. L. Rev. 511 (2004). See also *infra* notes 102-103 and accompanying text.

28. McKinley, *supra* note 14, at 243; see also Note, *Confusion in Cyberspace: Defending and Recalibrating the Initial Interest Confusion Doctrine*, 117 Harv. L. Rev. 2387, 2392 n.27 (2004) (“Although the factors of this test vary from circuit to circuit, there is little substantive variation among the tests.”); Cf. Bierman & Wexler, *supra* note 14, at 2 (“Although the circuits differ as to the precise formulation of the test for the likelihood of confusion, there is broad general agreement as to the factors which should be considered.”).

The idiosyncrasies of tradition rather than of reason governed the development of the multifactor tests across the circuits. Each of the circuits' current multifactor tests originated either directly or indirectly from the 1938 *Restatement (First) of Torts*,²⁹ and this may largely account for the current muddle. Due to a controversy of the time concerning the proper scope of trademark rights,³⁰ the *Restatement (First)* failed to set forth a single, unified multifactor test for trademark infringement. Instead, it proposed a four-factor test to be considered in cases in which the parties' goods were competitive, that is, substitutable,³¹ and an additional nine-factor test to be considered in cases where the parties' goods were not competitive.³² Initially, most circuits followed the example of the *Restatement (First)* and applied one multifactor test to competitive goods cases and another to non-competitive goods cases. This distinction eventually broke down, however, and the circuits each began to use a single, unified multifactor test regardless of whether the parties' goods were competitive or not.³³ For the most part, the peculiarities of the particular cases in which the circuit's multifactor test first coalesced determined which factors the circuit still considers today. This is certainly true of the influential *Polaroid* factors in the Second Circuit,³⁴ the *Roto-Rooter* factors

29. See, e.g., *AMF Inc. v. Sleekcraft Boats*, 599 F.2d 341, 348 n.11 (9th Cir. 1979) (citing to the *Restatement (First) of Torts* § 729); *Scott Paper Co. v. Scott's Liquid Gold, Inc.*, 589 F.2d 1225, 1229 (3d Cir. 1978) (citing to *Scott Paper Co. v. Scott's Liquid Gold, Inc.*, 439 F. Supp. 1022, 1036-37 (D. Del. 1977) (citing to the *Restatement (First) of Torts* §§ 729, 731)); *Roto-Rooter Corp. v. O'Neal*, 513 F.2d 44, 45 (5th Cir. 1975) (citing *Am. Foods, Inc. v. Golden Flake, Inc.*, 312 F.2d 619 (5th Cir. 1963) (citing to the *Restatement (First) of Torts* § 730)); *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492, 495 (2d Cir. 1961) (citing to the *Restatement (First) of Torts* §§ 729-31); See also 3 McCarthy, *supra* note 2, at §23:19.

30. The controversy concerned whether a trademark owner should have the right to exclude the use of a confusingly similar mark only on "directly competitive" (directly substitutable) goods or services, or should have the right to exclude the use of such a mark at least on "goods of the same descriptive properties," if not more broadly on any goods or services where consumer confusion was likely. Compare *Borden Ice Cream Co. v. Borden's Condensed Milk Co.*, 201 F. 510 (7th Cir. 1912) (finding no infringement of Borden as used on condensed milk by Borden as used on ice cream) with *Yale Elec. Corp. v. Robertson*, 26 F.2d 972 (2d Cir. 1928) (finding infringement of Yale as used on locks and keys by Yale as used on flashlights and batteries).

31. See *Restatement (First) of Torts* § 729 (1938).

32. See *id.* at § 731.

33. See, e.g., *A&H Sportswear v. Victoria's Secret Stores, Inc.*, 237 F.3d 198, 214-15 (3d Cir. 2000); *Physicians Formula Cosmetics, Inc. v. W. Cabot Cosmetics, Inc.*, 857 F.2d 80, 83 (2d Cir. 1988) ("Although the *Polaroid* test originally was applied to non-competing products . . . it has been expanded to apply where, as here, competing goods are involved."). See also Kirkpatrick, *supra* note 14, at § 2:7. Cf. Milton Handler, *A Personal Note on Trademark and Unfair Competition Law Before The Lanham Act*, 59 *Law & Contemp. Probs.* 5, 10 (1996) (criticizing courts' application of "the *Polaroid* dictum in cases where the products are the same, contrary to the very opening words of the *Friendly* formulation").

34. See *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492 (2d Cir. 1961). The plaintiff in *Polaroid* offered evidence that it might expand into the market for defendant's goods; consequently, the "bridge the gap" factor became one of the *Polaroid* factors, and was then taken up in various forms by the Third, Sixth, Ninth, and D.C. Circuits. *Id.* at 495-96. Similarly, the defendant in *Polaroid* argued

in the Fifth,³⁵ and the *Lapp* factors in the Third.³⁶ In some circuits, however, such as the Seventh, the adoption of specific factors appears to have occurred more or less randomly.³⁷ The result is the great diversity of factors considered by the circuits as represented in Table 1.

Common to all of the circuits' tests are four factors: the similarity of the marks, the proximity of the goods, evidence of actual confusion, and the strength of the plaintiff's mark. A fifth factor, the intent of the defen-

that "that there is no evidence that plaintiff has suffered either through loss of customers or injury to reputation, since defendant has conducted its business with high standards," and thus Judge Friendly considered the "quality of defendant's product" as well. *Id.* at 495. Courts ever since have wondered how the quality of defendants' goods speaks to the question of whether consumers are likely to be confused. See *supra* note 19 and accompanying text.

35. See *Roto-Rooter Corp. v. O'Neal*, 513 F.2d 44, 45 (5th Cir. 1975) ("In this circuit likelihood of confusion is determined by evaluating a variety of factors including the type of trademark at issue; similarity of design; similarity of product; identity of retail outlets and purchasers; identity of advertising media utilized; defendant's intent; and actual confusion."). The *Roto-Rooter* test, which strongly influenced the First, Fourth, and Eleventh Circuits' tests, failed to include a factor that was at the heart of the appeal in that case: the sophistication of the relevant consumer population. The district court in *Roto-Rooter* discounted the plaintiff's anecdotal evidence of actual confusion, finding that the four consumers who testified to having been confused had simply made an "error" which "resulted from carelessness or inadvertence rather than from any confusing similarity." *Id.* at 46. The Fifth Circuit reversed. Apparently, since the very issue on appeal was whether the four consumers who were actually confused were or were not representative of the general sophistication of the relevant consumer population, the consumer sophistication factor did not make it into the opinion's recitation of the (other) factors to be considered. As a result, the Fifth Circuit, and the Fourth and Eleventh Circuits along with it, still do not explicitly consider the consumer sophistication factor.

36. See *Interpace Corp. v. Lapp, Inc.*, 721 F.2d 460, 462-63 (3d Cir. 1983). *Lapp* copied its list of factors from *Scott Paper Co. v. Scott's Liquid Gold, Inc.*, 589 F.2d 1225, 1229 (3d Cir. 1978), where the Third Circuit first consolidated its multifactor test. *Scott Paper* set forth a ten-factor multifactor test. Why ten? The test originates from Judge Stapleton's thorough and lengthy 1977 Delaware District Court opinion in *Scott Paper Co. v. Scott's Liquid Gold, Inc.*, 439 F.Supp. 1022, 1036-37 (D. Del. 1977). Judge Stapleton observed that in a case involving non-competing goods, the court should consider "many of the same factors which are relevant in a competing goods case," of which he listed six. *Id.* at 1036. He then added that there "are additional factors, however, which are particularly relevant in a case involving non-competing goods," of which he listed four more. *Id.* at 1037. At the conclusion of the second list of factors, Judge Stapleton cited, among other authorities, sections 729 and 731 of the Restatement (First). *Id.* at 1037. On appeal, and in reversing, the Third Circuit reworded and bunched together the factors Judge Stapleton considered into one test of ten factors, which is still used in the circuit, *Scott Paper*, 589 F.2d at 1229, though the list is now generally known as the "*Lapp* factors." This ten-factor test is, of course, what the Restatement (First) test would have looked like had it be unified a half-century earlier into a single test.

37. See *Helene Curtis Industries, Inc. v. Church & Dwight Co., Inc.*, 560 F.2d 1325, 1330 (7th Cir. 1977). The Seventh Circuit's *Helene Curtis* multifactor test is drawn from Second Circuit case law, but not from *Polaroid*. Up until the 1970s, some Second Circuit opinions used an alternative seven-factor test that did not consider bridge the gap or quality, but did consider "the area and manner of concurrent use." The otherwise unremarkable 1970 Second Circuit *Carl Zeiss Stiftung* case referenced this alternative test. See *Carl Zeiss Stiftung v. VEB Carl Zeiss Jena*, 433 F.2d 686, 705 (2d Cir. 1970). Seven years later, in its *Helene Curtis* opinion, the Seventh Circuit quoted from *Carl Zeiss Stiftung* as the sole authority for its own seven-factor test, which is still the test used in that circuit. *Helene Curtis*, 560 F.2d at 1330. It is not clear from *Helene Curtis* whether the Seventh Circuit chose the alternative authority on principle or simply because it was the case most readily at hand.

dant, is found in all but the Federal Circuit's test.³⁸ For purposes of comparison, Table 1 also indicates the factors proposed by the original *Restatement (First) of Torts* and the current *Restatement (Third) of Unfair Competition*. Despite the great weight that many courts and commentators have long purported to place on the evidence of actual confusion factor,³⁹ it was never listed in the *Restatement (First)*. Similarly, despite the centrality of the proximity of the goods factor to the current likelihood of confusion analysis, it does not explicitly appear among the factors proposed by the *Restatement (Third)*.⁴⁰

Beyond the five core factors, the circuits currently consider a wide variety of additional factors. Also, notwithstanding the apparent intersections and equalities among certain circuits' sets of factors, the precise wording of their individual factors varies, sometimes dramatically. This often leads to strikingly different forms of analysis. Part IV below will address this phenomenon in detail.

Finally, with respect to each circuit's ordering of its factors, the similarity factor tends to come early on in the various tests as does the strength factor. Otherwise, the ordering of the factors is altogether haphazard. If the circuits gave any attention to the issue, it is not apparent from the opinions in which each first established its test. This is particularly surprising (or disturbing) in light of the well-established social science finding, if not the simple intuition, that the ordering of cues is of decisive importance to the success of a decision making strategy.⁴¹

B. *The Amenability of the Multifactor Tests to Empirical Study*

The story of American trademark doctrine over the past century is at least in part a story of flexible and intensely pragmatic practitioner- and judge-made rules of thumb slowly degenerating into inflexible and formal doctrine. Lawyers and judges undoubtedly remain as flexible and pragmatic in their decision making as they ever were, but the doctrine has tended to take on a life of its own and to demand that its forms be followed, however perfunctorily. This certainly appears to have been the case

38. Intent is not explicitly listed among the *DuPont* factors. However, the Federal Circuit will consider it when it is relevant. See Richard L. Kirkpatrick, *Likelihood of Confusion Issues: The Federal Circuit's Standard of Review*, 40 Am. U. L. Rev. 1221, 1223 n.5 (1991). See also *Univ. of Notre Dame du Lac v. J.C. Gourmet Foods Imports Co.*, 703 F.2d 1372, 1374 (Fed. Cir. 1983) (intent is "pertinent to a determination of likelihood of confusion").

39. See *infra* Part IV.A.5.

40. Cf. *Restatement (Third) of Unfair Competition* § 21 cmt. j, at 233 (1995) (discussing the "[c]ompetitive proximity of the goods" (italics omitted)).

41. See, e.g., Peter K. Todd & Anja Dieckmann, *Heuristics for Ordering Cue Search in Decision Making*, in *Advances in Neural Information Processing Systems* 1393 (L.K. Saul, Y. Weiss, & L. Bottou eds., 2005) (discussing the importance of cue ordering in the context of lexicographic decision heuristics).

TABLE 1
FACTORS CONSIDERED BY CIRCUIT

Factor	Circuit																			Total circuits considering the factor	Restatement (First) Torts Factors	Restatement (Third) Unfair Competition Factors
	1	2	3	4	5	6	7	8	9	10	11	D.C.	Fed.									
1 Similarity of the marks	(1)	(2)	(1)	(2)	(2)	(3)	(1)	(2)	(1)	(1)	(2)	(2)	(1)	13	\$ 729(a)	\$ 21(a)						
2 Proximity of the goods	(2)	(3)	(9)	(3)	(3)	(2)	(2)	(3)	(3)	(4)**	(3)	(3)	(2)	13	\$ 731(a),(c)	--						
3 Evidence of actual confusion	(6)	(5)	(6)	(7)	(7)	(4)	(6)	(5)	(7)	(3)	(7)	(5)	(7)	13	--	\$ 23(1)						
4 Strength of plaintiff's mark	(8)	(1)	(2)	(1)	(1)	(1)	(5)	(1)	(2)	(6)	(1)	(1)	(5)	13	\$ 731(f)	\$ 21(d)						
5 Defendant's intent	(7)	(6)	(5)	(6)	(6)	(7)	(7)	(4)	(6)	(2)	(6)	(6)	--	12	§§ 729(b), 731(i)	\$ 22						
6 Sophistication of the consumers	(5)	(8)	(3)	--	--	(6)	(4)	(6)	(5)	(5)	--	(8)	(4)	10	§§ 729(d), 731(g)	\$ 21(c)						
7 Similarity of advertising methods, marketing methods	(4)	--	(7)	(5)	(5)	(5)	(3)*	--	(4)	(4)**	(5)	--	--	9	§§ 729(c), 731(d)	\$ 21(b)						
8 Similarity of sales facilities	(3)	--	--	(4)	(4)	--	(3)*	--	--	--	(4)	--	(3)	6	--	--						
9 Likelihood of bridging the gap	--	(4)	(10)	--	--	(8)	--	--	(8)	--	--	(4)	--	5	\$ 731(b)	\$ 21(c)						
10 Comparative quality of the goods	--	(7)	--	--	--	--	--	--	--	--	--	(7)	--	2	--	--						
11 Length of time of concurrent use without evidence of actual confusion	--	--	(4)	--	--	--	--	--	--	--	--	--	(8)	2	\$ 731(b)	\$ 23(2)						
12 Extent to which targets of parties' sales efforts are the same	--	--	(8)	--	--	--	--	--	--	--	--	--	--	1	\$ 731(c)	--						
13 Third-party uses of plaintiff's mark	--	--	--	--	--	--	--	--	--	--	--	--	(6)	1	--	--						
14 Variety of goods on which mark is used	--	--	--	--	--	--	--	--	--	--	--	--	(9)	1	--	--						
15 Market interface between applicant and owner of prior mark	--	--	--	--	--	--	--	--	--	--	--	--	(10)	1	--	--						
16 Extent of applicant's right to exclude	--	--	--	--	--	--	--	--	--	--	--	--	(11)	1	--	--						
17 Extent of potential confusion, i.e., whether de minimis or substantial	--	--	--	--	--	--	--	--	--	--	--	--	(12)	1	--	--						
18 Any other established fact probative of effect of use	--	--	--	--	--	--	--	--	--	--	--	--	(13)	1	--	--						
19 Strength of plaintiff's mark in defendant's geographically-distinct market	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	21(f)						
Total factors considered:	8	8	10	7	7	8	7	6	8	6	7	8	13	--	10	9						

* The Seventh Circuit tends to consider both the similarity of advertising methods and the similarity of sales facilities under factor three. **The Tenth Circuit considers in factor four the "similarity of products and manner of marketing."

with the multifactor test for the likelihood of consumer confusion. The *Restatement (First)* merely stated that “the following factors are important,”⁴² and the cases in which the various circuits first promulgated their respective tests contain similarly precatory language.⁴³ Nevertheless, the multifactor analysis has since become an essentially compulsory and formal exercise.

While such formalism is regrettable, it also makes the empirical study of the multifactor test possible. The empirical study of judicial reasoning is notoriously problematic,⁴⁴ perhaps even more so than the empirical study of judicial outcomes (win rates).⁴⁵ Nevertheless, the highly routinized and explicit manner in which most district courts employ the multifactor test makes it especially susceptible to empirical methods. Typically, the district court begins the multifactor analysis by citing to or quoting from circuit authority and listing in numerical order the precise wording of the factors considered in its circuit. It may also state various general principles: the test should not be applied mechanically;⁴⁶ the list of factors is not exhaustive;⁴⁷ some factors may be more important than others, while some may be irrelevant;⁴⁸ the outcome of the test should not be driven simply by determining which party has won the most factors.⁴⁹ The court then proceeds methodically through each factor, often giving it its own heading, and typically states whether—or the degree to which—the factor favors or disfavors a likelihood of confusion, is neutral, irrelevant, or, in a summary judgment analysis, presents an issue of fact. After considering each of the

42. Restatement (First) of Torts §§ 729, 731 (1938).

43. See, e.g., *Helene Curtis Indus., Inc. v. Church & Dwight Co., Inc.*, 560 F.2d 1325, 1330 (7th Cir. 1977) (“In determining ‘likelihood of confusion’ several factors are important . . .”)

44. See, e.g., R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. Pa. L. Rev. 1105, 1125-30 (2004) (discussing and responding to “oft-cited concerns and limitations about using judicial opinions for systematic study” such as unobserved reasoning, selection bias, and strategic behavior).

45. See, e.g., Kevin M. Clermont & Theodore Eisenberg, *Do Case Outcomes Really Reveal Anything About the Legal System? Win Rates and Removal Jurisdiction*, 83 Cornell L. Rev. 581, 587-92 (1998) (discussing how the “case-selection effect” complicates inferences made from plaintiff win rate data); See also Steven Shavell, *Any Frequency of Plaintiff Victory at Trial Is Possible*, 25 J. Legal Stud. 493 (1996).

46. See, e.g., *Scott Fetzer Co. v. House of Vacuums Inc.*, 381 F.3d 477, 485 (5th Cir. 2004) (“[The factors] do not apply mechanically to every case and can only serve as guides, not as an exact calculus.”).

47. See, e.g., *Soc’y of Fin. Exam’rs v. Nat’l Ass’n of Certified Fraud Exam’rs Inc.*, 41 F.3d 223, 228 n.15 (5th Cir. 1995) (referring to the multifactor test as a “nonexhaustive collection of considerations that may be relevant to the ultimate factual determination—Are the actions of the defendant likely to create confusion?”).

48. See, e.g., *Shakespeare Co. v. Silstar Corp. of Am.*, 110 F.3d 234, 242 (4th Cir. 1997) (“The list of these factors, however, is neither exhaustive nor exclusive, and some of the factors may not always be relevant or equally emphasized in each case.”).

49. See, e.g., *Thane Int’l, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894, 901 (9th Cir. 2002) (“The list of factors is not a scorecard—whether a party ‘wins’ a majority of the factors is not the point.”).

factors, courts often explicitly balance the factors and summarize the results of their analysis.⁵⁰

In the process, district courts give every appearance of scrupulously following a basic *weighted additive* decision strategy,⁵¹ and they do so for good reason. In most circuits, district courts that fail to address each factor of the multifactor analysis risk remand or reversal on that basis.⁵² This is especially true in the Second Circuit where the multifactor test is most often applied and where appellate panels have repeatedly emphasized that the multifactor analysis must be exhaustive and explicit.⁵³ The Ninth Circuit appears to be alone in declining to establish a “rigid test for analyzing likelihood of confusion in trademark cases”⁵⁴ and allowing district courts to consider a “subset”⁵⁵ of factors. Particularly in the Internet context, the Ninth Circuit has counseled against “excessive rigidity.”⁵⁶ Overall, the circuit-wide mean of the proportion of factors not explicitly addressed per opinion is very low (.096),⁵⁷ and notwithstanding the Ninth Circuit’s stated liberality, the mean in the Ninth Circuit (.113)⁵⁸ is not significantly

50. At least one court has gone so far as to conclude its analysis with a helpful table showing which way each factor tilts. *See Therma-Scan, Inc. v. Thermoscan, Inc.*, 118 F. Supp. 2d 792, 804 (E.D. Mich. 2000).

51. *See* John W. Payne et al., *The Adaptive Decision Maker* 24 (1993) (“The weighted additive rule considers the values of each alternative on all the relevant attributes and considers all the relative importances or weights of the attributes to the decision maker. Further, the conflict among values is assumed to be confronted and resolved by explicitly considering the extent to which one is willing to trade off attribute values, as reflected by the relative importances or weights.”) (internal emphasis omitted).

52. *Cf.* Richard A. Posner, *Judicial Behavior and Performance: An Economic Approach*, 32 Fla. St. U. L. Rev. 1259, 1271 (2005) (discussing judges’ aversion to being reversed).

53. *See, e.g., Arrow Fastener Co., Inc. v. Stanley Works*, 59 F.3d 384, 400 (2d Cir. 1995). The *Arrow Fastener* court wrote:

[I]t is incumbent upon the district judge to engage in a deliberate review of each factor, and, if a factor is inapplicable to a case, to explain why. . . . The steady application of *Polaroid* is critical to the proper development of trademark law, for it is only when the *Polaroid* factors are applied consistently and clearly over time that the relevant distinctions between different factual configurations can emerge. Litigants are entitled to the illumination and guidance this common-law process affords, and appellate courts depend on it for the performance of their assigned task of review. . . . The efficacy of the multi-factor approach that Judge Friendly wisely set out to address this difficult situation depends on thorough, careful, and consistent application of the doctrine by district courts.

Id.

54. *Moscow Distillery Cristall v. Pepsico, Inc.*, 141 F.3d 1177 (Table), 1998 WL 101696, at *2 (9th Cir. 1998).

55. *See* *Thane Int’l, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894, 901 (9th Cir. 2002) (defending the analysis of a “subset” of factors in the non-internet context); *Brookfield Commc’ns, Inc. v. W. Coast Entm’t Corp.*, 174 F.3d 1036, 1054 (9th Cir. 1999) (endorsing a “subset” of the three “most important” factors in the internet context).

56. *Brookfield Commc’ns*, 174 F.3d at 1054.

57. SD=.177, n=331.

58. SD=.181, n=53.

different from the mean of all other circuits (.093),⁵⁹ or, for that matter, from the Second Circuit's (.072).⁶⁰

District courts also tend to adhere very closely to their circuits' specific tests. Of the 331 opinions sampled, only six (2%) explicitly considered factors beyond those included in their circuit's test,⁶¹ and in doing so, only one of these looked to precedent from other circuits.⁶²

What we are left with is a collection of multifactor tests each explicitly and uniformly applied in their respective circuits. This creates a stable platform for empirical analysis. The irony is that this uniformity of application within each circuit may very well generate disparities in outcomes across the circuits. This is the theme of Part II.

II

Summary Statistics

Most trademark lawyers have been content to assume—or, at least, hope—that given the same facts, the circuits' various tests would yield the same outcome. This Part suggests that this long-held assumption is incorrect. Different tests appear to yield different outcomes. First, however, to build a foundation for this claim, the Part briefly considers the venue and posture of the opinions sampled.⁶³

A. *Venue and Posture of Opinions by Circuit*

For the 331 opinions sampled, Table 2 sets out the number and percentage of opinions sampled per circuit, as well as their procedural posture. As expected, the district courts of the Second Circuit contributed a large plurality of opinions to the sample, producing nearly one-third of the total

59. SD=.176, n=278.

60. SD=.163, n=107.

61. See *1-800 Contacts, Inc. v. WhenU.com*, 309 F. Supp. 2d 467, 503-05 (S.D.N.Y. 2003) (considering "branding" as an additional factor); *Taylor Made Golf Co. v. MJT Consulting Group, LLC*, 265 F. Supp. 2d 732, 742-44 (N.D. Tex. 2003) (considering "(9) the extent and nature of the changes made to the product, (10) the clarity and distinctiveness of the labeling on the rebuilt product, and (11) the degree to which any inferior qualities associated with the reconditioned product would likely be identified by the typical purchaser with the manufacturer" (citation omitted)); *Savin Corp. v. Savin Group*, No. 02 Civ. 9377 (SAS), 2003 U.S. Dist. LEXIS 19220, at *35-37 (S.D.N.Y. Oct. 28, 2003) (considering initial interest confusion as an additional factor); *A&H Sportswear Co. v. Victoria's Secret Stores, Inc.*, 167 F. Supp. 2d 770, 798-99 (E.D. Pa. 2001) (considering as additional factors "(1) commercial impression; (2) survey evidence; (3) continued success of the Plaintiff's MIRACLESUIT line"); *Urgent Gear Inc. v. Savoia*, No. 3:01-CV-2190-D, 2001 U.S. Dist. LEXIS 20459, at *10-15 (N.D. Tex. Dec. 10, 2001) (considering consumer sophistication); *Amazon, Inc. v. Canondale, Inc.*, No. 99 N 571, 2000 U.S. Dist. LEXIS 17864, at *41-46 (D. Colo. July 24, 2000) (considering "prior relationship of the parties" and "de minimis use").

62. See *Amazon*, 2000 U.S. Dist. LEXIS 17864, at *41-46.

63. Appendix A describes the selection and coding of the opinions. Appendix B considers the reliability of certain trademark-related data contained in the Federal Court Cases: Integrated Data Base of the Administrative Office of the United States Courts.

opinions and nearly one-half of the bench trial opinions.⁶⁴ Also as expected, preliminary injunction opinions formed by far the most common posture. The court engaged in no further application of the multifactor test in all but four of the 145 cases that produced a preliminary injunction opinion. The data are thus consistent with the conventional view that New York is the primary venue and the preliminary injunction hearing the primary forum for the adjudication of trademark infringement claims. There was no significant variation across the years sampled in the number of opinions by circuit or posture.⁶⁵

B. *Multifactor Test Win Rates by Circuit and Posture*

Table 2 also reports the rate at which plaintiffs or, in the case of summary judgment motions, movants won the multifactor test in the opinions

64. The Southern District of New York in particular accounted for 78% of the opinions in the Second Circuit and 25% of the opinions nationally. The two other leading districts were the Northern District of Illinois and the Central District of California, accounting for 11% and 8% of the opinions in the sample, respectively (and accounting for 85% of the opinions in the Seventh Circuit and 47% of the opinions in the Ninth Circuit, respectively).

65. The data reveal certain other aspects of the opinions sampled that will be of particular interest to trademark specialists. With respect to the nature of the trademark property at issue, 81% (267) of the opinions addressed a claim for the infringement of a trademark only, 16% (53) addressed a claim for the infringement of trade dress only, and 3% (11) addressed a claim for the infringement of both forms of trademark property. Of the seven trademark cases the Supreme Court has considered since 1992, four have been trade dress cases. See *TrafFix Devices, Inc. v. Marketing Displays, Inc.*, 532 U.S. 23 (2001); *Wal-Mart Stores, Inc. v. Samara Bro., Inc.*, 529 U.S. 205 (2000); *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159 (1995); and *Two Pesos, Inc. v. Taco Cabana, Inc.*, 505 U.S. 763 (1992). The data suggest, however, that such cases are not in the mainstream of reported trademark litigation.

Despite the well-deserved attention paid to them in the commentary, see, e.g., Jennifer E. Rothman, *Initial Interest Confusion: Standing at the Crossroads of Trademark Law*, 27 *Cardozo L. Rev.* 105 (2005), claims of initial-interest and post-sale confusion also appear to be outside of the mainstream of trademark litigation. Only 12% (forty) of the opinions sampled considered a claim of initial-interest confusion and only 6% (twenty-one) of the opinions sampled found initial-interest confusion. The Second Circuit appears to be particularly hostile to claims of initial-interest confusion: it found initial-interest confusion in only two of the twelve opinions which considered it. In contrast, the Seventh Circuit found initial-interest confusion in each of the six opinions which considered it. As for post-sale confusion, only 6% (nineteen) of the opinions sampled considered a claim of post-sale confusion, and thirteen of these came from the Second Circuit, which found post-sale confusion in only three of these opinions. Overall, post-sale confusion was found in 1.5% (five) of the opinions sampled.

Also of interest is the exceptionally low proportion of cases in which the defendant made out a defense of parody, another area of the law to which trademark scholarship has long been especially attentive. See, e.g., Julie Zando-Dennis, Note, *Not Playing Around: The Chilling Power of the Federal Trademark Dilution Act of 1995*, 11 *Cardozo Women's L.J.* 599 (2005). Of the seven opinions which considered a defense of parody, five were from the Second Circuit, and only one of the seven found in favor of the plaintiff. This is consistent with the conventional wisdom that those defendants who have the wherewithal to litigate the issue are generally successful in doing so.

Finally, nineteen of the opinions sampled addressed claims of reverse confusion. Remarkably, the plaintiff prevailed in only two of these opinions for a win rate of .105 as against a win rate of .452 in opinions not addressing a claim of reverse confusion.

TABLE 2
DISTRIBUTION AND MULTIFACTOR TEST WIN RATE OF OPINION POSTURES BY CIRCUIT
(PW = plaintiff multifactor test win rate; DW = defendant multifactor test win rate)

Circuit	Distribution by Circuit		Preliminary Injunction		Bench Trial		SJ Plaintiff		SJ Defendant		SJ Cross		
	N	%	N	PW	N	PW	N	PW	N	DW	N	PW	DW
1	10	3.0	6	.833	0	---	1	1.000	3	1.000	0	---	---
2	107	32.3	42	.405*	22	.409	10	.800	23	.783	10	.100	.200
3	28	8.5	12	.417	6	.667	1	.500	5	1.000	4	.250	.500
4	16	4.8	4	.500	1	1.000	4	1.000	2	1.000	5	.600	.200
5	9	2.7	1	1.000	2	.500	1	1.000	2	.500	3	.667	.000
6	27	8.5	14	.500	2	.500	1	1.000	5	.800	5	.400	.400
7	41	12.4	17	.588	3	.667	6	.333	7	.429	8	.250	.250
8	19	5.7	11	.546	2	.500	1	.000	1	1.000	4	.000	.250
9	53	16.0	29	.689*	4	.500	8	.875	8	.625	4	.500	.250
10	5	1.5	3	.667	0	---	0	---	1	1.000	1	.000	.000
11	14	4.2	6	.500	3	.667	1	1.000	2	.500	2	.000	.500
D.C	2	6.0	1	.000	1	1.000	0	---	0	---	0	---	---
N	331	100.0	146		46		34		59		46		
% of Total Opinions:			44.1		13.9		10.3		17.8		13.9		
Circuit-Wide:			.534		.521		.743		.746		.283		.261

*Significantly different from the mean win rate of all other circuits at the .05 level of confidence.

sampled. It must be emphasized that the data set consisted only of litigated federal trademark infringement cases that produced written opinions available from the Westlaw and Lexis databases. Furthermore, as explained in Appendix A, the data set excluded counterfeiting and licensing fact-patterns.⁶⁶ For these reasons, the inferences we can draw from the multifactor test win rate results are limited.⁶⁷ Nevertheless, two general observations may be made.

First, there is substantial intercircuit variation in plaintiff multifactor test win rates. Across the opinions sampled, the Second Circuit yielded a significantly lower overall plaintiff multifactor test win rate (37%) than all other circuits (51%).⁶⁸ Meanwhile, the Ninth Circuit yielded a significantly higher overall plaintiff win rate (64%) than the rest of the circuits (43%).⁶⁹ This result may largely stem from the differences in the plaintiff multifactor test win rates in preliminary injunction opinions: in the Second Circuit, the plaintiff win rate in such opinions (41%) was significantly lower than that in all other circuits (59%),⁷⁰ while, in the Ninth Circuit, the plaintiff win rate (69%) was higher than all other circuits (50%), though the difference was marginally significant.⁷¹ Bench trial plaintiffs also fared relatively poorly in the Second Circuit. Their multifactor test win rate was the lowest among the circuits.⁷²

More specifically, the Southern and Eastern Districts of New York appeared to be particularly hazardous venues for the litigation of trademark infringement claims—or appeared to attract particularly foolhardy plaintiffs. The Southern District's overall plaintiff multifactor test win rate in the opinions sampled (36%) was well below the overall plaintiff win rate of

66. Furthermore, the win rates reported in Table 2 diverge substantially from those reported for a comparable time period by the Federal Judicial Center's Federal Court Cases: Integrated Data Base. This divergence is addressed in Appendix B.

67. Cf. Clermont & Eisenberg, *supra* note 45, at 581, 592 (referring to win rate data as "probably the slipperiest of all judicial data" and suggesting that interpretation of win rate data is "as much art as science—and it is a difficult and subjective art").

68. Second Circuit: .374; Ninth Circuit: .518; $p=.014$. Here, I am comparing the proportion of district court opinions from the Second Circuit in which the plaintiff won the multifactor test to the proportion of opinions from all other circuits in which the plaintiff won the multifactor test. One way to do so is to test the null hypothesis that, given the sample data, the two population proportions could in fact be equal. This test produces a test statistic z . The "P-value" of this test statistic expresses "the probability, when [the null hypothesis] is true, of a test statistic value at least as contradictory to [the null hypothesis] as the value actually observed. The smaller the P-value, the more strongly the data contradict [the null hypothesis]." Alan Agresti & Barbara Finlay, *Statistical Methods for the Social Sciences* 157 (3d ed. 1997). Cf. *Coates v. Johnson & Johnson*, 756 F.2d 524, 537 n. 13 (7th Cir. 1985) ("[S]tatistical significance also may be expressed as a probability value (P) on a continuous or relative scale ranging from 0 to 1.0. The level of statistical significance rises as the value of the (P) level declines. A (P) value below .05 is generally considered to be statistically significant. . . .").

69. $p=.007$.

70. Second Circuit: .405; all other circuits: .587; $p=.046$.

71. Ninth Circuit: .689; all other circuits: .496; $p=.061$.

72. The difference was not statistically significant.

all other districts (51%).⁷³ The Eastern District's overall plaintiff win rate was even worse (18%),⁷⁴ and was the lowest in the nation among the twenty-seven districts that contributed more than two opinions to the sample. No other districts emerged from the sample as especially pro- or anti-plaintiff.⁷⁵

Second, in the opinions sampled, motions for summary judgment that were not met with cross-motions for summary judgment were significantly more successful than those that were met with cross-motions for summary judgment. A simple selection bias may account for this result: a party is more likely to bring a summary judgment motion when it has strong grounds for doing so, and a judge is more likely to produce a written opinion when she grants a summary judgment motion rather than denies it. Yet the striking disparity in win rates leaves open the possibility that cross-motions for summary judgment tend to some degree to cancel each other out.⁷⁶

III

Interfactor Analysis

It is something of a pastime in trademark law to speculate on which factors, if any, drive the outcome of the multifactor test and how the factors interact. The circuits themselves may largely be responsible for this

73. Southern District of New York: .361, n=83; all other districts: .508, n=248; p=.021.

74. Eastern District of New York: .182, n=11.

75. The nine opinions sampled from the Eastern District of Virginia yielded an overall plaintiff win rate of .778, but five of these opinions involved a claim of trademark infringement by means of a domain name.

76. The overall plaintiff win rates in preliminary injunction and bench trial opinions are consistent with the "fifty percent hypothesis." See generally George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. Legal Stud. 1 (1984) (considering the hypothesis that only close cases will survive settlement and proceed to trial and thus that plaintiffs should win about half of such cases). See also Daniel Kessler et al., *Explaining Deviations from the Fifty-Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation*, 25 J. Legal Stud. 233 (1996); Keith N. Hylton, *Asymmetric Information and the Selection of Disputes for Litigation*, 22 J. Legal Stud. 187 (1993); Theodore Eisenberg, *Testing the Selection Effect: A New Theoretical Framework with Empirical Tests*, 19 J. Legal Stud. 337 (1990). William Landes hypothesizes that intellectual property plaintiffs should enjoy higher win rates at trial than other civil plaintiffs because intellectual property plaintiffs face the risk that their intellectual property will be declared invalid. Thus, they are more likely to settle or withdraw from even the near-close cases. See William M. Landes, *An Empirical Analysis of Intellectual Property Litigation: Some Preliminary Results*, 41 Hous. L. Rev. 749 (2004). Landes finds confirmation for this hypothesis in data provided by the Administrative Office of the United States Courts. *Id.* at 771-72. Appendix B explains why this data should not be trusted. Even so, if we except Landes's theoretical point, then how can we explain trademark plaintiffs' near-fifty percent win rates in the sampled preliminary injunction and bench trial opinions? Perhaps for trademark plaintiffs, the risk that their trademark will be declared invalid is offset by the well-established doctrinal principle that a trademark owner must aggressively protect its mark from use by third-parties lest the mark lose its source-distinctiveness and thus its validity. See, e.g., *Herman Miller, Inc. v. Palazzetti Imports and Exports, Inc.*, 270 F.3d 298, 317 (6th Cir. 2001) (discussing how failure to prosecute may lead to the loss or weakening of trademark rights). For this reason, trademark plaintiffs are prone to proceed to trial unless the defendant agrees to discontinue use of the mark at issue.

custom, as most have identified at one time or another certain individual factors or combinations of factors as being the most important in their particular circuit, if not in general. The Second Circuit, for example, has pointed to the similarity of the marks and the proximity of the goods,⁷⁷ or to the strength of the plaintiff's mark, similarity, and proximity;⁷⁸ the Third to similarity;⁷⁹ the Fourth to evidence of actual confusion;⁸⁰ the Sixth to proximity;⁸¹ the Seventh to similarity, the defendant's intent, and actual confusion;⁸² the Ninth to similarity, proximity, and the commonality of the parties' marketing channels;⁸³ and the Eleventh to the "type of mark" and actual confusion.⁸⁴ In identifying these factors as the leading factors, the circuits rarely specify whether they mean that these factors are in fact the most important, or simply should in principle be the most important, or both.

77. *Vitarroz Corp. v. Borden, Inc.*, 644 F.2d 960, 966 (2d Cir. 1981) ("Our cases demonstrate, moreover, that the likelihood of confusion, the 'crucial' issue in a case such as this, often depends on the similarity of the marks and the proximity of the products." (citation omitted)). *See also* *Lexington Mgmt. Corp. v. Lexington Capital Partners*, 10 F. Supp. 2d 271, 288 (S.D.N.Y. 1998) ("In this case I am influenced by the Court of Appeals' recognition that determining likelihood of confusion 'often depends on the similarity of the marks and the proximity of the products.'" (citing *Vitarroz*)).

78. *Pfizer, Inc. v. Y2K Shipping & Trading, Inc.*, No. 00 CV 5304(SJ), 2004 U.S. Dist. LEXIS 10426, at *9 (E.D.N.Y. Mar. 30, 2004) ("The first three factors are considered the most important; however, no one factor is determinative and the list is not exhaustive." (citing *Mobil Oil Corp. v. Pegasus Petroleum Corp.*, 818 F.2d 254, 256-58 (2d Cir. 1987))). *See also* *Starbucks Corp. v. Wolfe's Borough Coffee, Inc.*, No. 01 Civ. 5981 (LTS) (THK), 2005 U.S. Dist. LEXIS 35578, at *19 (S.D.N.Y. Dec. 23, 2005) (quoting *Pfizer*).

79. *See, e.g., Checkpoint Sys., Inc. v. Check Point Software Techs., Inc.*, 269 F.3d 270, 281 (3d Cir. 2001) ("The single most important factor in determining likelihood of confusion is mark similarity."); *A&H Sportswear, Inc. v. Victoria's Secret Stores, Inc.*, 237 F.3d 198, 216 (3d Cir. 2000); *M.D. On-Line, Inc. v. WebMD Corp.*, No. 05-CV-4081 (WJM), 2005 WL 2469668, at *3 (D.N.J. Oct. 6, 2005); *Bijur Lubricating Corp. v. Devco Corp.*, 332 F. Supp. 2d 722, 728 (D.N.J. 2004); *Penn. Bus. Bank v. Biz Bank Corp.*, 330 F. Supp. 2d 511, 518 (E.D. Pa. 2004).

80. *Lyons P'ship, L.P. v. Morris Costumes, Inc.*, 243 F.3d 789, 804 (4th Cir. 2001). *See also* *Giant Brands, Inc. v. Giant Eagle, Inc.*, 228 F. Supp. 2d 646, 654 (D. Md. 2002) (citing *Lyons* for the proposition that "[a]ctual confusion is the final and most important *Pizzeria Uno* factor used to determine likelihood of confusion.") (emphasis in original).

81. *See* *Homeowners Group, Inc. v. Home Mktg. Specialists, Inc.*, 931 F.2d 1100, 1109 (6th Cir. 1991) (referring to the "relatedness" of the goods as "the important inquiry in the likelihood of confusion determination"). *See also* *Comerica Inc. v. Fifth Third Bankcorp*, 282 F. Supp. 2d 557, 570 (E.D. Mich. 2003) (citing *Homeowners Group* for same).

82. *See* *Eli Lilly & Co. v. Natural Answers, Inc.*, 233 F.3d 456, 462 (7th Cir. 2000) ("[A]lthough no one factor is decisive, the similarity of the marks, the intent of the defendant, and evidence of actual confusion are the most important considerations."); *G. Heileman Brewing Co. v. Anheuser-Busch, Inc.*, 873 F.2d 985, 999 (7th Cir. 1989) (same).

83. *See* *GoTo.com, Inc. v. Walt Disney Co.*, 202 F.3d 1199, 1205 (9th Cir. 2000). *See also* *Tillamook Country Smoker, Inc. v. Tillamook County Creamery Ass'n*, 311 F. Supp. 2d 1023, 1041 (D. Or. 2004) (referring to the similarity, proximity, and marketing channels factors as the "controlling troika" of factors); *CSC Brands LP v. Herdez Corp.*, 191 F. Supp. 2d 1145, 1149 (E.D. Cal. 2001) (same). *But see* *Playmakers, LLC v. ESPN, Inc.*, 297 F. Supp. 2d 1277, 1280 (W.D. Wash. 2003) (stating that the "controlling troika" analysis is confined to internet cases).

84. *See* *Dieter v. B & H Indus. of Sw. Fla., Inc.*, 880 F.2d 322, 326 (11th Cir. 1989) ("[T]he type of mark and the evidence of actual confusion are the most important in this circuit.").

This Part seeks to settle the debate, at least with respect to the question of *is* rather than *ought*. It shows that, in practice, a limited number of core factors determine the outcome of the test, and tend in the process to stampede the rest of the factors. The similarity of the marks factor is by far the most influential. Two other factors are decisive: the defendant's intent factor, but only when it favors a likelihood of confusion, and the proximity (or relatedness) of the parties' goods factor, but only when it disfavors a likelihood of confusion. The intent and actual confusion factors also appear to exert an inordinate degree of influence on the outcomes of the rest of the factors.

To demonstrate the importance of the core factors, I present the data in Subpart A from a variety of perspectives. Regression analysis is inappropriate—or at least unhelpful—here, not only because the data are so skewed,⁸⁵ but also because regression analysis tends to assume a fully integrative model of decision making, a model that this Article rejects.⁸⁶ Instead, I resort to a series of simple classification trees of the multifactor test outcome that classify that outcome according to certain factor outcomes. I then present crosstabulations of the test outcome by factor outcomes and factor outcomes by test outcome. This approach facilitates the

85. As Tables 3 and 4 show, the overall test outcome was invariant for certain factor outcomes. This raises the problem of “zero cell count” in which the dependent variable, here, the outcome of the multifactor test, is invariant for one or more values of an independent variable, for example, the similarity factor. See Scott Menard, *Applied Logistic Regression Analysis* 78-81 (2d ed. 2002). This invariance produces enormous regression coefficients and standard errors that severely limit the utility of the regression results. Furthermore, the factor outcomes tend very strongly to correlate with each other, as is shown in Subpart III.A.4, which raises the problem of multicollinearity. See Jacob Cohen et al., *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* 390 (3d ed. 2003) (describing multicollinearity as occurring “in data sets in which one (or more) of the [independent variables] is highly correlated with the other [independent variables] in the regression equation. The estimate of the regression coefficient B_i for this correlated predictor will be very unreliable because little unique information is available from which to estimate its value.”).

86. This problem is not generally recognized in the legal literature and has only recently been addressed elsewhere. See generally Simple Heuristics, *supra* note 25; Dhami & Harries, *infra* note 99; Gigerenzer & Goldstein, *supra* note 25. Cf. Kenneth Hammond, *Upon Reflection*, 2 *Thinking & Reasoning* 239, 244 (1996) (“[A] sin of commission on my part was to overemphasize the role of the multiple regression (MR) technique as a model for organising information from multiple fallible indicators into a judgment.”). If we accept the proposition that the boundedly rational decision maker does not integrate all relevant information and frequently engages in noncompensatory decision strategies, flexibly choosing primary cues based on the decision task, then a fully integrative model of decision making may lead us astray. Cf. *Societe Anonyme de la Grande Distillerie v. Julius Wile Sons & Co.*, 161 F. Supp. 545, 547 (S.D.N.Y. 1958) (Likelihood of confusion “does not readily lend itself to resolution by scientific appraisal or comparison. Rather, a finding on infringement is by necessity a subjective determination by the trial judge based on his visceral reactions. . . .”). Furthermore, if we accept the proposition that judges make a decision based upon a limited number of core factors and then stampede the rest of the factors to support that decision, then regression analysis may not adequately account for the degree to which the factors influence each other. At the very least, such an analysis may produce a suboptimal fit with the data. See Dhami & Harries, *infra* note 99. Indeed, no regression analysis of the factors considered in this study was able to achieve the degree of fit that a simple five-factor—or two-factor, for that matter—classification tree could achieve.

assessment of which factor outcomes necessitate a particular test outcome and which test outcomes necessitate particular factor outcomes. Finally, I present the results of correlation analysis of the factor outcomes and test outcomes. This analysis will prepare the ground for our consideration in Subpart B of the phenomenon of stampeding.

A. *The Importance of the Core Factors*

A diverse body of empirical work supports the intuition that when confronted with complex decision tasks we seldom seek to consider all relevant information or reduce uncertainty to the maximum extent conceivable, even if we were capable of doing so. Instead, we use various strategies to decide when to stop acquiring and analyzing information and commit to a course of action. Empirical studies of decision making generally,⁸⁷ and of judicial decision making in particular,⁸⁸ consistently show that decision makers, even when making complex decisions, reach their stopping threshold and make a decision after considering a remarkably low number of decision-relevant factors. Social science researchers have demonstrated that in regression-based modeling of human decision making, only a small number of cues—three on average, one author has

87. See, e.g., Gad Saad & J. Edward Russo, *Stopping Criteria in Sequential Choice*, 67 *Organizational Behav. & Human Decision Processes* 258 (1996); Jacob Jacoby et al., *Tracing the Impact of Item-by-Item Information Accessing on Uncertainty Reduction*, 21 *J. Consumer Res.* 291 (1994) (observing that decision makers made uncertain decisions based on a limited number of cues rather than access all cues available to them); R.N. Shepard, *On Subjectively Optimum Selections Among Multi-Attribute Alternatives*, in *Decision Making* 257 (W. Edwards & A. Tversky eds., 1967) (arguing that decision makers rarely consider more than one or two factors at any one time).

88. See, e.g., Mandeep K. Dhami & Peter Ayton, *Bailing and Jailing the Fast and Frugal Way*, 14 *J. Behav. Decision Making* 141 (2001) (demonstrating that “magistrates’ policies were better described and predicted by a fast and frugal model characterized by noncompensatory cue use, than by either of two compensatory integration models”); Vladimir J. Konečni & Ebbe B. Ebbesen, *The Mythology of Legal Decision Making*, 7 *Int’l J. L. & Psychiatry* 5, 7, 15 (1984) (concluding, based on review of empirical literature, that “[e]xtremely simple decision strategies are the rule rather than the exception” in judicial decision making and that “the few factors which are taken into account are generally not those which the decision makers claim they are responsive to”). For studies more specifically of judges’ bounded rationality, see Adam J. Hirsch, *Cognitive Jurisprudence*, 76 *S. Cal. L. Rev.* 1331 (2003) (arguing that judges are boundedly rational); Stephen M. Bainbridge & G. Mitu Gulati, *How Do Judges Maximize? (The Same Way Everybody Else Does—Boundedly): Rules of Thumb in Securities Fraud Opinions*, 51 *Emory L.J.* 83 (2002) (arguing that judges use doctrinal rules of thumb to simplify decision making in securities cases); Hillary A. Sale, *Judging Heuristics*, 35 *U.C. Davis L. Rev.* 903 (2002) (analyzing judges’ use of heuristics in securities fraud cases); Chris Guthrie et al., *Inside the Judicial Mind*, 86 *Cornell L. Rev.* 777 (2001) (demonstrating that judicial decision making is influenced by anchoring, framing, hindsight bias, representativeness, and egocentric bias); Jeffrey J. Rachlinski, *Heuristics and Biases in the Courts: Ignorance or Adaptation?*, 79 *Or. L. Rev.* 61, 93 (2000) (observing that “courts tend to identify and adapt to the influence of cognitive illusion on the determination of issues that juries are likely to resolve and ignore or fall prey to the influence of cognitive illusions on the determination of issues that judges are likely to resolve”); Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 *U. Chi. L. Rev.* 571 (1998) (discussing the legal systems adaptation to and accommodation of hindsight bias).

speculated⁸⁹—emerge as statistically significant.⁹⁰ Perhaps more interestingly for our purposes, empirical work suggests that decision makers tend to use a *core attributes heuristic* by which they stop acquiring and analyzing information once the last in their set of most important, determinant attributes has been acquired and analyzed.⁹¹

Empirical work also suggests that such decision making strategies are rational. Social science research dating back to the 1950s has demonstrated that the consideration of too much information impairs decision-making accuracy.⁹² More recently, a strong body of research in the fast and frugal tradition shows that in certain contexts, we make complex decisions based on a single cue, and that this simple strategy, known as *Take The Best*, is as good as or even outperforms computation-intensive decision making.⁹³

The data collected for this study support both the general hypothesis that decision makers, even when making complex decisions, consider only a small number of factors and the more specific hypothesis that, in doing so, decision makers use a core attributes heuristic.⁹⁴ Furthermore, the data suggest that this decision making strategy need not be understood as a regrettable expression of human cognitive limitations. On the contrary,

89. See Berndt Brehmer, *The Psychology of Linear Judgement Models*, 87 Acta Psychologica 137 (1994).

90. See, e.g., Annica Brehmer & Berndt Brehmer, *What Have We Learned About Human Judgment From Thirty Years of Policy Capturing?*, in *Human Judgment: The SJT View* 75 (Berndt Brehmer & C.R.B. Joyce eds., 1988); Robert Libby & Barry L. Lewis, *Human Information Processing Research in Accounting: The State of the Art in 1982*, 7 Acct., Org., & Soc'y 231 (1982); Paul Slovic & Sarah Lichtenstein, *Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment*, 6 Organizational Behav. & Human Performance 649 (1971).

91. See Saad & Russo, *supra* note 87.

92. See Everett L. Kelly and Donald W. Fiske, *The Prediction of Performance in Clinical Psychology* (1951) (finding that the inclusion of too many predictors makes clinical prediction less efficient and may result in overconfidence). See also David A. Sheluga, James Jaccard, & Jacob Jacoby, *Preference, Search and Choice: An Integrative Approach*, 6 J. Cons. Res. 166 (1979) (finding that consumers who used only the relevant portions of available information made more accurate decisions). See generally Jacob Jacoby, Donald E. Speller, & Carol Kohn Berning, *Brand Choice Behavior as a Function of Information Load: Replication and Extension*, 1 J. Cons. Res. 33 (1974) (demonstrating that decision-making accuracy decreases after a certain degree of information load); Jacob Jacoby, Donald E. Speller, & Carol A. Kohn, *Brand Choice Behavior as a Function of Information Load*, 11 J. Marketing Res. 63 (1974) (studying the effects of information load on brand choice behavior).

93. See, e.g., Jean Czerlinski, Gerd Gigerenzer & Daniel G. Goldstein, *How Good are Simple Heuristics?*, in *Simple Heuristics*, *supra* note 25, at 97 (showing that the Take The Best heuristic outperforms multiple regression across a variety of decision environments). But see Ben R. Newell & David R. Shanks, *Take the Best or Look at the Rest? Factors Influencing "One-Reason" Decision Making*, 29 J. Experimental Psychol.: Learning, Memory, & Cognition 53 (2003) (reporting results that challenge the validity and pervasiveness of the Take The Best model).

94. Cf. J.E. Russo & B.A. Doshier, *Strategies for Multiattribute Binary Choice*, 9 J. Experimental Psychol.: Learning, Memory & Cognition 676, 683 (1983) (discussing "dimensional reduction" in which less important attributes relevant to a decision task are ignored).

judges take advantage of the ecology⁹⁵ of the multifactor test and, in particular, the redundancy⁹⁶ of many of its factors to reach a conclusion efficiently. The heuristic they employ is evidence, I suggest, of human ingenuity rather than human fallibility.

1. *The Multifactor Test Outcome Classification Tree*

The circuits' various multifactor tests studied in this paper use from six to ten factors—the Federal Circuit's majestic thirteen factor test is not considered.⁹⁷ The average number of factors across the twelve circuits studied is 7.5, but the data suggest that knowledge of far fewer factor outcomes is sufficient to predict the overall outcome of any circuit's particular multifactor test. Figures 1 and 2 set forth simple classification trees of the multifactor test outcome in the 192 preliminary injunction and bench trial opinions sampled.⁹⁸ Figure 1 shows that a one-factor classification scheme based on the outcome of the similarity factor alone (if the similarity factor favors a likelihood of confusion, predict that plaintiff wins; if it does not, predict that defendant wins) would yield an accurate classification of 173 or 90% of the outcomes of the 192 preliminary injunction and bench trial opinions sampled. In Figure 2, the addition of a second node, going to the proximity factor, would increase the accuracy of the classification scheme to 96%, an accuracy rate far above what any regression analysis of the test outcome on all the core factor outcomes proved capable of yielding.⁹⁹

95. See Gigerenzer & Todd, *supra* note 25, at 5 (describing “ecological rationality” as “rationality that is defined by its fit with reality”); *id.* at 13 (“A heuristic is ecologically rational to the degree that it is adapted to the structure of an environment. . . . Thus, simple heuristics and environmental structure can both work hand in hand to provide a realistic alternative to the ideal of optimization, whether unbounded or constrained.”).

96. See *Entrepreneur Media v. Smith*, 279 F.3d 1135, 1141 (9th Cir. 2002) (“Since each factor represents only a facet of the single dispositive issue of likely confusion, the factors, not surprisingly, tend to overlap and interact, and the resolution of one factor will likely influence the outcome and relative importance of other factors. . . . [T]he determination of one factor is often, in essence, only another way of viewing the same considerations already taken into account in finding the presence or absence of another one.”). See also Gigerenzer & Goldstein, *supra* note 25, at 654-55 (discussing pairwise correlation as a means of measuring cue-redundancy); *id.* at 665 (concluding that “[h]igh cue redundancy . . . does seem sufficient but is not necessary for the successful performance of the satisficing algorithms”).

97. None of the district court opinions sampled used the Federal Circuit's *Du Pont* factors.

98. For much of the analysis in this Subpart, I grouped together the 146 preliminary injunction opinions with the forty-six bench trial opinions in order to achieve a sufficiently large sample size to yield meaningful results. Quite obviously, preliminary injunction and bench trial opinions arise out of very different stages of the litigation process. However, as is reflected in the remarkably similar win rates of the two groups of opinions, courts' application of the multifactor test does not appear appreciably to differ between the two groups. A closer comparison of the data for the two groups would confirm this.

99. See Mandeep K. Dhali & Clare Harries, *Fast and Frugal Versus Regression Models of Human Judgement*, 7 *Thinking & Reasoning* 5, 6 (2001) (“Fast and frugal models are easier to understand, and are psychologically more plausible than regression models because they are more

Though it is hard to believe that a federal district court judge would employ, even under exceptional time pressure, anything like a “take the best, ignore the rest”¹⁰⁰ heuristic, both figures offer tantalizing evidence of such a heuristic in action.

Figure 3 sets forth a more elaborate five-node classification tree of the outcomes of the 192 preliminary injunction and bench trial opinions sampled. Here, the five nodes are ordered at each stage according to the accuracy of their classification of the remaining opinions, and if their accuracy is equal, according to the number of remaining opinion outcomes they accurately classify.¹⁰¹ This node order produces, for this data, the most accurate classification scheme (97%) among all possible node orders.

The distribution of the classification tree in Figure 3 is quite stunning. Of the seventy-one opinions in which the similarity factor did not favor a likelihood of confusion, the defendant won the multifactor test in each of these opinions. Of the 121 remaining opinions, the plaintiff won the multifactor test in each of the sixty-five opinions in which the court found (in addition to similarity) that the defendant intended to confuse consumers as to source. Of the remaining fifty-six opinions, twenty-one found evidence of actual confusion, and in each of these, plaintiff won the multifactor test. Thus, by the third node, we have classified 82% of the 192 opinion outcomes with none misclassified. Of the thirty-five remaining opinions, in thirteen the court found that the proximity of the goods factor disfavored a likelihood of confusion, and in twelve of these, the defendant won the multifactor test. This leaves twenty-two opinions. In twelve of these, the court found that the strength factor favored a likelihood of confusion, and in eleven of these twelve, the court found in favor of the plaintiff. Of the ten opinions in which the court found that the strength factor did not favor a likelihood of confusion, the court found in favor of the defendant in six of them. Thus, the application of five nodes accurately classifies 186 of 192 opinion outcomes for an overall accuracy of 97%. The average number of nodes (factor outcomes) required to classify an opinion outcome under this scheme was 2.22.

A word of caution: though the results of the classification trees set forth in Figures 1, 2, and 3 are compelling, there is little in the opinions sampled to suggest that this rudimentary outcome analysis accurately

compatible with humans' cognitive limitations and flexible use of information.”). See also Gigerenzer & Goldstein, *supra* note 25.

100. Gigerenzer & Goldstein, *supra* note 25, at 653.

101. Cf. Laura Martignon & Ulrich Hoffrage, *Fast, Frugal, and Fit: Simple Heuristics For Paired Comparison*, 52 *Theory & Decision* 29, 36-37 (2002) (“Conditional validity” is “computed for each cue just on the set of pairs not discriminated by the cues already used. The first cue used by this type of search is the most valid. . . . The second cue is the most valid on the set of pairs that the first cue did not discriminate, and so on. . . .”)

FIGURE 1
A ONE-NODE CLASSIFICATION TREE OF THE OUTCOME OF THE MULTIFACTOR TEST FOR THE LIKELIHOOD OF CONSUMER CONFUSION IN 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
(PW = plaintiff multifactor test win rate; SS = stamped score; SD = standard deviation of SS)

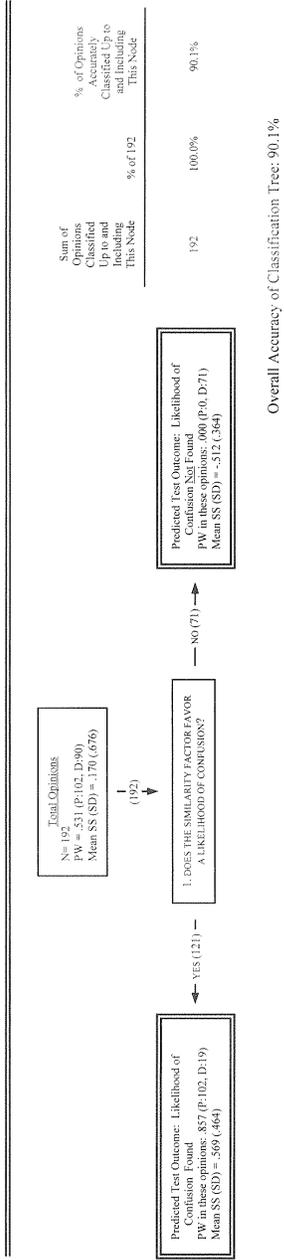


FIGURE 2
A TWO-NODE CLASSIFICATION TREE OF THE OUTCOME OF THE MULTIFACTOR TEST FOR THE LIKELIHOOD OF CONSUMER CONFUSION IN 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
(PW = plaintiff multifactor test win rate; SS = stamped score; SD = standard deviation of SS)

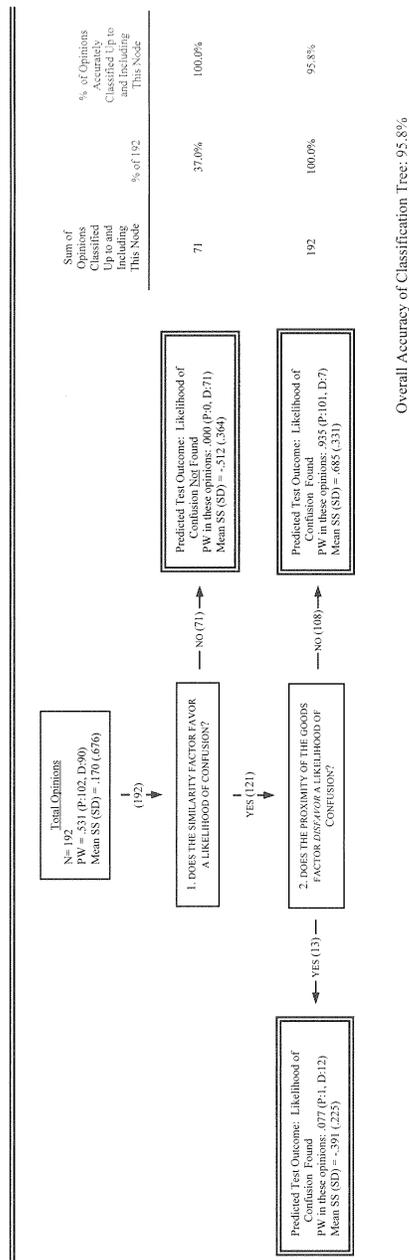
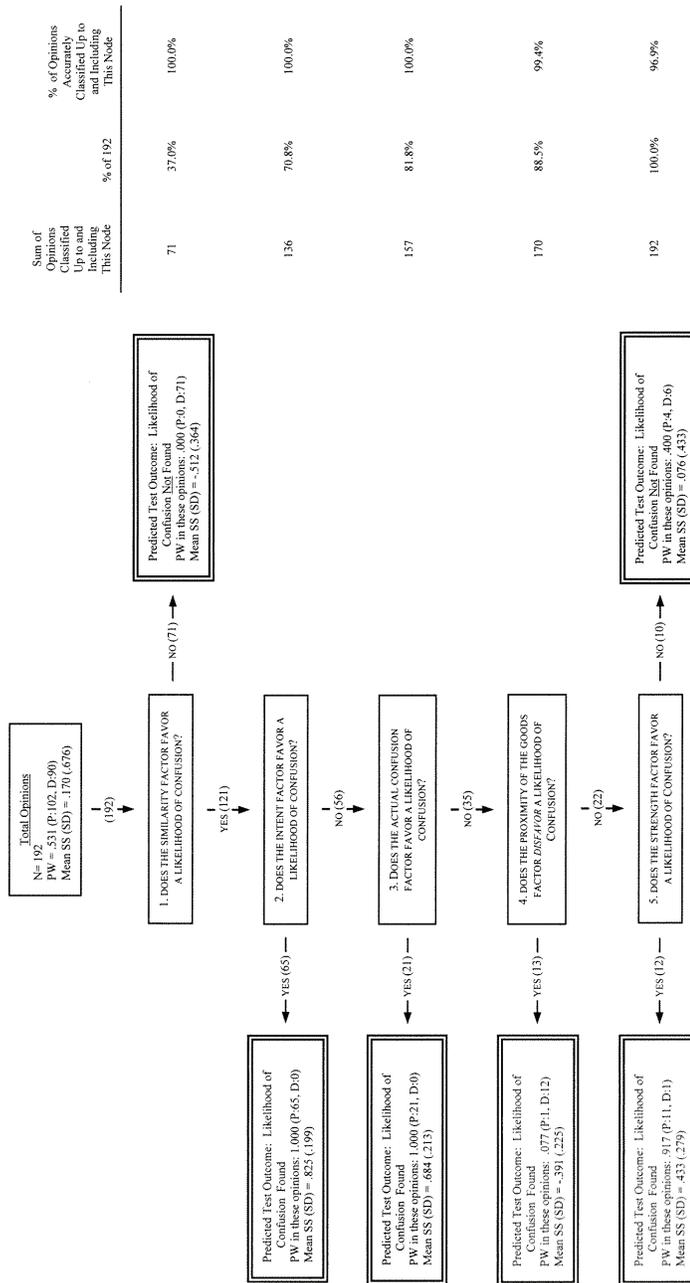


FIGURE 3
A FIVE-NODE CLASSIFICATION TREE OF THE OUTCOME OF THE MULTIFACTOR TEST FOR THE LIKELIHOOD OF CONSUMER CONFUSION IN 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
(PW = plaintiff multifactor test win rate; SS = stamped score; SD = standard deviation of SS)



reflects the *process* of reasoning the judges actually employed. That is, there is little to suggest that the judges conducted a separate, independent analysis of each successive factor in the order and with the stopping points represented. On the contrary, there is much to suggest that the judges engaged in what Professor Dan Simon has called coherence-based reasoning,¹⁰² that certain factor outcomes affected the outcomes of certain other factors, and that the test outcome may have influenced various factor outcomes as much as various factor outcomes influenced the test outcome. Of this, I will have more to say below.

Nevertheless, the distributions of the classification trees suggest that at least some aspects of the multifactor analysis of the likelihood of confusion are noncompensatory in nature. Put differently, certain factor outcomes are weighted so strongly as to outweigh the combined weights of all other factor outcomes.¹⁰³ For example, a finding that the similarity of the marks factor does not favor a likelihood of confusion is sufficient to trigger an overall finding of no likelihood of confusion, regardless, it appears, of the outcomes of any other factors. Similarly, two findings—that the similarity of the marks factor favors a likelihood of confusion and the defendant’s intent factor also favors a likelihood of confusion—are together sufficient to trigger an overall finding of a likelihood of confusion, again, regardless of the outcomes of any other factors.¹⁰⁴

2. Test Outcome by Factor Outcome

The multifactor test outcome classification trees do not do justice to the association between the test outcome and certain individual factors listed lower—or not at all—in Figure 3. To see this association, we must

102. See Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. Chi. L. Rev. 511 (2004). See also *infra* notes 102-126 and accompanying text.

103. Compensatory decision making weighs the positive attributes of a decision alternative against its negative attributes. In the process, a “good value on one attribute can compensate for a poor value on another. A compensatory strategy thus requires explicit trade-offs among attributes.” James R. Bettman et al., *Constructive Consumer Choice Processes*, 25 J. Consumer Res. 187, 190 (1998). Noncompensatory decision making does not engage in the weighing of attributes and may instead decide based on only one attribute or some limited combination of attributes. “[A]lternatives may be eliminated (or chosen) based on the value of one attribute without considering the values of other potentially compensating attributes.” Barbara E. Kahn & Jonathan Baron, *An Exploratory Study of Choice Rules Favored for High-Stakes Decisions*, 4 J. Consumer Psychol. 305, 307 (1995). An example of a noncompensatory decision-making strategy is the lexicographic rule (consider the process of ordering words in a dictionary), in which the decision maker compares decision alternatives according to attributes ordered by importance, and decides based on the first attribute that discriminates between the alternatives. See Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. Chi. L. Rev. 1203, 1223-32 (2003) (discussing non-compensatory decision making). See generally Payne et al., *supra* note 51, at 26-30.

104. In this connection, it is interesting to note that there was one opinion in the initial sample of 337 opinions in which the plaintiff won both the similarity and intent factors but lost the test. See *Sally Beauty Co. v. Beautyco, Inc.*, No. CIV-99-1372-C, 2001 U.S. Dist. LEXIS 25294 (W.D. Okla. 2001), *rev’d*, *Sally Beauty Co., Inc. v. Beautyco, Inc.*, 304 F.3d 964 (10th Cir. 2002).

take a closer look at the data. For the 192 preliminary injunction and bench trial opinions sampled, Table 3 sets out by factor the number and proportion of opinions that held that the factor favored or disfavored a likelihood of confusion. The table also sets out the plaintiff multifactor test win rates in those opinions. Thus, taking the similarity factor as an example, 121 opinions (63%) found that the similarity factor favored a likelihood of confusion, and plaintiff won the multifactor test in 84% of these opinions; sixty-five opinions (34%) found that the similarity factor disfavored confusion, and plaintiff won the multifactor test in none of these opinions.

Compare these plaintiff win rates under the similarity factor to plaintiff win rates under the actual confusion and intent factors. The court found an intent to confuse consumers in sixty-seven opinions. In sixty-five (97%) of these opinions, the court found an overall likelihood of confusion.¹⁰⁵ A finding of bad faith appears to exert a substantial, if not dispositive, influence on the outcome of the test. With respect to the actual confusion factor, the court found that the factor favored a likelihood of confusion in sixty-six opinions. In sixty-one (92%) of these opinions, the court found an overall likelihood of confusion. Finally, compare the consumer sophistication factor win rates to the similarity factor win rates. The plaintiff win rate in the fifty-five cases in which it won the consumer sophistication factor is comparable to the 121 cases in which it won the similarity factor.

We find dramatically skewed results as well in opinions in which the plaintiff lost certain factors. In forty-one of the 192 opinions, the court found that the parties' goods were not proximate. The plaintiff lost the multifactor test in all but one of these opinions.¹⁰⁶ As a practical matter, in order to win the multifactor test, the plaintiff must not lose this factor—or alternatively, when the judge finds an overall likelihood of confusion, the judge almost invariably finds that the proximity factor favors this result. The same may be said of the strength factor. The plaintiff lost this factor in fifty-three of the 192 opinions and lost the overall test in fifty of these fifty-three opinions.

105. For a discussion of the two outlying opinions, see *infra* notes 117-126 and accompanying text.

106. In the one outlying case, *Welch Allyn Inc. v. Tyco Int'l Servs. AG*, 200 F. Supp. 2d 130 (N.D.N.Y. 2002), the plaintiff sold stethoscopes and sphygmomanometers under the mark TYCOS while the defendant sold medical "disposable items or small items purchased in bulk," *id.* at 140-41, under the mark TYCO/HEALTHCARE or TYCO with a corporate tag line "A Tyco International Ltd. Company." *Id.* at 139-41. Though the court found that "[p]laintiff has not presented sufficient evidence to suggest that this [proximity] factor weighs in its favor." *Id.* at 141. The court found a likelihood of confusion and granted a preliminary injunction against the defendant with respect to the sale of any "non-disposable medical products or medical instruments . . ." *Id.* at 151.

TABLE 3
 PLAINTIFF WIN RATE AND STAMPEDE SCORES BY FACTOR OUTCOME IN
 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
 (LOC = likelihood of confusion; PW = plaintiff multifactor test win rate; SS = stampede score;
 SD = standard deviation of SS; \hat{s} = skewness; \hat{k} = kurtosis)

Factor	Outcome	Distribution by Outcome		PW	Stampede Score			
		N	%		SS	SD	\hat{s}	\hat{k}
1. Similarity of the marks	Favors LOC	121	63.0	.843	.569	.464	-1.159	3.423
	Disfavors LOC	65	33.9	.000	-.502	.365	0.393	2.375
	Other	6	03.1	.000	-.616	.371	0.523	2.366
2. Proximity of the goods	Favors LOC	134	69.8	.716	.470	.508	-0.725	2.255
	Disfavors LOC	41	21.4	.024	-.627	.290	0.474	2.317
	Other	17	8.9	.294	-.281	.713	0.627	2.022
3. Evidence of actual confusion	Favors LOC	66	34.4	.924	.765	.340	-2.240	8.550
	Disfavors LOC	88	45.8	.193	-.351	.499	0.571	2.378
	Other	38	19.8	.632	.341	.515	-0.400	1.912
4. Strength of plaintiff's mark	Favors LOC	116	60.4	.793	.525	.507	-1.078	2.996
	Disfavors LOC	53	27.6	.057	-.517	.356	0.402	2.373
	Other	23	12.5	.304	-.076	.699	0.114	1.858
5. Defendant's Intent	Favors LOC	67	34.9	.970	.798	.255	-1.793	7.564
	Disfavors LOC	99	51.6	.253	-.270	.516	0.482	2.161
	Other	26	13.5	.462	.222	.673	-0.464	2.071
6. Sophistication of the Consumers	Favors LOC	55	31.4	.855	.682	.374	-1.240	3.817
	Disfavors LOC	70	40.0	.186	-.386	.492	0.815	2.872
	Other	50	28.6	.640	.306	.613	-0.664	2.218
7. Similarity of advertising methods	Favors LOC	50	56.2	.800	.640	.442	-1.075	3.036
	Disfavors LOC	23	25.6	.217	-.360	.516	0.684	2.499
	Other	16	18.0	.562	.209	.740	-0.482	1.608
8. Similarity of sales facilities	Favors LOC	30	69.8	.700	.512	.468	-0.651	2.171
	Disfavors LOC	5	11.6	.000	-.569	.248	0.429	2.153
	Other	8	17.1	.750	.518	.289	-1.175	2.540
9. Likelihood of bridging the gap	Favors LOC	37	27.8	.649	.427	.505	-0.578	1.914
	Disfavors LOC	26	19.5	.231	-.535	.445	0.836	2.489
	Other	70	53.0	.514	.155	.671	-0.201	1.598
10. Comparative Quality of the Parties' Goods	Favors LOC	13	19.7	.692	.495	.524	-0.774	2.242
	Disfavors LOC	19	28.8	.158	-.497	.502	1.543	4.686
	Other	34	51.5	.441	-.008	.570	-0.112	1.773
11. Length of time of concurrent use	Favors LOC	4	22.2	1.000	.917	.167	-1.155	2.333
	Disfavors LOC	7	38.9	.143	-.536	.367	-0.167	1.692
	Other	7	38.9	.571	.299	.416	-0.568	1.715
12. Similarity in targets of sales efforts	Favors LOC	15	83.3	.600	.307	.545	-0.128	1.531
	Disfavors LOC	1	5.6	.000	-.600	--	--	--
	Other	2	11.1	.000	-1.000	.000	--	--
Total opinions:		192		.531	.170	.676	-0.259	1.633
LOC found:		102	53.1		.723	.291	-1.235	4.632
LOC not found:		90	46.9		-.458	.369	.293	2.326

3. *Factor Outcome by Test Outcome*

While Table 3 tabulates the overall multifactor test outcome by the outcome of each individual factor, Table 4 does the inverse. It tabulates the outcomes of the individual factors by outcome of the overall multifactor test. In doing so, it provides some insight, however limited, on the degree to which test outcomes may drive specific factor outcomes.

Consider, for example, the 102 opinions that found a likelihood of confusion. Of these, only sixty-five (64%) found that the intent factor favored a likelihood of confusion, which suggests that, while probably nearly sufficient, a finding of bad faith intent is by no means necessary to trigger an overall finding of a likelihood of confusion. By comparison, 94% of the 102 opinions that found a likelihood of confusion found that the proximity of the goods factor favored this result and 90% found that the strength factor favored the result—and, of course, all of them found that the similarity factor favored the result. From this, we can infer, albeit weakly, that judges tend to rely on these three factor outcomes to form the foundation for a finding of a likelihood of confusion. Indeed, ninety (88%) of the 102 opinions that found a likelihood of confusion found that each of these factors favored that result.

Consider next the ninety opinions that found no likelihood of confusion. Nineteen of these (21%) nevertheless found that the similarity of the marks disfavored that result. This confirms that a finding that the similarity factor favors a likelihood of confusion is necessary but not sufficient to trigger an overall finding of a likelihood of confusion. Similarly, thirty-eight (42%) found that the proximity of the goods factor favored a likelihood of confusion. Thus, while the plaintiff must not lose the proximity factor in order to win the multifactor test, winning the factor does not guarantee success.

4. *Correlations Between the Factor and Test Outcomes*

Table 5 shows, with respect to the 192 preliminary injunction and bench trial opinions sampled, the pairwise correlation coefficients¹⁰⁷ between, on the one hand, each of the two outcomes of the multifactor test and on the other, each of the two most common outcomes (favors or

107. Generally speaking, correlation analysis produces a correlation coefficient from -1, which, if the coefficient is statistically significant, demonstrates a perfect inverse relation between two variables, to +1, which, again if the coefficient is statistically significant, demonstrates a perfect positive relation between two variables. See generally Cohen et al., *supra* note 85, at 19-32. Pairwise correlation analysis analyzes the correlation between two variables in all observations in which they coexist. Observations in which either one or both of the variables are not present are excluded from the analysis. Pairwise correlation analysis is appropriate here because certain circuits do not consider certain factors.

TABLE 4
 FACTOR OUTCOME AND STAMPEDE SCORE BY TEST OUTCOME IN 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
 (SS = stampede score; SD = standard deviation of SS)

Factor	Outcome	Test Outcome							
		Opinions in Which LOC Found				Opinions in Which LOC Not Found			
		N	%	SS	SD	N	%	SS	SD
1. Similarity of the marks	Favors LOC	102	100.0	.723	.291	19	21.1	-.257	.318
	Disfavors LOC	0	0.0	---	---	65	72.2	-.502	.365
	Other	0	0.0	---	---	6	6.7	-.616	.371
2. Proximity of the goods	Favors LOC	96	94.1	.737	.270	38	42.2	-.204	.299
	Disfavors LOC	1	1.0	-.375	---	40	44.4	-.633	.290
	Other	5	4.9	.677	.316	12	13.3	-.680	.335
3. Evidence of actual confusion	Favors LOC	61	59.8	.841	.194	5	5.6	-.158	.396
	Disfavors LOC	17	16.7	.397	.281	71	78.9	-.531	.351
	Other	24	23.5	.656	.314	14	15.6	-.198	.288
4. Strength of plaintiff's mark	Favors LOC	92	90.2	.745	.259	24	26.7	-.317	.294
	Disfavors LOC	3	2.9	-.037	.157	50	55.6	-.546	.344
	Other	7	6.9	.767	.326	17	18.9	-.423	.473
5. Defendant's Intent	Favors LOC	65	63.7	.826	.199	2	2.2	-.095	.337
	Disfavors LOC	25	24.5	.422	.302	74	82.2	-.503	.330
	Other	12	11.8	.799	.273	14	15.6	-.272	.489
6. Sophistication of the Consumers	Favors LOC	47	51.1	.801	.230	8	9.6	-.022	.267
	Disfavors LOC	13	14.1	.396	.368	57	68.7	-.564	.309
	Other	32	34.8	.702	.252	18	21.7	-.398	.383
7. Similarity of advertising methods	Favors LOC	40	74.1	.813	.268	9	56.3	-.036	.289
	Disfavors LOC	5	9.3	.430	.252	5	31.3	-.569	.248
	Other	9	16.7	.781	.184	2	15.4	-.786	.303
8. Similarity of sales facilities	Favors LOC	21	77.8	.747	.302	10	28.6	-.052	.294
	Disfavors LOC	0	0.0	---	---	18	50.0	-.580	.308
	Other	6	22.2	.952	.117	7	20.0	-.526	.450
9. Likelihood of bridging the gap	Favors LOC	24	36.4	.757	.200	13	19.4	-.181	.269
	Disfavors LOC	6	9.1	.149	.267	20	29.9	-.740	.224
	Other	36	54.5	.729	.274	34	50.7	-.452	.353
10. Comparative Quality of the Parties' Goods	Favors LOC	9	33.3	.796	.213	4	10.3	-.183	.304
	Disfavors LOC	3	11.1	.375	.650	16	41.0	-.660	.257
	Other	15	55.6	.530	.233	19	48.7	-.432	.350
11. Length of time of concurrent use	Favors LOC	4	44.4	.916	.167	0	0.0		
	Disfavors LOC	1	11.1	-.111	---	6	66.7	-.607	.346
	Other	4	44.4	.617	.033	3	33.3	-.125	.217
12. Similarity in targets of sales efforts	Favors LOC	9	100.0	.669	.345	6	66.7	-.236	.225
	Disfavors LOC	0	0.0	---	---	1	11.1	-.600	---
	Other	0	0.0	---	---	2	22.2	-1.000	.000
LOC found		102		.723	.291				
LOC not found						90		-.458	.369

disfavors a likelihood of confusion) of the nine factors most commonly used among the circuits. It is prudent not to place too much weight on these results, first, because a correlation coefficient can underreport the strength of correlations between dichotomous variables,¹⁰⁸ and second, because the correlations reported are weakened by their failure to take into account outcomes other than favors or disfavors a likelihood of confusion, such as findings that a factor was neutral, irrelevant, or not argued.¹⁰⁹ Even so, the coefficients given in the first two columns of the table are generally consistent with the inferences made above. The outcomes of the similarity factor enjoy the strongest correlation with the overall test outcomes. Additionally, the outcomes of the other four core factors also each correlate fairly strongly with the larger test outcomes, with strength and intent correlating slightly more strongly than actual confusion and proximity.¹¹⁰

Outcomes under the remaining four factors listed in the table—purchaser sophistication, similarity of advertising/marketing, similarity of sales facilities, and likelihood of bridging the gap—show weak correlations with the test outcomes, with one exception. A finding that the consumer sophistication factor disfavors a likelihood of confusion correlates fairly strongly with an overall finding of no likelihood of confusion. This is consistent with the plaintiff's relatively low multifactor test win rate (19%) in opinions in which the court made this specific finding. As for the three other factors used by the circuits but not listed in Table 5, both the comparative quality of the marks factor and the similarity of the targets of the parties' sales efforts factor showed no strong correlation with the outcome of the test.¹¹¹ Finally, as expected, the factor analyzing the length of time of

108. See, e.g., Cohen et al., *supra* note 85, at 53-55. A dichotomous variable is a binary variable typically coded either as one or zero.

109. For this correlation analysis, each factor outcome is represented with two binary variables: favors a likelihood of confusion (1=yes, 0=no) and disfavors a likelihood of confusion (1=yes, 0=no). Thus, if the first variable is coded as one, then the second variable will be coded as zero, and vice-versa. But if the court found the factor to be neutral, irrelevant, or not argued, then both variables were coded as zero. This explains why, in Table 5, the sum of the absolute values of the correlation coefficients for any particular factor variable as against another factor (or test) outcome do not equal one.

110. The correlation coefficient of the similarity factor with the test outcome is relatively high. Its confidence interval at the .05 level does not overlap with the confidence intervals of any other factors' correlation coefficients.

111. For preliminary injunction and bench trial opinions, the correlation between an overall finding of a likelihood of confusion and a finding that the comparative quality factor favored that result was .289 ($p=.020$, $n=66$); the correlation for preliminary injunction and bench trial opinions between an overall finding of no likelihood of confusion and a finding that the comparative quality factor favored that result was .324 ($p=.008$, $n=66$). The correlations between the outcomes of the similarity of the targets of the parties' sales efforts and the outcomes of the multifactor test were not significant at the .05 level for preliminary injunction and bench trial opinions.

TABLE 5
PAIRWISE CORRELATION COEFFICIENTS AMONG TEST AND FACTOR OUTCOMES IN 192 PRELIMINARY INJUNCTION AND BENCH TRIAL OPINIONS
(Number in parentheses indicates number of opinions correlated; "LOC" is "Likelihood of Confusion";
Correlation coefficients greater than .500 are rendered in boldface type)

	LOC Found		LOC Not Found		Similarity of the Marks		Proximity of the Goods		Actual Confusion		Strength of the Plaintiff's Mark		Defendant's Intent		Purchaser Sophistication		Similarity of Advertising		Similarity of Sales Facilities		Likelihood of Bridging the Gap	
	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC	Supports LOC	Does Not Support LOC
LOC Found	1																					
LOC Not Found		-1.000** (192)																				
Similarity of the Marks		.815** (192)																				
Does Not Support LOC			1																			
Supports LOC				-0.934** (192)																		
Proximity of the Goods																						
Does Not Support LOC																						
Supports LOC																						
Actual Confusion																						
Does Not Support LOC																						
Supports LOC																						
Strength of the Plaintiff's Mark																						
Does Not Support LOC																						
Supports LOC																						
Defendant's Intent																						
Does Not Support LOC																						
Supports LOC																						
Purchaser Sophistication																						
Does Not Support LOC																						
Supports LOC																						
Similarity of Advertising/Marketing																						
Does Not Support LOC																						
Supports LOC																						
Similarity of Sales Facilities																						
Does Not Support LOC																						
Supports LOC																						
Likelihood of Bridging the Gap																						
Does Not Support LOC																						
Supports LOC																						

* Correlation is significant at the .05 level of confidence. **Correlation is significant at the .01 level of confidence.

concurrent use without evidence of actual confusion essentially reiterated the correlations shown under the actual confusion factor.¹¹²

5. Summary

In general, the data suggest, first, that when district judges use the multifactor test, very few factor outcomes, often merely two or three, are sufficient to trigger one of the two test outcomes, but second, that both of these test outcomes require or nearly require certain factor outcomes. Can we conclude, then, that judges tend to short-circuit the multifactor test? The answer is yes. In theory, the multifactor test is a full-fledged balancing test. In practice, it is a complex of per se rules. But in relying only on certain leading factors, are judges making flawed decisions? If recent research in human decision making is any guide, then the answer is very likely no. Like any human decision makers, district judges attempt to decide both efficiently and accurately.¹¹³ In pursuit of efficiency, they consider only a few factors. In pursuit of accuracy, they consider the most decisive factors. In essence, as consummate pragmatists, they “take the best,” a strategy which empirical work suggests is an altogether successful—and rational—approach to decision making.¹¹⁴

The crucial problem, however, is that in explaining the reasoning that led to their decision, district judges are not permitted to “ignore the rest.” Each factor must be addressed, not simply those which actually formed the basis for their ruling. This creates the conditions for the stampeding of factor outcomes, a phenomenon to which I now turn.

B. Stampeding

The correlation coefficients reported in Table 5 among the factor outcomes themselves, rather than with the test outcomes, show that all of the core factor outcomes bear a statistically significant association with each other. Additionally, these coefficients reveal that many of the non-core factor outcomes also produce a statistically significant association with the core factor outcomes, if not also to some extent with each other’s outcomes. This hints at an interesting phenomenon: judges tended to stampede the factor outcomes to favor the test outcome, especially when they found a

112. For preliminary injunction and bench trial opinions, the correlation between an overall finding of a likelihood of confusion and a finding that the length of time of concurrent use factor favored that result was .535 ($p=.022$, $n=18$); the correlation between an overall finding of no likelihood of confusion and a finding that the length of time of concurrent use factor favored that result was .570 ($p=.014$, $n=18$).

113. See, e.g., Jens Förster, E. Tory Higgins, & Amy Taylor Bianco, *Speed/Accuracy Decisions in Task Performance: Built-In Tradeoff or Separate Strategic Concerns?*, 90 *Organizational Behav. & Human Decision Processes* 148 (2003) (discussing speed/accuracy decisions from the perspective of regulatory focus theory).

114. See *supra* notes 25 and 93 and accompanying text.

likelihood of confusion. What emerges is a picture of legal multifactor decision making in which certain factors drive the outcome and the rest of the factors subsequently fall in line to support that outcome. The data are all the more compelling in light of the fact that the sampled opinions arise out of cases that we would generally assume to be close cases, that is, cases that failed to settle before the issuance of the opinion.¹¹⁵ Furthermore, the data set excluded counterfeiting and licensing opinions, which we would expect to stampede.¹¹⁶ Before turning to the data, however, I need to address more generally the practice of coherence-based reasoning.

1. *Coherence-Based Reasoning and Coherence Shifts*

In a remarkable series of law review and experimental psychology journal articles, researchers have developed and tested a model of what they have termed coherence-based reasoning.¹¹⁷ This model hypothesizes that the “decision-making process progresses bi-directionally: premises and facts both determine conclusions and are affected by them in return. A natural result of this cognitive process is a skewing of the premises and facts toward inflated support for the chosen decision.”¹¹⁸ As Professor Simon explains, the decision-making process begins with a “mental model of the decision task,” which:

contains a myriad of variables that point in more than one direction and thus do not all fit into a coherent mental model. One subset of variables [a₁, a₂, . . . a₃] supports conclusion A, and the other subset [b₁, b₂, . . . b₃] supports the opposite conclusion B. . . . In all but easy cases, neither subset of variables dominates the other. Since each variable has some bearing on the task, it can be said to impose a *constraint* on the network. . . . Each and every constraint influences, and is influenced by, the entire network, so that every processing cycle results in a slightly modified mental model.¹¹⁹

The decision maker cycles through the mental model in an effort to satisfy, and in the process conform, the model’s myriad constraints until “the

115. See generally Priest & Klein, *supra* note 76. I thank Mark Lemley for proposing, though not necessarily endorsing, this point.

116. See Appendix A.

117. Dan Simon, Chadwick J. Snow & Stephen J. Read, *The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction*, 86 *J. Personality & Soc. Psychol.* 814 (2004); Dan Simon, Daniel C. Krawczyk & Keith J. Holyoak, *Construction of Preferences by Constraint Satisfaction*, 15 *Psychol. Sci.* 331 (2004); Simon, *supra* note 102; Dan Simon, Lien B. Pham, Quang A. Le, & Keith J. Holyoak, *The Emergence of Coherence Over the Course of Decision Making*, 27 *J. Experimental. Psychol.: Learning, Memory, & Cognition* 1250 (2001); Keith J. Holyoak & Dan Simon, *Bidirectional Reasoning in Decision Making by Constraint Satisfaction*, 128 *J. Experimental. Psychol.: Gen.* 3 (1999); Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 *Rutgers L.J.* 1 (1998).

118. Simon, *supra* note 102, at 511.

119. *Id.* at 521 (internal citations omitted).

constraints settle at a point of maximal coherence.”¹²⁰ By means of a kind of “reversed induction,” “[c]onstraint satisfaction processes force the task variables to change toward a better fit with the gradually emerging state of coherence.”¹²¹ In time, the mental model shifts toward this state of coherence, in which the considerations favoring an alternative are seen strongly to do so, and those disfavoring [that] alternative are seen to do so weakly, if at all.¹²² “In sum,” Professor Simon writes,

the ultimate state of coherence is essentially a byproduct of the cognitive system’s drifting toward either one of two skewed mental models. Within each of these models, the initially complex and incoherent mental model has been spread into two subsets, one of which dominates the other, thereby enabling a relatively easy and confident choice. This skewed representation reflects an artificial polarization between the inflated representation of the variables that support the chosen conclusion and the deflated ones that support the rejected conclusion; it differs considerably from the way the task variables were perceived before the decision-making process got underway, and it differs also from the way they will be perceived some time after the completion of the task.¹²³

Two aspects of the coherence-based reasoning model must be emphasized from the outset. First, while the model shares in the general spirit of the biases and heuristics research that underlies behavioral law and economics, it addresses different cognitive phenomena. It is more concerned with the underlying cognitive processes that drive complex decision making than with the biases and heuristics that inform various discrete judgments and decisions. Second, coherence-based reasoning theory is distinct from cognitive dissonance theory.¹²⁴ The former does not conceive, as the latter does, of coherence shifts as simply a matter of post hoc rationalization, as something triggered in an effort to reduce post-decision regret. On the contrary, empirical work suggests that coherence shifts precede the decision and form the basis for it. In fact, they can occur early on in the decision-making process and can even be triggered by a single attribute.¹²⁵ “Much as forcing one interpretation on a vertex of a Necker cube can cause

120. *Id.* at 522.

121. *Id.*

122. *Id.* at 516. (“A mental model of a decision task is deemed ‘coherent’ when the decision-maker perceives the chosen alternative to be supported by strong considerations while the considerations that support the rejected alternative are weak. . . . A mental model is considered ‘incoherent’ when the decision-maker perceives the considerations as providing equivocal support for both alternatives. As defined, coherence is an empirical phenomenon, not a jurisprudential ideal.” (internal citations omitted)).

123. Simon, *supra* note 102, at 522-23.

124. On cognitive dissonance theory, *see generally* Leon Festinger, Conflict, Decision, and Dissonance (1964). For Simon’s response to cognitive dissonance theory, *see, for example*, Simon, *supra* note 102, at 534-535; Simon et al., *Construction of Preferences*, *supra* note 117, at 333-34.

125. *See* Simon et al., *Construction of Preferences*, *supra* note 117.

the entire set of vertices to be perceived as a particular three-dimensional form, biasing the assessment of one particular point of dispute should initiate spreading coherence, causing systematic changes in the entire set of assessments related . . .”¹²⁶ to the decision task.

2. *Stampeding by Disposition and Posture*

The data collected for this study support the coherence-based reasoning model, and specifically, the hypothesis that coherence shifts trigger an artificial polarization between the set of considerations favoring a decision and the set disfavoring a decision. Table 6 reports the mean *multifactor stampede scores* for the 331 opinions sampled by outcome and posture. An opinion’s multifactor stampede score is the difference between the proportion of factors considered that favored a finding of a likelihood of confusion and the proportion of factors considered that did not favor a finding of a likelihood of confusion. Thus, in an opinion where all factors in the multifactor test were held to favor a likelihood of confusion, the opinion yielded a stampede score of 1.000, and in an opinion where all factors were held not to favor a likelihood of confusion, the opinion yielded a stampede score of -1.000. In opinions where the factors were tied (for example, four in favor of a likelihood of confusion and four against), the stampede score was 0.000. Figure 4 sets forth the distribution of stampede scores by disposition for the 287 dispositive opinions sampled.

The remarkably high mean stampede scores for opinions that found a likelihood of confusion provide strong evidence of polarization, as does the strong skew in the distribution of the stampede scores of such opinions. The twenty-five bench trial opinions that found a likelihood of confusion yielded a mean stampede score of .788,¹²⁷ with nine (36%) yielding a stampede score of 1.000.¹²⁸ Similarly, the seventy-eight preliminary injunction opinions that found a likelihood of confusion yielded a mean stampede score of .706, with twenty-eight (36%) yielding a stampede score of 1.000. Opinions granting plaintiff summary judgment provided similarly high

126. Holyoak & Simon, *supra* note 117, at 12. The Swiss geologist Louis Albert Necker is often given credit for developing the following optical illusion:



See, e.g., Frederick Burwick, *The Grottesque: Illusion vs. Delusion*, in *Aesthetic Illusion: Theoretical and Historical Approaches* 122, 132 (Frederick Burwick & Walter Pape eds., 1990) (discussing Necker’s first encounter with the optical illusion that bears his name).

127. In an eight factor test, this is roughly the equivalent of, among other combinations, seven factors in favor of confusion and one against.

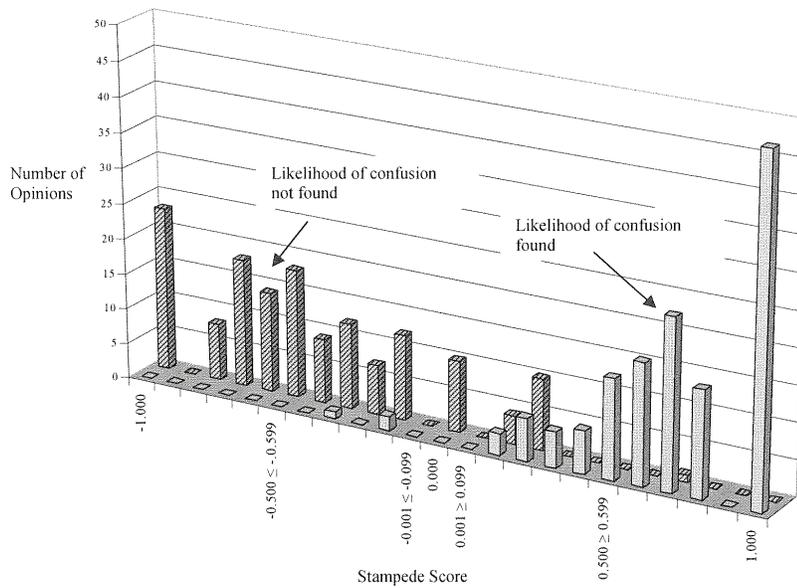
128. Again, I emphasize that counterfeiting, licensing, and similar cases where we would expect such scores were excluded from the sample.

TABLE 6
MULTIFACTOR TEST STAMPEDE SCORES BY DISPOSITION AND POSTURE IN 331 DISTRICT COURT OPINIONS
(SS = stampede score)

Disposition	Posture	N	Mean SS	SD	Min. SS	Median SS	Max. SS	N With [SS] of 1.000	N With Irregular SS*
LOC Found	Preliminary Injunction	78	.706	.302	-.375	.732	1.000	28	2
	Bench Trial	24	.779	.251	-.111	.845	1.000	8	1
	Plaintiff SJ Motion	26	.720	.244	.200	.714	1.000	8	0
	Cross SJ Motions	13	.726	.253	.250	.714	1.000	4	0
LOC Not Found	Preliminary Injunction	68	-.489	.382	-1.000	-.500	.375	11	7
	Bench Trial	22	-.361	.314	-1.000	-.354	.143	1	2
	Defendant SJ Motion	44	-.507	.415	-1.000	-.625	.714	9	4
	Cross SJ Motions	12	-.426	.465	-1.000	-.646	.250	2	2
SJ Denied	Plaintiff SJ Denied	8	.092	.327	-.625	.167	.429	0	0
	Defendant SJ Denied	15	.296	.389	-.625	.286	.833	0	0
	Cross SJs Denied	21	.065	.408	-.800	.000	1.000	1	0
Total		331	.120	.655	-1.000	.167	1.000	72	18

*"Irregular Count" describes opinions in which the winner of the multifactor test won fewer factors than the loser of the test.

FIGURE 4
THE DISTRIBUTION OF MULTIFACTOR TEST STAMPEDE SCORES FROM 287 DISPOSITIVE DISTRICT COURT OPINIONS



scores, even when the defendant submitted its own cross-motion for summary judgment. This suggests that the tendency of courts to stampede the factors when they find an overall likelihood of confusion is robust against the various presumptions and standards of the opinion postures.

When the multifactor test led to a finding of no likelihood of confusion, however, there was little, if any stampeding of the factors. This is apparent in each disposition's relatively modest stampede score and in the low proportion of opinions that yielded a score of -1.000. The relatively flat distribution of the stampede scores of such opinions also supports this characterization. Once again, this characterization is robust against the posture of the opinion.

Why do the factors tend to stampede in one direction but not in the other, regardless of posture? There may be a simple explanation, one that goes to the underlying merits of the cases brought. Generally, a plaintiff will not bring an action for trademark infringement unless the facts of its case are such that it will win at least a few of the multifactor test factors. The overall mean stampede score, regardless of outcome, for the 331 opinions sampled is .120,¹²⁹ whereas the plaintiff multifactor test win rate for these opinions is about even at .471. Similarly, the overall mean stampede score for the 192 preliminary and bench trial opinions sampled is .170,¹³⁰ whereas the plaintiff multifactor test win rate for these opinions is also about even at .531. These results are consistent with an overall lean in the test by one or two factors toward a likelihood of confusion.

Furthermore, a finding of no likelihood of confusion generally represents an endorsement of the status quo, while a finding of a likelihood of confusion generally leads to an injunctive intervention in the status quo. Regardless of the presumptions and standards of the preliminary injunction, summary judgment, or bench trial postures, courts may feel, perhaps as a function of "status quo bias"¹³¹ or "omission bias,"¹³² that the multifactor test must tilt strongly toward a likelihood of confusion to justify such an

129. SD=.655.

130. SD=.676.

131. For more information on "status quo bias," in which individuals tend to prefer the status quo to alternative states, all other things being equal, see Russell Korobkin, *The Status Quo Bias and Contract Default Rules*, 83 Cornell L. Rev. 608, 625-30 (1998) (reviewing status quo bias literature). See generally William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. Risk & Uncertainty 7 (1988) and Raymond S. Hartman et al., *Consumer Rationality and the Status Quo*, 106 Q.J. Econ. 141, 158-60 (1991). On the role of status quo bias in appellate decision making, see Chris Guthrie & Tracey E. George, *The Futility of Appeal: Disciplinary Insights into the "Affirmance Effect" on the United States Courts of Appeals*, 32 Fla. St. U. L. Rev. 357, 377-79 (2005).

132. On omission bias, see, for example, Laura Y. Niedermayer & Gretchen B. Chapman, *Action, Inaction, and Factors Influencing Perceived Decision Making*, 14 J. Behav. Decision Making 295, 296 (2001) ("The omission bias is the tendency to judge actions as worse than omissions when they both have the same bad consequences."). See also Guthrie & George, *supra* note 131, at 379-80; Robert A. Prentice & Jonathan J. Koehler, *A Normality Bias in Legal Decision Making*, 88 Cornell L. Rev. 583 (2003).

intervention, whereas an endorsement of the status quo does not require as strong of a foundation. As their mental models cycle toward a finding for the plaintiff, judges may thus feel the need to bolster their findings under various factors, if only to insulate their more activist judgment from appeal.

Table 6 also shows the highest and lowest stampede scores for each disposition. Note that in eighteen of the 331 cases sampled (5%), the loser of the multifactor test won more factors than the winner of the test.¹³³ For example, Table 6 shows that in two preliminary injunction opinions, the court found a likelihood of confusion even though the defendant won more factors in the multifactor test than the plaintiff, and that in seven preliminary injunction opinions, the court found no likelihood of confusion even though the plaintiff won more factors than the defendant. This confirms courts' frequent observation that the outcome of the test is not necessarily determined by which party wins the most factors.¹³⁴

3. *Stampeding by Factor*

In light of the distribution of the outcome classification tree shown in Figure 3 and the prevalence of certain factor outcomes in opinions that found a likelihood of confusion as shown in Table 4, it is tempting to suggest that the multifactor test is essentially a two-stage test, the first stage of which is *not* a balancing test. Instead, the first stage could be conceived of as consisting of certain requirements, each of which the plaintiff must meet: The marks must be similar, the plaintiff's mark must be strong, and the goods must be proximate. If these requirements are met, then the test proceeds to the second stage, which balances other factors, the most important of which are intent, actual confusion, and perhaps also consumer sophistication.

The problem with this characterization is that it fails properly to account for the strong influence of two core factors, from stage two of the above hypothesized test, on their stage one peers. These two factors are intent and actual confusion, whose outcomes appear to stampede the rest of the test factors. There is thus yet another possible explanation for the divergence by outcome in the degree of stampeding in the opinions sampled. As we saw above, Simon suggests that a single variable can initiate spreading coherence in the sense that featuring one vertex of a Necker cube can shift the viewer's perception of all the other vertices. Intent and, to a lesser degree, actual confusion appear to exert such a coherence-shifting influ-

133. See, e.g., *Butcher Co., Inc. v. Bouthot*, 124 F. Supp. 2d 750, 760 (D. Me. 2001) ("It is true, of course, that in terms of simple numbers there are only three factors weighing against finding a likelihood of confusion but five factors in favor of finding a likelihood of confusion. Nonetheless, the former are determinative in this case.").

134. See *supra* note 46 and accompanying text.

ence when they favor a likelihood of confusion. Indeed, in the forty-nine opinions in which both findings were made, thirty-four (69%) of them found that all the factors favored a likelihood of confusion.

In addition to setting out the test outcomes by factor outcome, Table 3 reports the stampede scores for the preliminary injunction and bench trial opinions in which a factor was found to favor or disfavor a likelihood of confusion (or some other outcome). When the intent factor is found to favor a likelihood of confusion, it reports a remarkably high mean stampede score of .798 with severe negative skewing¹³⁵ and peakedness.¹³⁶ Similarly, when the actual confusion factor is found to favor a likelihood of confusion, it reports a mean stampede score of .765, also with severe negative skewing and peakedness. The other factors do not show such characteristics.

4. Intercircuit Variation in Stampeding

The data show some variation among the circuits in their district courts' propensity to stampede the factors. Due perhaps in part to the limited number of opinions sampled from certain circuits, there are few statistically significant differences between the mean stampede score of an individual circuit as compared to that of all other circuits. The main exception is the Second Circuit. Its districts' mean stampede score for preliminary injunction and bench trial opinions that found a likelihood of confusion (.593)¹³⁷ is the lowest among the circuits reporting more than two opinions, and is significantly different from the mean stampede score of such opinions from all other circuits (.768).¹³⁸ If, as the data suggest, Second Circuit district courts appear far less prone to stampede the factors when they find a likelihood of confusion, then this may be the result of these courts' and their bars' more frequent exposure to and greater sophistication in the use of the multifactor test. However, the Ninth Circuit, which also produced a relatively large number of multifactor opinions, reports a relatively high stampede score (.801)¹³⁹ in preliminary injunction

135. Skewing refers here to the asymmetrical nature of the distribution of stampede scores. A skewed distribution of values is one that when plotted on a graph produces an asymmetrical curve. "A positively (or upward or right) skewed distribution is one in which the infrequent [values] are on the high or right side of the *x* axis, such as the scores on a difficult test. A left (or downward or negatively skewed) distribution is one in which the rare values are on the low or left side of the *x* axis, such as the scores on an easy test." W. Paul Vogt, *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences* 297 (3d ed. 2005).

136. Peakedness describes the *kurtosis* of the distribution. Kurtosis refers to "the extent to which a distribution departs from the bell-shaped or normal curve by being either pointier (leptokurtosis) or flatter (platykurtosis)." *Id.* at 166. "The basic rule for interpreting [numerical expressions] of kurtosis is that negative numbers mean flatter than normal and positive numbers mean more peaked than normal. The number for a normal distribution is zero." *Id.*

137. SD=.315, n=26.

138. SD=.270, n=76, p=.007.

139. SD=.241, n=22.

and bench trial opinions that found a likelihood of confusion, one that is significantly different from the Second Circuit's, but not from that of all other circuits.¹⁴⁰ More speculatively, the Second Circuit's relatively low mean stampede score may be further evidence of what Part II.B suggested was the circuit's slight bias as compared to other circuits against finding a likelihood of confusion.

For preliminary injunction and bench trial opinions in which no likelihood of confusion was found, intercircuit variation in mean stampede scores is muted. Interestingly, however, the Second Circuit reports a relatively large negative mean stampede score (-.502)¹⁴¹—though we cannot say that this difference is statistically significant from the mean stampede score in such cases of all other circuits (-.426).¹⁴²

IV

Factor-Specific Analysis

We have seen that the core factors drive the outcome of the test, and that some factors are far more influential than others. This Part looks more closely at how each of the five core factors operates, beginning with the similarity of the marks factor, and then considering, in rough order of the importance of the factors, the defendant's intent, the proximity of the goods, the strength of the plaintiff's mark, and the evidence of actual confusion. The Part reports a number of findings that contravene conventional wisdom in trademark law. Most notably, it demonstrates that the intent factor, thought by some to be irrelevant, is of decisive importance, that survey evidence, thought by many to be highly influential, is in practice of little importance, and that the doctrine of trademark strength, particularly as it relates to the concept of inherent distinctiveness, has broken down. In the process of reviewing the factor-specific data, this Part also briefly speculates on the implications of the data for the reform of the multifactor analysis. As originally conceived, the multifactor analysis was simply a heuristic to aid the judge in making an accurate finding of fact as to the likelihood of consumer confusion. Much that is external or contrary to that purpose has since insinuated itself into the multifactor analysis. The goal of any reform should be to restore to the multifactor analysis its empirical, fact-finding purpose.

140. The Ninth Circuit's mean stampede score was significantly different from the Second Circuit's at the .01 level of significance.

141. SD=.343, n=38.

142. SD=.386, n=52.

A. *The Core Factors*

1. *The Similarity of the Marks*

The data clearly show that the similarity of the marks factor is by far the most important factor in the multifactor test. Of course, courts and commentators have long said as much, if not more. Courts have suggested that the similarity factor can be “dispositive,”¹⁴³ and an authoritative treatise has asserted that this factor “is usually controlling.”¹⁴⁴ However, these characterizations are accurate only in the limited sense that the plaintiff must win the similarity factor in order to have any chance of winning the multifactor test.¹⁴⁵ As some courts have recognized, the similarity inquiry is a threshold inquiry.¹⁴⁶ This makes intuitive sense. It is hard to imagine a judge finding that the marks are not similar, and yet that consumers are likely to confuse them.¹⁴⁷ Remarkably, we might expect cases which produce reported preliminary injunction or bench trial opinions to be close cases that would turn on factors other than this threshold requirement. Yet, in sixty-five out of the 192 opinions sampled, the defendant failed to over-

143. See, e.g., *Nabisco, Inc. v. Warner-Lambert Co.*, 220 F.3d 43, 48 (2d Cir. 2000) (“Having determined that the parties’ use of their DENTYNE ICE and ICE BREAKERS marks is so dissimilar as to require judgment for Warner-Lambert, we need not examine the remaining *Polaroid* factors and express no view of the district court’s analysis of them.”). See also *Kaufman & Fisher Wish Co. v. F.A.O. Schwarz*, 184 F. Supp. 2d 311, 323 (S.D.N.Y. 2001) (“Here, the ‘similarity of the marks’ factor is dispositive, for the design and packaging of Alluwishes bears so little resemblance to that of Amanda Love that no reasonable factfinder could conclude that there is any likelihood of confusion between them.”); *E. Am. Trio Prods., Inc. v. Tang Elec. Corp.*, 97 F. Supp. 2d 395, 414 (S.D.N.Y. 2000) (“The strength of plaintiff’s trade dress, the proximity of the products, and even plaintiff’s evidence of actual confusion are, in the eyes of the Court, far outweighed by the fact that the overall appearance of the packaging at issue is so dissimilar. The remainder of the *Polaroid* factors carry little weight in this analysis for the reasons set forth above.”).

144. See 3 *McCarthy*, *supra* note 2, at § 23.33. (2005), cited in *CAE, Inc. v. Clean Air Eng’g, Inc.*, No. 97-C-3264, 2000 WL 28274, at *16 (N.D. Ill. Jan. 10, 2000).

145. *But see Earthquake Sound Corp. v. Bumper Indus., Inc.*, No. 98-17160, 1999 WL 638681 (9th Cir. Aug. 18, 1999). In affirming the district court’s granting of summary judgment to plaintiff, the appellate court stated:

[W]e have neither required district courts to make specific findings on this factor nor mandated reversal when a district court failed to do so. Thus, in the present case, so long as the other factors the district court considered support its ultimate determination that a likelihood of confusion exists, reversal is not mandated simply because the district court failed to make a determination of the similarity, or lack thereof, between the competing marks.

Id. at *4. The Ninth Circuit then found that based on the record before it, the marks were indeed similar. *Id.* at *5.

146. See, e.g., *Sun-Fun Prods., Inc. v. Suntan Research & Dev., Inc.*, 656 F.2d 186, 189 (5th Cir. 1981) (“The two marks must bear some threshold resemblance in order to trigger inquiry into extrinsic factors. . . .”). See also *Kirkpatrick*, *supra* note 14, at § 4:1 (“Without a threshold similarity of the marks that might result in confusion, it may even be unnecessary to weigh the other factors.”). *Cf.* *Fisons Horticulture, Inc. v. Vigoro Indus., Inc.*, 30 F.3d 466, 476 n.11 (3d Cir. 1994) (“We have emphasized the importance of the similarity of the marks in likelihood of confusion, but we have not ranked the factors otherwise.” (citation omitted)).

147. See *Playmakers, LLC v. ESPN, Inc.*, 297 F. Supp. 2d 1277, 1282 (W.D. Wash. 2003) (“Without similarity, there can be no confusion.”).

come even this initial hurdle. Perhaps in these opinions, courts decided the outcome of the close case and then conformed their finding under the similarity factor to that outcome. However, in the sixty-five opinions in which courts found that the similarity factor disfavored a likelihood of confusion, courts nevertheless found that only about half of the other factors considered also disfavored a likelihood of confusion.¹⁴⁸ The courts declined, in other words, to conform the outcomes of other, less important factors to the outcome of the overall test.

As for the degree of similarity between the marks, the data call into question the conventional wisdom that plaintiffs stand a better chance of establishing a likelihood of confusion when the parties' marks are identical rather than merely similar.¹⁴⁹ In seventy (21%) of the 331 opinions sampled, the court explicitly found that the marks were "identical," the "same," or "nearly" so.¹⁵⁰ For opinions in which the parties' goods were found to be competitive, the plaintiffs' multifactor test win rate when the parties' marks were found to be identical (.881)¹⁵¹ was higher than, but not significantly different from, the plaintiffs' win rate when the parties' marks were found merely to be similar, but not identical (.806).¹⁵² The same is true for opinions in which the parties' goods were found not to be competitive. In these opinions, the plaintiffs' multifactor test win rate when the parties' marks were found to be identical (.363)¹⁵³ was again higher than, but not significantly different from, the plaintiffs' win rate when the parties' marks were found to be merely similar (.214).¹⁵⁴

The implications of the similarity factor data for the reform of the multifactor test are not completely clear. What is certain, though not easily shown empirically, is that the similarity inquiry in trademark law, as in

148. As Table 3 shows, the multifactor stampede score in such opinions was -.502.

149. See, e.g., 3 McCarthy, *supra* note 2, at § 23:20 ("Cases where a defendant uses an identical mark on competitive goods hardly ever find their way into the appellate reports. Such cases are 'open and shut' and do not involve protracted litigation to determine liability for trademark infringement."), quoted in *Wynn Oil Co. v. Thomas*, 839 F.2d 1183, 1191 (6th Cir. 1988).

150. See, e.g., *GTFM, Inc. v. Solid Clothing, Inc.*, 215 F. Supp. 2d 273, 295 (S.D.N.Y. 2002) ("Solid directly copied GTFM's mark and used it in an identical way on the same kind of garments."); *K2 Advisors, LLC v. K2 Volatility Fund, LP*, No. 02 CIV. 3984, 2002 WL 31235701, at *11 (S.D.N.Y. Oct. 4, 2002) ("substantially identical"); *American Fid. & Liberty Ins. Co. v. American Fid. Group*, No. 97-4307, 2000 U.S. Dist. LEXIS 13863, at *19 (E.D. Pa. Sep. 25, 2000) ("virtually identical"); *Patsy's Brand, Inc. v. I.O.B. Realty, Inc.*, 99 Civ. 10175, 1999 U.S. Dist. LEXIS 19376, at *5 (S.D.N.Y. Dec. 17, 1999) ("The name used is exactly the same.").

151. n=59.

152. n=98.

153. n=11.

154. n=28. Cf. Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 Va. L. Rev. 2099, 2136 (2004) (arguing that "a rule conclusively presuming confusion when marks are identical and the defendant competes directly for the same consumers will reduce administrative costs and eliminate erroneous acquittals and their associated costs").

copyright law,¹⁵⁵ is a frustratingly nebulous and unsystematic inquiry, one that is typically little more than an exercise in abstract formal comparison.¹⁵⁶ Cognitive science has struggled to define and discipline the notion of similarity,¹⁵⁷ and trademark doctrine has done no better. The doctrine urges judges to consider similarities of “sound, sight, and meaning,”¹⁵⁸ to view each mark as a whole,¹⁵⁹ and to emphasize similarities over differences.¹⁶⁰ But beyond that, the inquiry is wide-open and not well-suited to the fact-finding purpose of the multifactor test. It may be reassuring to discover that, in practice, a court must find the marks to be similar if it is also to find an overall likelihood of confusion, and furthermore, that the degree of similarity of the marks does not appear to significantly affect the outcome of the test; instead, having crossed the similarity threshold, the plaintiff must rely on other factors, some of them intensely empirical in orientation, to prevail in the ultimate analysis. Nevertheless, it is undeniable that those plaintiffs who do win the similarity factor tend also to win the multifactor test. In preliminary injunction and bench trial opinions, 83% of plaintiffs who won the similarity factor won the test. The corresponding figure is 90% in opinions addressing summary judgment motions brought by the plaintiff. These kinds of win rates are tantalizing, though far from conclusive, evidence that the formal inquiry as to similarity exerts an inordinate degree of influence, amounting to a de facto strong presumption, on the outcome of what should otherwise be an essentially empirical rather than formal fact-finding analysis.

155. See, e.g., William Patry, *Does the Substantial Similarity Analysis Make Sense?*, The Patry Copyright Blog, Sept. 8, 2005, <http://williampatry.blogspot.com/2005/09/does-substantial-similarity-analysis.html> (last visited Sept. 16, 2006).

156. See, e.g., Thane Int'l, Inc. v. Trek Bicycle Corp., 305 F.3d 894, 903 (9th Cir. 2002) (“OrbiTrek contains the two syllable prefix ‘Orbi,’ while TREK does not. So OrbiTrek has three times as many syllables as TREK and twice as many letters. In *Entrepreneur Media* we held that a reasonable fact finder could find ‘Entrepreneur’ dissimilar from both ‘Entrepreneur Illustrated’ and ‘EntrepreneurPR.’ In particular, we reasoned that ‘Entrepreneur Illustrated’ is almost twice as long—to both the eye and ear—as ‘Entrepreneur,’ and ‘EntrepreneurPR’ contains two more syllables than ‘Entrepreneur.’ Moreover, the TREK trademark appears with all four letters capitalized, distinguishing it visibly from OrbiTrek.” (citation omitted)). See also *FW OmniMedia Corp. v. Toyota Motor Sales, U.S.A., Inc.*, No. 04-8624, 2004 U.S. Dist. LEXIS 27464, at *4-5 (C.D. Cal. Dec. 8, 2004) (comparing the marks “FW>>forward” and “FW forward>>”); *Playmakers, LLC v. ESPN, Inc.*, 297 F. Supp. 2d 1277, 1282-83 (W.D. Wash. 2003) (comparing plaintiff’s and defendant’s use of the mark playmakers).

157. See generally Robert L. Goldstone & Ji Yun Son, *Similarity*, in *Cambridge Handbook of Thinking and Reasoning* 13 (Keith J. Holyoak & Robert G. Morrison eds., 2005) (surveying cognitive science research on the human perception of similarity); Douglas L. Medin, Robert L. Goldstone, & Dedre Gentner, *Respects for Similarity*, 100 *Psych. Rev.* 254 (1993) (calling for research into the dynamic cognitive processes by which similarity is perceived).

158. *Henri’s Food Prods. Co., Inc. v. Kraft, Inc.*, 717 F.2d 352, 355 (7th Cir. 1983).

159. *Sweetwater Brewing Co. v. Great Am. Restaurants, Inc.*, 266 F. Supp. 2d 457, 462 (E.D. Va. 2003) (discussing the “anti-dissection” rule in trademark infringement doctrine).

160. See *Universal Money Ctrs., Inc. v. Am. Tel. & Tel. Co.*, 22 F.3d 1527, 1531 (10th Cir. 1994) (stating that, in evaluating a trademark infringement claim, “similarities between marks should be given more weight than differences”).

In their treatment of the similarity factor, then, judges may often quite successfully employ a kind of *Take The Best* heuristic to come quickly to a conclusion, especially when that conclusion is that there is no likelihood of confusion in light of the clear formal dissimilarity of the marks.¹⁶¹ But judges should also be wary of allowing their initial intuitions, formed in considering the formal similarity of the marks, to predispose the remainder of their analysis. Here, the data with respect to a finding that the marks are identical is indeed reassuring. Such a finding does not appear to bias the remainder of the analysis any more than does a finding that the marks are merely similar.¹⁶²

2. *The Defendant's Intent*

Courts have long expressed conflicting views on what role the intent factor should and does play in the multifactor test. Some circuits have held, as the Second has, that a finding of bad faith intent creates a “rebuttable legal presumption that the actor’s intent to confuse will be successful.”¹⁶³ Others have held that a finding of bad faith intent may “justify the inference”¹⁶⁴ of confusing similarity or is entitled to “great weight,”¹⁶⁵ but have declined to establish a presumption.¹⁶⁶

161. *Cf.* Nat'l Collegiate Athletic Ass'n. v. Bd. of Regents of Univ. of Okla., 468 U.S. 85, 110 n.39 (1984) (“The essential point is that the rule of reason can sometimes be applied in the twinkling of an eye.” (citation omitted)).

162. There is no substantial variation among the circuits in the wording of the similarity factor itself or in the doctrine underlying the factor. The circuits tend to phrase the factor as simply the “similarity of the marks” or the “degree of similarity between the marks.” *See, e.g.*, I.P. Lund Trading ApS v. Kohler Co., 163 F.3d 27, 45 (1st Cir. 1998) (“similarity of the marks”); Sally Beauty Co. v. Beautyco, Inc., 304 F.3d 964, 972 (10th Cir. 2002) (“degree of similarity between the marks”). *But see In re E. I. Du Pont de Nemours & Co.*, 476 F.2d 1357, 1361 (C.C.P.A. 1973) (considering “[t]he similarity or dissimilarity of the marks in their entireties as to appearance, sound, connotation and commercial impression”). Regression analysis suggests that there is no statistically significant variation among the circuits in how the factor is applied.

163. *Samara Bros., Inc. v. Wal-Mart Stores, Inc.*, 165 F.3d 120, 127 (2d Cir. 1998), *rev'd on other grounds*, *Wal-Mart Stores, Inc. v. Samara Bros., Inc.*, 529 U.S. 205 (2000); *Mobil Oil Corp. v. Pegasus Petroleum Corp.*, 818 F.2d 254, 258 (2d Cir. 1987) (intentional copying gives rise to a presumption of a likelihood of confusion); *Harlequin Enters. v. Gulf & W. Corp.*, 644 F.2d 946, 949 (2d Cir. 1981) (“Evidence of conscious imitation is pertinent because the law presumes that an intended similarity is likely to cause confusion.”); *My-T Fine Corp. v. Samuels*, 69 F.2d 76, 77 (2d Cir. 1934) (Hand, J.) (“[S]uch an intent raises a presumption that customers will be deceived.”). *Cf.* Kirkpatrick, *supra* note 14, at § 8:3.2 (“Some panels of the Second Circuit Court of Appeals still honor the presumption, sometimes in the breach.”). *See also Osem Food Indus. Ltd. v. Sherwood Foods, Inc.*, 917 F.2d 161, 16 U.S.P.Q.2d 1646, 1649 (4th Cir. 1990) (“Logic requires . . . that from such intentional copying arises a presumption that the newcomer is successful and that there is a likelihood of confusion. It would be inconsistent not to require one who tries to deceive customers to prove they have not been deceived.”).

164. *Blue Bell Bio-Med. v. Cin-Bad, Inc.*, 864 F.2d 1253, 1259 (5th Cir. 1989) (internal emphasis omitted).

165. *Kendall-Jackson Winery Ltd. v. E. & J. Gallo Winery*, 150 F.3d 1042, 1052 n.11 (9th Cir. 1998) (“[P]roof of intent to cause confusion is entitled to *great weight*, not that it creates a presumption of confusion that shifts the burden of proof to the other party”) (emphasis in original). *Cf.* *Nautilus Group, Ind. v. Icon Health & Fitness, Ind.*, 372 F.3d 1330, 1337 (Fed. Cir. 2004) (“Recently . . . the

At the same time, certain circuits have declared that “intent is largely irrelevant in determining if consumers likely will be confused as to source.”¹⁶⁷ As one district court put it, “the presence or absence of intent does not impact the perception of consumers whose potential confusion is at issue.”¹⁶⁸ Notwithstanding the presumptive force it appears to accord to the intent factor, the Second Circuit has been particularly critical of it. Judge Leval, the circuit’s most influential intellectual property jurist, has addressed intent at length:

Bad faith on the part of a party can influence the court in at least two ways. First, where a substantive issue such as irreparable harm or likelihood of confusion is a close question that could reasonably be called either way, a party’s bad faith could cause it to lose the benefit of the doubt. Second, if prospective entitlement to relief has been established, the good or bad faith with which the parties had conducted themselves could influence the court in the fashioning of appropriate equitable relief, or even cause it to deny equitable relief to a party that had conducted itself without clean hands. A preliminary injunction can have drastic consequences—potentially putting a party out of business prior to trial on the merits. A court may be less concerned about imposing such drastic consequences on a party that had conducted itself in bad faith.¹⁶⁹

Other circuits, such as the Fifth, have declared the intent factor to be “critical”¹⁷⁰ or “important”¹⁷¹ in the limited sense that while the “presence of intent constitutes strong evidence of confusion, the absence of intent is

Ninth Circuit may have de-emphasized somewhat the role of the intent factor in the likelihood of confusion analysis.”)

166. *See, e.g.*, *Frehling Enter. v. Int’l Select Group*, 192 F.3d 1330, 1340, 52 U.S.P.Q.2d 1447 (11th Cir. 1999) (bad faith intent “alone may be enough to justify the inference that there is confusing similarity”). *But see* *Bauer Lamp Co. v. Shaffer*, 941 F.2d 1165, 1172 (11th Cir. 1991) (“Intent to copy in itself creates a rebuttable presumption of likelihood of confusion.”).

167. *Lois Sportswear, U.S.A., Inc. v. Levi Strauss & Co.*, 799 F.2d 867, 875 (2d Cir. 1986).

168. *Perfumania, Inc. v. Perfulandia, Inc.*, 279 F. Supp. 2d 86, 101 (D.P.R. 2003). *See also* *Chrysler Corp. v. Silva*, 118 F.3d 56, 59 n.3 (1st Cir. 1997) (“Strictly, intent, or lack thereof, does not affect the eyes of the viewer. . . . Proof of bad intent may, psychologically, hurt as an admission. Proof of good intent does not change appearance.” (citation omitted)). *Cf.* *Tillamook Country Smoker, Inc. v. Tillamook County Creamery Ass’n*, 311 F. Supp. 2d 1023, 1044 (D. Or. 2004) (“The Ninth Circuit, however, has emphasized the minimal importance of the intent factor. . . .” (internal quotations omitted)).

169. *TCPIP Holding Co. v. Haar Commc’ns*, 244 F.3d 88, 102-03 (2d Cir. 2001). *See also* *Virgin Enter. v. Nawab*, 335 F.3d 141, 151 (2d Cir. 2003) (Bad faith is not “of high relevance to the issue of likelihood of confusion. A finding that a party acted in bad faith can affect the court’s choice of remedy or can tip the balance where questions are close. It does not bear directly on whether consumers are likely to be confused.”).

170. *Amstar Corp. v. Domino’s Pizza, Inc.*, 615 F.2d 252, 263 (5th Cir. 1980).

171. *See, e.g.*, *Computer Care v. Serv. Sys. Enters., Inc.* 982 F.2d 1063, 1070-1071 (7th Cir. 1992) (referring to the intent factor as an “important factor bearing on the likelihood of confusion” (internal quotations omitted)).

irrelevant in determining likelihood of consumer confusion.”¹⁷² Fifth Circuit courts frequently state that “if the mark was adopted with the intent of deriving benefit from the reputation of [the plaintiff,] that fact alone may be sufficient to justify the inference that there is confusing similarity.”¹⁷³ However, they add, “the lack of guilt is immaterial to the evaluation.”¹⁷⁴ Courts that follow this approach typically reason that “one who intends to confuse is more likely to succeed in doing so”¹⁷⁵ or that the defendant’s intent “is relevant because it demonstrates the junior user’s true opinion as to the dispositive issue, namely, whether confusion is likely.”¹⁷⁶ Some hold more precisely that “defendant’s intent will indicate a likelihood of confusion only if an intent to confuse consumers is demonstrated via purposeful manipulation of the junior mark to resemble the senior’s.”¹⁷⁷

The data strongly reject the hypothesis that the intent factor is irrelevant to the outcome of the multifactor test. In fact, they suggest that a finding of bad faith intent creates, if not in doctrine, then at least in practice, a nearly un-rebuttable presumption of a likelihood of confusion.¹⁷⁸ All but one of the fifty preliminary injunction opinions in which the court found bad faith intent resulted in a finding of a likelihood of confusion,¹⁷⁹ and all but one of the seventeen bench trials in which the court found bad faith

172. *Gaylord Entm’t Co. v. Gilmore Entm’t Group*, 187 F. Supp. 2d 926, 949 (M.D. Tenn. 2001).

173. *Sunbeam Prods. v. W. Bend Co.*, 123 F.3d 246, 258 (5th Cir. 1997) (quoting *Chevron Chem. Co. v. Voluntary Purchasing Groups, Inc.*, 659 F.2d 695, 703-04 (5th Cir. 1981) (quoting *Amstar Corp. v. Domino’s Pizza, Inc.*, 615 F.2d 252, 263 (5th Cir. 1980)) (internal quotations omitted), *rev’d on other grounds*, *Traffix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23, 28 (2001); *see also* *Fuji Photo Film Co., Inc. v. Shinohara Shoji Kabushiki Kaisha*, 754 F.2d 591, 596 (5th Cir. 1985) (“Good faith is not a defense to trademark infringement. . . . The reason for this is clear: if potential purchasers are confused, no amount of good faith can make them less so. Bad faith, however, may, without more, prove infringement.” (citations omitted)); *Kodak Prods. Co., Inc. v. Tie Down, Inc.*, No. Civ.A.4:03-CV-1474-Y, 2004 WL 2599353, at *9 (N.D. Tex. Nov 12, 2004); *RaceTrac Petroleum, Inc. v. J.J.’s Fast Stop, Inc.*, No. Civ.A. 3:01-CV-1397, 2003 WL 251318, at *8 (N.D. Tex. Feb. 3, 2003); *Neles-Jamesbury, Inc. v. Valve Dynamics, Inc.*, 974 F.Supp. 964, 973 n.25 (S.D. Tex., 1997).

174. *Ironclad, L.P. v. Poly-America, Inc.* No. 3:98-CV-2600-P, 2000 U.S. Dist. LEXIS 10728, at *13-14 (N.D. Tex. July 28, 2000).

175. *Cable News Network L.P. v. CNNNews.com*, 177 F. Supp. 2d 506, 520 (E.D. Va. 2001).

176. *Kemp v. Bumble Bee Seafoods, Inc.*, 398 F.3d 1049, 1057 (8th Cir. 2005) (emphasis omitted). *See also* *Little Caesar Enters., Inc. v. Pizza Caesar, Inc.*, 834 F.2d 568, 572 (6th Cir. 1987) (“[A] defendant who purposely chooses a particular mark because it is similar to that of a senior user is saying, in effect, that he thinks that there is at least a possibility that he can divert some business from the senior user—and the defendant ought to know at least as much about the likelihood of confusion as the trier of fact.”).

177. *A & H Sportswear, Inc. v. Victoria’s Secret Stores, Inc.*, 237 F.3d 198, 226 (3d Cir. 2000).

178. This result is all the more remarkable in light of the fact that the data set excluded counterfeiting cases, in which the defendant’s bad faith intent is typically quite clear.

179. For the one preliminary injunction opinion sampled in which bad faith intent was found but a likelihood of confusion was not found, *see* *Do the Hustle, LLC v. Rogovich*, 03 Civ. 3870(VM), 2003 U.S. Dist. LEXIS 10445 (S.D.N.Y. June 19, 2003).

intent produced the same result.¹⁸⁰ In both of the outlying opinions, a finding of dissimilarity between the marks trumped the finding of bad faith intent,¹⁸¹ and in both, the finding of intent was weak.¹⁸² The force of a finding of bad faith intent is no less apparent in the district courts of the Second Circuit, which routinely cite to Judge Leval's reasoning on the intent factor,¹⁸³ but are otherwise no different from other district courts in appearing to treat a finding of bad faith intent as nearly dispositive.¹⁸⁴

As with the similarity factor, the precise wording and application of the intent factor does not vary substantially by circuit with one exception: the Seventh Circuit. Logistic regression analysis suggests that, all else being equal, district courts in the Seventh Circuit are less likely to find that the intent factor favors a likelihood of confusion than are the district courts of the other circuits—the coefficient was only marginally significant, however.¹⁸⁵ It is not clear why the Seventh Circuit should stand out in this way. One possible explanation goes to the wording of the intent factor in that circuit. While most circuits phrase their intent factor as simply “the defendant's intent”¹⁸⁶ or “the defendant's intent in selecting the mark,”¹⁸⁷ the Seventh Circuit is unique in inquiring into the defendant's “intent to palm

180. For the one bench trial opinion sampled in which bad faith intent was found but a likelihood of confusion was not found, see *Choice Hotels Int'l, Inc. v. Kaushik*, 147 F. Supp. 2d 1242 (M.D. Ala. 2000).

181. See *Do the Hustle*, 2003 U.S. Dist. Lexis 10445, at *34-35; *Choice Hotels*, 147 F. Supp. 2d at 1249-50.

182. See *Do the Hustle*, 2003 U.S. Dist. Lexis 10445, at *34-35 (“The bad faith element may weigh in favor of Plaintiffs, since Rogovich was admittedly involved from near the inception of Breakfast Club and was likely responsible for many of the similarities between the two establishments, in violation of the Purchase Agreement. Despite Rogovich's assertion in his submissions and at the Hearing that he mistakenly understood the Purchase Agreement to mean 30 miles from Manhattan as opposed to the plain language, ‘New York City,’ such a lack of understanding of the plain words in the contract does not excuse his breach.”); *Choice Hotels*, 147 F. Supp. 2d at 1253-54 (“The court notes that it makes this finding in the absence of any evidence of actual intent to infringe . . . [Kaushik's] acts were intentional under the law, while not intentional in the literal sense of the word.”).

183. See, e.g., *Madison Avenue Caviarateria, Inc. v. Caviaria.Com*, No. 04 Civ.00493 RO, 2004 WL 744481, at *2 n. 2 (S.D.N.Y. Apr. 7, 2004); *Echo Design Group, Inc. v. Zino Davidoff S.A.*, 283 F. Supp. 2d 963, 966 (S.D.N.Y. 2003).

184. For a scholarly treatment of the intent factor supporting its importance, see Beverly W. Pattishall, *The Impact of Intent in Trade Identity Cases*, 65 Nw. U. L. Rev. 421 (1970).

185. -0.814 (.456), $z=-1.78$, $p=.074$. Logistic regression of the intent factor was performed on a dummy variable for the Seventh Circuit and dummy variables for the two most common outcomes (favors or disfavors confusion) of each of the other four core factors ($N=331$, $\chi^2=140.08$, $p>\chi^2=.000$, pseudo $R^2=.343$). The same regression was performed for each of the circuits with dummy variables for the circuit. Only the Seventh Circuit yielded a marginally significant coefficient.

186. See, e.g., *Elvis Presley Enters., Inc. v. Capece*, 141 F.3d 188, 194 (5th Cir. 1998); *Pizzeria Uno Corp. v. Temple*, 747 F.2d 1522, 1527 (4th Cir. 1984).

187. See, e.g., *Frisch's Rests., Inc. v. Elby's Big Boy of Steubenville, Inc.*, 670 F.2d 642, 648 (6th Cir. 1981); *AMF Inc. v. Sleekcraft Boats*, 599 F.2d 341, 349 (9th Cir. 1979).

off its goods” as those of the plaintiff.¹⁸⁸ This phrasing, which comes from the original *Helene Curtis* opinion that established the Seventh Circuit’s factors, refers to the old doctrine of “palming off”—or “passing off”—as it is otherwise known.¹⁸⁹ Arguably, the Seventh Circuit’s phrasing of the intent factor calls for something more than a mere general finding of bad faith, which may be made in situations where the defendant knowingly adopted a mark to which the plaintiff had a prior claim, but where the defendant did not, strictly speaking, intend to cause consumer confusion. The Seventh Circuit’s phrasing calls instead for a more specific finding that the defendant knowingly adopted a mark similar to the plaintiff’s with the intent to confuse consumers as to source.¹⁹⁰

It is black-letter doctrine across the circuits, including the Seventh Circuit, that bad faith intent may be inferred solely from the fact that the parties’ marks are similar and the fact that the defendant had knowledge of the plaintiff’s mark when it adopted its own, similar mark.¹⁹¹ The data suggest that this circumstantial inference is the leading basis for a finding of bad faith intent. District courts found bad faith in 102 of the 331 opinions sampled. In fifty-eight of these 102 opinions, the court based its finding of bad faith at least in part on the combination of similarity and defendant’s knowledge.¹⁹² In thirty of these fifty-eight opinions, the court also based its finding on direct evidence of bad faith, such as documents produced by the defendant or actions of the defendant after receiving a cease and desist de-

188. See *Helene Curtis Indus., Inc. v. Church & Dwight Co.*, 560 F.2d 1325, 1330 (7th Cir. 1977). But see *Anheuser-Busch, Inc. v. Balducci Publ’ns*, 28 F.3d 769, 774 (8th Cir. 1994) (considering “the alleged infringer’s intent to confuse the public”).

189. See 1 McCarthy, *supra* note 2, at § 5:2.

190. For a recent example of how this fine distinction works out in practice, see *Starbucks Corp. v. Wolfe’s Borough Coffee, Inc.*, No. 01 Civ. 5981 (LTS) (THK), 2005 WL 3527126, at *6 (S.D.N.Y. Dec. 23, 2005). See also *Yankee Candle Co., Inc. v. Bridgewater Candle Co.*, 259 F.3d 25, 45 (1st Cir. 2001) (“[T]he relevant intent is not just the intent to copy, but to ‘pass off’ one’s goods as those of another. Given that Bridgewater prominently displayed its trade name on its candles, we do not think that the evidence of copying was sufficiently probative of secondary meaning.” (citation omitted)). Cf. *Blanchard v. Hill*, 2 Atk. 484, 26 Eng. Rep. 692 (Dec. 18, 1742) (holding that the defendant’s use of a mark identical to plaintiff’s on identical goods was not actionable in the absence of fraudulent intent to pass off defendant’s goods as those of plaintiff). Cf. Kirkpatrick, *supra* note 14, at § 1:1.3 (“In sum, the terms ‘passing off’ and ‘palming off’ principally serve plaintiffs as colorful, pejorative accusations of literally underhanded misconduct by defendants.”).

191. See 3 McCarthy, *supra* note 2, at § 23:115.

192. See, e.g., *Bliss Clearing Niagara, Inc. v. Midwest Brake Bond Co.*, 339 F. Supp. 2d 944, 967 (W.D. Mich. 2004) (“Direct evidence of intentional copying is not necessary to prove intent. Rather, the use of a contested mark with knowledge of the protected mark at issue can support a finding of intentional copying.” (citation omitted)); *Pfizer, Inc. v. Y2K Shipping & Trading, Inc.*, No. 00 CV 5304(SJ), 2004 WL 896952, at *5 (E.D.N.Y. Mar. 26, 2004) (“Bad faith . . . is established where there is evidence of actual knowledge of the senior user’s mark and the marks are so similar that it seems clear that deliberate copying has occurred.” (citation omitted)).

mand from the plaintiff.¹⁹³ In only thirty-seven of the 102 opinions did the court base its finding solely on direct evidence without explicitly mentioning the combination of similarity and defendant's knowledge.¹⁹⁴

Finally, what light do the data shed on the conflict among the courts concerning the proper role of intent in the multifactor analysis? Even more so than the similarity factor data, the intent factor data suggest that a finding of bad faith intent exerts excessive influence on the outcome of the multifactor test. The facile assumption, evidently quite pervasive among the courts, that if the defendant intended to confuse, then it succeeded in doing so does not do justice to the great diversity of trademark infringement fact patterns before the courts. Further, it loosens the focus of the multifactor analysis on what should be the overriding empirical question of whether consumers are likely to be confused.¹⁹⁵ To be sure, in light of the defendant's bad faith, courts employ the multifactor test to reach what they deem to be the right result. But if trademark law seeks to prevent commercial immorality, then it should do so explicitly. An injunction should issue and damages be granted on that basis alone, and not on the basis of possibly distorted findings of fact as to the likelihood of consumer confusion.

3. *The Proximity of the Goods*

As they have with the intent factor, courts have expressed conflicting views about the importance of the proximity of the goods factor. The purpose of the proximity factor is to consider whether "the [parties'] goods are similar enough that a customer would assume they were offered by the same source,"¹⁹⁶ and also to consider whether "buyers and users of each parties' goods are likely to encounter the goods of the other, creating an assumption of common source affiliation or sponsorship."¹⁹⁷ We saw above that the Sixth Circuit has called the proximity factor "the most important inquiry" in the multifactor analysis.¹⁹⁸ Another court has referred to it as

193. See, e.g., *Telerep Caribe, Inc. v. Zambrano*, 146 F. Supp. 2d 134, 143 (D.P.R. 2001); *CFM Majestic, Inc. v. NHC, Inc.*, 93 F. Supp. 2d 942, 956-58 (N.D. Ind. 2000); *Teaching Co. Ltd. P'ship v. Unapix Entm't, Inc.*, 87 F. Supp. 2d 567, 582-83 (E.D. Va. 2000).

194. See, e.g., *Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc.*, 275 F. Supp. 2d 543, 572 (D.N.J. 2003); *Discovery Commc'ns, Inc. v. Animal Planet, Inc.*, 172 F. Supp. 2d 1282, 1290 (C.D. Cal. 2001).

195. Compare the approach the Trademark Trial and Appeal Board takes to intent when it considers trademark deceptiveness. See *In re House of Windsor, Inc.*, 221 U.S.P.Q. 53, 56 (T.T.A.B. 1983) ("[I]t seems to us that intent of the user of the mark should not be an element of a case of geographical deceptiveness. . . . The better approach, we believe, is to determine whether the deception is material to the purchasing decision. If so, the mark is deceptive within the meaning of Section 2(a)."). See also Robert Brauneis & Roger E. Schechter, *Geographic Trademarks and the Protection of Competitor Communication*, 96 Trademark Rep. 782 (2006).

196. *Checkpoint Sys., Inc. v. Check Point Software Techs., Inc.*, 269 F.3d 270, 286 (3d Cir. 2001).

197. *Id.*

198. See *Homeowners Group, Inc. v. Home Mktg. Specialists, Inc.*, 931 F.2d 1100, 1109 (6th Cir. 1991).

“extremely important.”¹⁹⁹ Yet the Second Circuit has speculated that “since modern marketing methods tend to unify widely different types of products in the same retail outlets or distribution networks, this factor is not of over-riding importance.”²⁰⁰

The Sixth Circuit has also more precisely characterized the proximity factor as important at the extremes, when it weighs strongly in favor of one party or the other, but not important otherwise:

This court has identified three categories regarding the relatedness of the goods or services with which trademarks are associated: First, if the parties compete directly by offering their goods or services, confusion is likely if the marks are sufficiently similar; second, if the goods or services are somewhat related but not competitive, the likelihood of confusion will turn on other factors; third, if the goods or services are totally unrelated, confusion is unlikely.²⁰¹

The data support this account. In preliminary injunction and bench trial opinions in which the parties’ goods were found to be identical and their marks were found to be similar, plaintiffs won the multifactor test 100% of the time (n=28). In summary judgment opinions of this nature, plaintiffs’ multifactor test win rate was also very high (.813).²⁰² On the other extreme, in all opinions, regardless of posture, in which the proximity factor was found to disfavor a likelihood of confusion, the plaintiffs’ multifactor test win rate was exceedingly low (.027).²⁰³ Finally, in those opinions in which the proximity factor was found to favor a likelihood of confusion but the goods were not identical, the plaintiffs win rate was unexceptional (.624),²⁰⁴ suggesting that the outcome of the test did indeed turn on other factors.

Overall, the degree of proximity of the goods appears significantly to affect the outcome of the test. We saw above, under the similarity factor, that the plaintiffs’ win rate in opinions in which the parties’ marks were found to be identical was not significantly different from their win rate in opinions in which the parties’ marks were found to be similar but not identical. Under the proximity factor, however, the story is different. The overall plaintiff win rate in identical goods cases (.808)²⁰⁵ was significantly

199. See, e.g., *Playmakers, LLC v. ESPN, Inc.*, 297 F. Supp. 2d 1277, 1282 (W.D. Wash. 2003) (“Whether the parties’ services are related is extremely important—perhaps second only to similarity—and the Court places great weight on this factor.”).

200. *Vitaroz Corp. v. Borden, Inc.*, 644 F.2d 960, 967 (2d Cir. 1981) (citation omitted).

201. *Therma-Scan, Inc. v. Thermoscan, Inc.*, 295 F.3d 623, 632 (6th Cir. 2002).

202. n=16.

203. n=73.

204. n=157.

205. n=52.

higher than the plaintiff win rate (.624)²⁰⁶ in opinions that found the parties' goods to be similar, yet not identical.²⁰⁷

If the ideal multifactor analysis seeks only to discover the facts on the ground, then the proximity data should be seen as encouraging. Courts appear still to pay attention to the specific circumstances of the marketplace rather than rely on such general principles as those articulated above by the Second Circuit. Relatedly, the proximity data are also an interesting corrective to the general tendency of trademark policymaking²⁰⁸ and scholarship²⁰⁹ to focus on the big marks and big marketplaces. The proximity data suggest that the daily workings of trademark litigation are still characterized by the clash of small, or at least non-famous, marks, for which the proximity of the goods consideration is still important.

4. *The Strength of the Plaintiff's Mark*

Courts assess the degree of strength or distinctiveness of the plaintiff's mark on the assumption that the stronger the mark, the more protection it should receive. In conducting their analysis, courts are generally instructed to consider two forms of trademark distinctiveness:

[I]nherent distinctiveness[] examines a mark's theoretical potential to identify plaintiff's goods or services without regard to whether it has actually done so. . . . [A]cquired distinctiveness[] refers to something entirely different. This measure looks solely to that recognition plaintiff's mark has earned in the marketplace as a designator of plaintiff's goods or services.²¹⁰

The data on the strength factor yield what are probably the most interesting factor-specific results in the study, if also the most ambiguous. The data suggest that, at least in the context of the multifactor test, the doctrine of trademark strength has broken down. Basic concepts are no longer consistently applied and mistakes of doctrine are common.²¹¹ Nevertheless, in

206. n=157.

207. p=.015. The data show no significant intercourt variation in the application of the proximity factor.

208. See, e.g., Trademark Dilution Revision Act of 2005: Hearing on H.R. 683 Before the Subcomm. on Courts, the Internet, and Intellectual Property of the House Comm. on the Judiciary, 109th Cong. (2005) (considering the reform of U.S. antidilution protection for famous marks).

209. See, e.g., Barton Beebe, *Search and Persuasion in Trademark Law*, 103 Mich. L. Rev. 2020 (2005) (concentrating primarily on trademark doctrine as it relates to famous marks) [hereinafter Beebe, *Search and Persuasion*]; Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. Rev. 621 (2004) (same) [hereinafter Beebe, *Semiotic Analysis*].

210. *Brennan's, Inc. v. Brennan's Rest., L.L.C.*, 360 F.3d 125, 131 (2d Cir. 2004). See also *Walmart Stores, Inc. v. Samara Bros.* 529 U.S. 205, 210 (2000) ("[A] mark is inherently distinctive if '[its] intrinsic nature serves to identify a particular source.'" (citation omitted)); *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159, 163 (1995) ("The imaginary word 'Suntost,' or the words 'Suntost Marmalade,' on a jar of orange jam immediately would signal a brand or a product 'source. . . .").

211. See, e.g., *Int'l Bancorp, L.L.C. v. Societe des Bains de Mer et du Cercle des Etrangers a Monaco*, 192 F. Supp. 2d 467, 483 (E.D. Va. 2002) ("First, it has been established that SBM's 'Casino

opinions that do address the issue of trademark strength, and inherent strength in particular, there is a surprisingly good correlation between inherent strength and success in the multifactor test.

a. The Breakdown of Abercrombie

Established by Judge Friendly in 1976, the *Abercrombie* spectrum of trademarks classifies marks according to their degree of inherent distinctiveness.²¹² *Fanciful marks* are coined terms and are thought to have the highest degree of inherent distinctiveness (i.e., xerox). Second in the hierarchy are *arbitrary marks*, which have no semantic connection to the products to which they are affixed (i.e., apple computers). Third are *suggestive marks* that are suggestive of or metaphorically related to their products' characteristics (i.e., ivory soap). *Descriptive marks* (i.e., coca-cola) are thought to lack inherent distinctiveness, as are *generic marks* (i.e., aspirin²¹³). It remains black-letter doctrine that the more inherently distinctive a mark is, the greater the scope of protection it should receive. Nevertheless, courts have occasionally expressed their dissatisfaction with the *Abercrombie* hierarchy. Judge Easterbrook in particular has criticized its formalism:

We have said before that “arbitrary,” “suggestive” and the other words in the vocabulary of trademark law may confuse more readily than they illuminate. . . a caution litigants should take

de Monte Carlo' mark has secondary meaning and is, therefore, strong and distinctive.”); Guinness United Distillers & Vintners B.V. v. Anheuser-Busch, Inc., 02 Civ. 0861 (LMM), 2002 U.S. Dist. LEXIS 12722, at *8 (S.D.N.Y. July 12, 2002) (“Both ‘Red Label’ and ‘Johnnie Walker Red Label’ are federally registered marks for scotch whiskey. . . . As such, they are entitled to a rebuttable presumption of distinctiveness, i.e., that the marks are suggestive or arbitrary rather than descriptive.”); Chum Ltd. v. Lisowski, 98 Civ. 5060(KMW), 2001 U.S. Dist. LEXIS 2462, at *28-29 (S.D.N.Y. Mar. 12, 2001) (“Because plaintiff’s marks are generic, the strength of these marks must be demonstrated through secondary meaning.”); Ironclad, L.P. v. Poly-America, Inc., No. 3:98-CV-2600-P, 2000 U.S. Dist. LEXIS 10728, at *10 (N.D. Tex. July 28, 2000) (“Parties agreed Iron Sak(R) is suggestive, which means neither strong nor weak.”); First Jewellery Co. of Can., Inc. v. Internet Shopping Network LLC, 2000 U.S. Dist. LEXIS 794, at *12-13 (S.D.N.Y. Feb. 1, 2000) (“In this Circuit, however, ‘[a] term that is merely self-laudatory, such as “plus” or “super,” seeking to convey the impression that a product is excellent or of especially high quality, is generally deemed suggestive,” quoting Estee Lauder Inc. v. The Gap, Inc., 108 F.3d 1503, 1509 (2d Cir. 1997)).

212. See *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, 9 (2d Cir. 1976) (“The cases, and in some instances the Lanham Act, identify four different categories of terms with respect to trademark protection. Arrayed in an ascending order which roughly reflects their eligibility to trademark status and the degree of protection accorded, these classes are (1) generic, (2) descriptive, (3) suggestive, and (4) arbitrary or fanciful.”). See also *Wal-Mart Stores, Inc.*, 529 U.S. at 210-11 (“[W]ord marks that are ‘arbitrary’ (‘Camel’ cigarettes), ‘fanciful’ (‘Kodak’ film), or ‘suggestive’ (‘Tide’ laundry detergent) are held to be inherently distinctive.”).

213. The Bayer Company originally coined the term “aspirin” as a trademark for acetyl salicylic acid. In 1921, Judge Learned Hand found that aspirin had lost its significance as the designation of a particular source of acetyl salicylic acid and had become a generic term for the substance itself. See *Bayer Co. v. United Drug Co.*, 272 F. 505 (D.N.Y. 1921). The same fate befell, for example, the term “cellophane.” See *DuPont Cellophane Co. v. Waxed Prods. Co.*, 85 F.2d 75 (2d Cir. 1936).

seriously before arguing cases so that everything turns on which word we pick. It is better to analyze trademark cases in terms of the functions of trademarks. That frees the arguments from the clutches of Webster's Third and the conflicting advice of text writers.²¹⁴

As for the bar, one respected trademark practitioner and commentator, the late Beverly W. Pattishall, went so far as to refer to the "artificial and regrettable 'four pigeon hole' rule" established in *Abercrombie* as "[o]ne of the worst blights [on the law] . . . which has spread from the Second Circuit and now appears to be settling in generally."²¹⁵

The data suggest that Mr. Pattishall need not have worried, and that Judge Easterbrook's altogether sensible call has been answered. First, in the context of the likelihood of confusion inquiry, district courts appear to make little use of the *Abercrombie* spectrum and the concept of inherent distinctiveness that underlies it. Courts failed to specify whether or not the mark at issue was inherently distinctive in 40% of the 192 preliminary injunction and bench trial opinions sampled and in 50% of the 139 summary judgment opinions sampled, for an overall failure rate of 44% in the 331 opinions examined. Overall, only 193 or 58% of the 331 opinions sampled made some use of the *Abercrombie* spectrum, and twenty-nine of these opinions neglected to place the plaintiff's mark in one specific *Abercrombie* category. Instead, they opted to make such estimations as that the mark was "suggestive . . . even though it contains descriptive elements,"²¹⁶ "fanciful and arbitrary,"²¹⁷ "arbitrary or suggestive,"²¹⁸ or is "at least a suggestive mark and is arguably an arbitrary or fanciful mark."²¹⁹ The breakdown of the *Abercrombie* analysis is even more apparent in the context of claims for trade dress infringement, in other words, for the infringement of a product's packaging or configuration.²²⁰ The court failed to specify where

214. *Scandia Down Corp. v. Euroquilt, Inc.*, 772 F.2d 1423, 1431 n.3 (7th Cir. 1985).

215. Beverly W. Pattishall, *The Lanham Trademark Act—Its Impact over Four Decades*, 76 Trademark Rep. § 24, at 57.

216. *See Macia v. Microsoft Corp.*, 335 F. Supp. 2d 507, 515 (D. Vt. 2004).

217. *See, e.g., PACCAR, Inc. v. TeleScan Techs., L.L.C.*, 115 F. Supp. 2d 772, 776 (E.D. Mich. 2000) ("The marks were derived as variations of the names of the original developers and fall into the categories of fanciful and arbitrary."); *E. Am. Trio Prods., Inc. v. Tang Elec. Corp.*, 97 F. Supp. 2d 395, 410 (S.D.N.Y. 2000).

218. *Eurotech, Inc. v. Cosmos European Travels Aktiengesellschaft*, 213 F. Supp. 2d 612, 620 n.16 (E.D. Va. 2002) ("[B]ecause it is clear that these marks are either arbitrary or suggestive, they are deemed distinctive without requiring a showing of secondary meaning.>").

219. *KeyCorp v. Key Bank & Trust*, 99 F. Supp. 2d 814, 823 (N.D. Ohio 2000).

220. In *Wal-Mart Stores, Inc. v. Samara Bros.*, 529 U.S. 205 (2005), Justice Scalia discussed the meaning of the term "trade dress":

The breadth of the definition of marks registrable under § 2 [of the Lanham Act], and of the confusion producing elements recited as actionable by § 43(a), has been held to embrace not just word marks, such as "Nike," and symbol marks, such as Nike's "swoosh" symbol, but also "trade dress"—a category that originally included only the packaging, or "dressing," of a product, but in recent years has been expanded by many courts of appeals to encompass the design of a product.

See id. at 209.

in the *Abercrombie* spectrum the trade dress fell in twenty-seven of the fifty-three opinions (51%) that adjudicated a claim only for trade dress infringement and in nine of the eleven opinions (82%) that adjudicated a claim for both trademark and trade dress infringement.²²¹

Second, and more strikingly, in those opinions in which the court's assessment of the mark's inherent strength was at odds with its assessment of the mark's acquired strength, a finding of acquired strength (or weakness) almost invariably trumped a finding of inherent weakness (or strength). For example, courts found marks to be inherently weak but commercially strong in twenty-three of the opinions sampled. In twenty-two of these opinions, the court found that the strength factor favored confusion. The one outlying opinion found, somewhat ambiguously, "a mark strength somewhere between strong and weak."²²² As for the inverse situation, courts found marks to be inherently strong but commercially weak in twenty-seven of the opinions sampled. In twenty-four of these opinions, the court found that the strength factor did not favor confusion. Of the three outliers, one found that the strength factor "favors neither party,"²²³ another explicitly narrowed its finding of strength,²²⁴ and a third found as it did because the plaintiff failed to present any evidence of commercial strength.²²⁵ Ultimately, these results should not be surprising. Though most appellate courts have not yet come around to acknowledging it, district courts appear in practice to have recognized that the mark's acquired or "actual strength" in the marketplace logically incorporates the effects of the mark's inherent strength. Indeed, the results suggest that courts need not even consider inherent strength in their assessment of the strength factor, or that if they do, inherent strength should properly be understood as merely one factor—among others such as advertising expenditure, length of time of use of the mark, revenues associated with the mark, and third-party uses²²⁶—that a court should consider in assessing a mark's actual strength.²²⁷

221. Cf. Graeme Dinwoodie, *Reconceptualizing the Inherent Distinctiveness of Product Design Trade Dress*, 75 N.C. L. Rev. 471, 508-20 (1997) (discussing the inaptness of the *Abercrombie* spectrum for the analysis of the inherent distinctiveness of product packaging and configuration).

222. *Pocono Int'l Raceway, Inc. v. Pocono Mountain Speedway, Inc.*, 171 F. Supp. 2d 427, 430 (M.D. Pa. 2001).

223. *Nat'l Distillers Prods. Co. v. Refreshment Brands, Inc.*, 198 F. Supp. 2d 474, 481 (S.D.N.Y. 2002).

224. See *Welch Allyn, Inc. v. Tyco Int'l Servs. AG*, 200 F. Supp. 2d 130, 139 (N.D.N.Y. 2002) ("[D]ue to the inherent distinctiveness of the Tyco's mark, this factor favors Plaintiff to a significant degree, but only with respect to the narrow field of stethoscopes and sphygmomanometers.").

225. See *Friesland Brands, B.V. v. Vietnam Nat'l Milk Co.*, 228 F. Supp. 2d 399, 405 (S.D.N.Y. 2002).

226. It is well-accepted that third-party uses of a mark identical or similar to plaintiff's mark is a factor in mitigation of the mark's strength and, thus, of the likelihood of confusion. See, e.g., *Amstar Corp. v. Domino's Pizza, Inc.*, 615 F.2d 252, 259-60 (5th Cir. 1980) ("The greater the number of identical or more or less similar trade-marks already in use on different kinds of goods, the less is the likelihood of confusion." (quoting Restatement of Torts § 729 (1938))). Of the sixty-three opinions that

b. The Correlation Between Inherent Strength and Success

Despite the apparent infirmity of the concept of inherent strength, we cannot close the book on it entirely. When courts did address the issue of the inherent strength of the plaintiff's mark, their findings correlated quite cleanly with the plaintiff multifactor test win rate, as Table 7 reveals. The multifactor test win rate in dispositive opinions for inherently distinctive marks (.628)²²⁸ was significantly higher than the win rate in such opinions for non-inherently distinctive marks (.242).²²⁹ More specifically, in the ninety dispositive opinions in which the court placed the plaintiff's mark in one of the five *Abercrombie* categories, the plaintiff multifactor test win rate steadily declined with the inherent strength of its mark: fanciful marks enjoyed the highest win rate, followed by arbitrary marks, suggestive marks, descriptive marks, and then generic marks. More specifically still, and underlying these win rate results, inherently distinctive marks did better on each of the core factors, and the degree of their inherent strength often closely tracked the proportion of opinions in which each of the core factors favored a likelihood of confusion.

These data support two important propositions, one championed by the trademark bar and the other by commentators on trademark doctrine. First, the data support the trademark lawyer's common advice to her clients that, at least as a matter of trademark law, if not of marketing, firms should choose inherently distinctive marks and, ideally, marks that are fanciful. As for the proposition of trademark commentators, courts frequently group into the same category trademarks that are fanciful and those that are arbitrary. Judge Friendly made this mistake when he first formulated the *Abercrombie* spectrum,²³⁰ and the Supreme Court has since failed to correct the

considered the effect of third-party uses on the strength of the plaintiff's mark, thirty-nine found that these uses mitigated strength (and thirty-three of the thirty-nine found no overall likelihood of confusion), while twenty-four found that third-party uses did not mitigate strength.

227. Another formalism has had a deleterious effect on the strength inquiry. This is the principle that incontestable marks are presumptively strong. *See, e.g., Data Concepts, Inc. v. Digital Consulting, Inc.*, 150 F.3d 620, 625 (6th Cir. 1998) ("A mark that has been registered and uncontested for five years . . . is entitled to a presumption that it is a strong mark."). The data suggest that courts make only limited use of this principle. Of the thirty out of 331 opinions that addressed the incontestable status of the plaintiff's mark as part of the strength inquiry, twenty-two found that this status supported a finding of strength (and 21 eventually found that the mark was strong), while eight found that this status did not support a finding of strength. Most of these opinions came from the Second, Sixth, and Eleventh Circuits.

228. n=78.

229. n=33.

230. *See Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, 9 (2d Cir. 1976) (referring to the fourth class of trademarks as "arbitrary or fanciful").

TABLE 7
 FREQUENCY, PLAINTIFF MULTIFACTOR TEST WIN RATE, AND PROPORTION OF FACTOR OUTCOMES
 FAVORING A LIKELIHOOD OF CONFUSION BY DISTINCTIVENESS CLASSIFICATION IN 192 PRELIMINARY
 INJUNCTION AND BENCH TRIAL OPINIONS
 (PW = plaintiff multifactor test win rate)

Panel A: Frequency, Plaintiff Multifactor Test Win Rate, and Proportion of Factor Outcomes Favoring a Likelihood of Confusion by General Inherent Distinctiveness Classification

General Distinctiveness Classification	N	PW	Proportion of Factor Outcomes Favoring a Likelihood of Confusion				
			Similarity of the Marks	Proximity of the Goods	Evidence of Actual Confusion	Strength of the Plaintiff's Mark	Defendant's Intent
Inherently Distinctive	78	.628	.705	.756	.410	.756	.333
Not Inherently Distinctive	33	.242	.424	.515	.212	.394	.212
Not Specified	76	.579	.658	.724	.355	.566	.447
Unclear	5	.200	.400	.600	.000	.200	.000
Total	192	.531	.630	.698	.344	.604	.349

Panel B: Frequency, Plaintiff Multifactor Test Win Rate, and Proportion of Factor Outcomes Favoring a Likelihood of Confusion by Specific *Abercrombie* Classification*

Specific <i>Abercrombie</i> Classification	N	PW	Proportion of Factor Outcomes Favoring a Likelihood of Confusion				
			Similarity of the Marks	Proximity of the Goods	Evidence of Actual Confusion	Strength of the Plaintiff's Mark	Defendant's Intent
Fanciful	5	.800	1.000	.600	.200	1.000	.600
Arbitrary	19	.684	.737	.737	.579	.684	.316
Suggestive	34	.588	.647	.767	.294	.588	.206
Descriptive	31	.258	.419	.548	.226	.258	.194
Generic	1	.000	1.000	.000	.000	.000	.000
Total	90	.500	.611	.667	.322	.644	.244

*The data is limited to opinions in which plaintiff's mark was classified in one, single *Abercrombie* category. Opinions in which courts speculated that the plaintiff's mark should be placed in either one *Abercrombie* category or another (e.g., "the mark is either arbitrary or suggestive") are not included.

error, if not compounded it.²³¹ Trademark commentators have argued that the two kinds of trademarks are in fact quite different, and that fanciful marks deserve a heightened degree of protection over arbitrary marks.²³² This argument is typically made on the basis of functionality concerns.²³³ The data bolster this proposition from a different angle. Perhaps most fanciful marks manage to achieve, due to their fanciful nature, greater actual strength in the marketplace than arbitrary marks, and thus they do better than arbitrary marks in trademark infringement litigation. But we should be clear that the relative success of fanciful marks in trademark infringement litigation is not due simply to their ability to satisfy some category of trademark doctrine—the data show that courts place little weight on the doctrine of inherent strength. Rather, their relative success appears to be due to the degree to which their inherent strength manifests itself in the form of actual marketplace strength.

Ultimately, as with the intent factor, the doctrine of inherent strength threatens to distort the fact-finding inquiry as to the likelihood of consumer confusion by insinuating into that inquiry policy-oriented goals that are better served elsewhere. With intent, the goal is to discourage commercial immorality. Here, with the doctrine of inherent strength, the goal is encourage the use of inherently distinctive rather than descriptive marks. These are both worthy objectives, but courts should not pursue them when they are making findings of fact about the likelihood of consumer confusion. Perhaps the standard for a finding of a likelihood of confusion should be lowered when the defendant has acted in bad faith or when the plaintiff is using an inherently distinctive mark, but the court's estimate of the likelihood of consumer confusion in the marketplace should not simply be raised in light of the presence, without more, of bad faith intent or an inherently distinctive mark. It is at least reassuring that courts appear already to have recognized this principle when they have encountered an inherently distinctive mark.

c. *Intercircuit Variation*

The doctrine underlying the strength factor varies considerably among the circuits. District courts in the First, Sixth, Eighth, and Ninth Circuits consider somewhat eccentric sub-factors when evaluating trademark strength.²³⁴ The Ninth Circuit's are arguably the most peculiar:

231. See *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 426 (2003). *But see* *Brockmeyer v. Hearst Corp.*, 248 F. Supp. 2d 281, 294 (S.D.N.Y. 2003) (“Arbitrary or fanciful are sometimes described as separate categories.” (citing *Two Pesos, Inc. v. Taco Cabana, Inc.*, 505 U.S. 763, 768 (1992))).

232. See Beebe, *Semiotic Analysis*, *supra* note 209, at 673-74.

233. See *id.*

234. For the First Circuit, see *Boston Athletic Ass'n. v. Sullivan*, 867 F.2d 22, 32 (1st Cir. 1989) (“We have found the following factors useful in determining a trademark's relative strength: the length

Two tests are commonly used to measure the strength of a mark, the “imagination test” and the “need test.” The “imagination test” focuses on the amount of imagination required in order for a consumer to associate a given mark with the goods or services it identifies. If a consumer must use more than a small amount of imagination to make the association, the mark is suggestive and not descriptive. The “need test” focuses on the extent to which a mark is actually needed by competitors to identify their goods or services. . . . The two tests are related, because the more imagination that is required to associate a mark with a product or service, the less likely the words used will be needed by competitors to describe their products or services.²³⁵

It is unclear how either of these tests helps to determine the actual strength of the mark in the marketplace. For example, “United Airlines” is hardly imaginative, and the term “united” is certainly needed by competitors, yet it is generally thought to be a very strong mark.²³⁶

Despite the wide diversity of circuit-specific doctrine underlying the strength factor, regression analysis demonstrates no significant intercircuit variation in the application of the factor, not even in the Ninth Circuit.

5. Evidence of Actual Confusion

If the factor-specific results relating to the strength factor are the most interesting in the study, the results relating to the actual confusion factor are the most disturbing. Courts consider two forms of evidence of actual confusion: (1) survey evidence and (2) direct evidence, such as testimony

of time a mark has been used and the plaintiff’s relative renown in its field; the strength of the mark in plaintiff’s field of business, especially by looking at the number of similar registered marks; and the plaintiff’s actions in promoting its mark.” (citations omitted)). For the Sixth Circuit, see *Therma-Scan, Inc. v. Thermoscan, Inc.*, 295 F.3d 623, 631 (6th Cir. 2002) (“Generally, the strength of a mark is the result of its unique nature, its owner’s intensive advertising efforts, or both.”). See also *Midwest Guar. Bank v. Guar. Bank*, 270 F. Supp. 2d 900, 910 (E.D. Mich. 2003) (“The Sixth Circuit has commented that a mark’s strength is generally a result of (1) its unique nature; (2) its owner’s intensive advertising efforts; and (3) which of the four categories the mark occupies—generic, descriptive, suggestive or arbitrary/fanciful.” (citing *Therma-Scan*, 295 F.3d at 631)). For the Eighth Circuit, see *Gateway, Inc. v. Companion Prods.*, Civ. 01-4096-KES, 2003 U.S. Dist. LEXIS 21461, at *14-15 (D.S.D. Aug. 19, 2003) (“Three factors determine the strength of a trademark: the classification of the trademark; whether the trademark is registered; and the nature, extent, and use of the trademark by others.”).

235. *Rodeo Collection, Ltd. v. W. Seventh*, 812 F.2d 1215, 1218 (9th Cir. 1987) (citations omitted).

236. At least one Ninth Circuit district court has tried to apply the imagination and need tests to determine marketplace strength. Its analysis is not encouraging. See *Ellison Educ. Equip., Inc. v. Chen*, No. SACV 02-1184-JVS(ANx), 2004 U.S. Dist. LEXIS 26947, at *35 (C.D. Cal. Dec. 21, 2004) (“Placement of the mark on the continuum, however, is only the first step; the Court must also determine the strength of the mark in the marketplace. To this end, the Court applies both the ‘imagination test’ and the ‘need test.’” (citations omitted)); *Id.* (“As opposed to ‘bold’ or ‘italics,’ the word ‘lollipop’ does not describe the font; rather, it takes imagination to associate the two. Moreover, the need aspect is low because competitors such as CommCut do not need to use the word ‘lollipop’ to describe their products. Thus, this factor weighs in favor of Ellison.”).

by consumers who were confused by the defendant's use of its mark or documents indicating such confusion. I focus here on survey evidence.

It is generally thought that survey evidence is the best evidence of actual confusion, and indeed, that a good survey has the potential to supersede the rest of the multifactor analysis.²³⁷ In a recent statement before Congress, the American Bar Association set forth the conventional view: "survey evidence is traditionally one of the most classic and most persuasive and most informative forms of trial evidence that trademark lawyers utilize in both prosecuting and defending against trademark claims of various sorts."²³⁸ Some circuits even apply an adverse inference of no likelihood of confusion if the plaintiff has the resources and time to produce survey evidence but fails to do so²³⁹—though only six opinions out of the 331 sampled drew such an inference.²⁴⁰

The data suggest that the conventional view of the utility of survey evidence may be incorrect and that the application of an adverse inference may be inappropriate. Of the 331 opinions sampled, only sixty-five (20%) addressed survey evidence, only thirty-four (10%) credited the survey evidence, and only twenty-four (7%) ultimately ruled in favor of the outcome that the credited survey evidence itself favored. More specifically, of the fifty-three opinions that addressed survey evidence presented by the plaintiff, twenty-two (42%) credited that evidence, with one opinion using the evidence against the plaintiff. Of the nineteen opinions that addressed survey evidence presented by the defendant, twelve (63%) credited that evidence, but three used the evidence against the defendant. Finally, of the seven opinions that addressed survey evidence presented by both the plaintiff and the defendant, three credited the defendant's evidence and none credited the plaintiff's.

237. See, e.g., Edward George Epstein, *Surveys: Growing Admissibility But Narrow Utilization*, 83 Trademark Rep. 863 (1993).

238. Committee Print to Amend the Federal Trademark Dilution Act: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Property of the Comm. on the Judiciary, 108th Cong. 14 (2004) (statement of Robert W. Sacoff, Chair, Section of Intellectual Property Law, American Bar Association)). See also *Government Employees Ins. Co. v. Google, Inc.*, No. 1:04CV507, 2005 WL 1903128, at *5 (E.D. Va. Aug. 8, 2005) ("To prove likelihood or absence of confusion, initial or otherwise, parties commonly introduce the results of customer or potential customer surveys."); Edelman, *supra* note 238, at 747 ("[S]urvey evidence has become de rigeur in trademark infringement cases"); Itamar Simonson, *The Effect of Survey Method on Likelihood of Confusion Estimates: Conceptual Analysis and Empirical Test*, 83 Trademark Rep. 364, 364 (1993) ("[S]urveys are now routinely employed to prove likelihood of confusion, and a failure to introduce a survey into evidence often leads to harsh criticism by the courts.").

239. See generally Sandra Edelman, *Failure to Conduct a Survey in Trademark Infringement Cases: A Critique of the Adverse Inference*, 90 Trademark Rep. 746 (2000).

240. The data set included two relevant variables: whether the court drew a general adverse inference from the plaintiff's failure to present any evidence of actual confusion and whether the court drew a specific adverse inference from the plaintiff's failure to present survey evidence. Of the 331 opinions sampled, nine drew only a general adverse inference, two drew only a specific adverse inference, and four drew both kinds of adverse inference.

It may be objected that trademark litigation is typically resolved at the preliminary injunction stage before either party has had the time or can be expected to conduct a creditable survey. It is true that the percentage of bench trial opinions that addressed survey evidence (24%) was higher than the percentage of preliminary injunction opinions that did so (16%), and that the percentage of preliminary injunction opinions that explicitly held that it was too soon to expect any evidence of actual confusion (11%) was somewhat higher than the percentage of bench trial opinions that espoused this view (9%). Yet it is still striking that survey evidence played a relatively minor role even in the bench trial context. In only thirteen out of forty-six bench trial opinions did the court address survey evidence, which it credited in eight of those opinions. Further, as stated above, only six opinions out of the 331 sampled explicitly drew an adverse inference from plaintiff's failure to present survey evidence. Four of these were preliminary injunction opinions.

There was no significant intercircuit variation in the courts' consideration of survey evidence. It may be of interest, however, that the district courts of the Second Circuit credited the plaintiff's survey evidence in only seven of the twenty-five opinions in which they considered it, and credited the defendant's survey evidence in only six of the ten opinions in which they considered it.²⁴¹

B. *The Non-Core Factors*

We saw above that the non-core factors generally have little, if any, effect on the outcome of the multifactor test. This should not be surprising. Several of these factors have no business being in the multifactor test in the first place, and the courts appear to have recognized this in practice, if not yet in doctrine. Happily, one sees here how district courts, in applying a multifactor test, may resist the dead-hand influence of various idiosyncratic and rarely relevant factors that tend to accumulate over time.

Probably the only non-core factor that deserves to be in the multifactor test is the consumer sophistication factor. It makes sense, and has been confirmed empirically, that the more sophisticated the consumers, the more care with which they will treat their search and purchasing decisions.²⁴² Nevertheless, a fairly high proportion (16%) of the opinions sampled from circuits that include the factor in their multifactor test either failed to address the factor or stated that it was not argued by the parties. The majority of these opinions were preliminary injunction or bench trial opinions. On

241. The court ruled in favor of the plaintiff in all but one of the seven opinions in which it credited plaintiff's survey evidence, see *Malaco Leaf, AB v. Promotion In Motion, Inc.*, 287 F. Supp. 2d 355 (S.D.N.Y. 2003), and ruled in favor of the defendant in all but one of the six opinions in which it credited defendant's survey evidence. See *Guinness United Distillers & Vintners B.V. v. Anheuser-Bush, Inc.*, No. 02 CIV. 0861, 2002 WL 1543817 (S.D.N.Y. 2002).

242. See generally Beebe, *Search and Persuasion*, *supra* note 209, at 2020.

the whole, across the 292 opinions sampled from circuits that explicitly consider the factor, the factor was found to disfavor a likelihood of confusion (that is, consumers were seen as sufficiently sophisticated not to be confused) 39% of the time and to favor a likelihood of confusion 28% of the time. Interestingly, although no variation exists among the circuits in the doctrine underlying the factor, the data reveal some variation among the circuits in terms of test outcomes: Logistic regression analysis suggests that the Second Circuit is significantly less likely than other circuits to find that the consumer sophistication factor disfavors a likelihood of confusion, which, if win rates are any guide, runs counter to the Second Circuit's apparent bias against finding a likelihood of confusion.²⁴³ It is not clear, and probably doubtful, however, that this had any effect on the overall outcome of the multifactor test.

The factors relating to the similarity of the parties' advertising, marketing, and sales facilities all tended to be redundant of the proximity of the goods factor in the circuits that consider these issues separately from the proximity factor. For example, among the seven possible outcomes coded for each factor, the similarity of the sales facilities factor produced exactly the same outcome as the proximity factor in 74% of the opinions sampled (n=90) that considered both factors. Of the twenty opinions in which the two factors produced divergent outcomes, six did so because they found that the proximity factor favored a likelihood of confusion and then simply did not address the sales facilities factor. The similarity of advertising or marketing methods factor produced exactly the same outcome as the proximity factor in 66% of the relevant cases sampled (n=156), and of the fifty-three opinions in which these two factors produced divergent outcomes, twelve did so because, again, they found that the proximity factor favored a likelihood of confusion and then neglected to address the advertising or marketing methods factor. The Second, Eighth, Tenth, and D.C. Circuits already consider proximity, advertising, marketing, and sales facilities together under the proximity factor.²⁴⁴ The data suggest that, at least in practice, the district courts of the other circuits do so as well.

Also redundant of the core factors are two factors used by the Third Circuit. Of the twenty-eight opinions sampled from the Third Circuit,

243. Logistic regression of the consumer sophistication factor was performed on a dummy variable for the Second Circuit and dummy variables for the two most common outcomes (favors or disfavors confusion) of each of the five core factors. The same regression was performed for each of the circuits with dummy variables for the circuit. Only the Second Circuit yielded a marginally significant coefficient ($z=-1.89$, $p>|z|=0.058$, $N=292$, $\chi^2=95.96$, $p>\chi^2=.000$, pseudo $R^2=.245$).

244. See, e.g., *Best Cellars, Inc. v. Grape Finds at Dupont, Inc.* 90 F. Supp. 2d 431, 456 (S.D.N.Y. 2000) ("The 'proximity-of-the-products' inquiry concerns whether and to what extent the two products compete with each other. The court must consider 'the nature of the products themselves and the structure of the relevant market,' including 'the class of customers to whom the goods are sold, the manner in which the products are advertised, and the channels through which the goods are sold.'" (citations and internal quotations omitted)).

twenty (71%) explicitly grouped their analysis of the factor relating to the length of time of concurrent use without evidence of actual confusion with their analysis of the actual confusion factor. Of the eight opinions that did not group their analysis, none reached substantively different outcomes under the two factors. The Third Circuit also considers the extent to which the targets of the parties' sales efforts are the same. Of the twenty-eight opinions from the circuit, seventeen (61%) produced exactly the same outcome under this factor as under the proximity factor, and only three of the remaining eleven opinions produced substantively different outcomes under the two factors.

Finally, the likelihood of bridging the gap factor, and the comparative quality of the parties' goods factor, are remarkable for the degree to which courts either ignore them or bend them to conform to the outcome of the test. Of the 217 opinions sampled from the five circuits that consider the bridge the gap factor, fifty-seven (26%) did not address the factor and another forty-four (20%) explicitly found it to be irrelevant. Strangely, twenty (nearly half) of the opinions that found the factor to be irrelevant nevertheless held that it favored a likelihood of confusion on the grounds that "there is no gap to bridge."²⁴⁵

As for the quality factor, which is considered only by the Second and D.C. Circuits, twenty-one of the 109 opinions from these circuits did not address the factor. Of the eighty-eight opinions that did, forty-five subscribed solely to the tarnishment theory of the factor,²⁴⁶ three subscribed solely to the similarity theory of the factor,²⁴⁷ and twenty-one opinions subscribed to *both* theories.²⁴⁸ Perhaps more than any other, the quality factor

245. *Id.* at 456 ("Here, there is no gap to bridge: Best Cellars and Grape Finds sell the same products in the same field. This factor, therefore, also favors Best Cellars."). *See also* Blue & White Food Prods. Corp. v. Shamir Food Indus., Ltd., 350 F. Supp. 2d 514, 521 (S.D.N.Y. 2004) ("The fourth factor favors Blue & White because if Shamir Food were permitted to bring its own 'Shamir' products into the market, there would be no gap to bridge."); *Macia v. Microsoft Corp.*, 335 F. Supp. 2d 507, 518 (D. Vt. 2004) ("As the products compete directly there is no 'gap' to be bridged.").

246. The tarnishment approach to the comparative quality factor posits that if the defendant's goods are of lower quality, then this will tarnish plaintiff's reputation and the factor should thus somehow favor a likelihood of confusion. *See, e.g.*, Prof'l Sound Servs., Inc. v. Guzzi, 349 F. Supp. 2d 722, 735 (S.D.N.Y. 2004) ("The seventh factor, quality of defendants' products, asks 'whether the senior user's reputation could be jeopardized by virtue of the fact that the junior user's product is of inferior quality.'" (citation omitted)).

247. The similarity approach to the comparative quality factor posits that if the defendant's goods are of similar quality to or even higher quality than the plaintiff's, then consumers are more likely to confuse the two parties' marks. *See, e.g.*, Landscape Forms, Inc. v. Columbia Cascade Co., 117 F. Supp. 2d 360, 367 (S.D.N.Y. 2000) ("There is also little dispute that the two product lines are of similar quality—a factor also weighing in plaintiff's favor.").

248. *See* Hasbro, Inc. v. Lanard Toys, Ltd., 858 F.2d 70, 78 (2d Cir. 1988). ("The next factor, quality of the junior user's product, is the subject of some confusion. One view is that an inferior quality product produced by the junior user injures the senior user's reputation insofar as consumers might think that the source of the inferior product is the senior user. Another view is that a junior user's product of equal quality to a senior user's product injures the senior user by the increased tendency of similar quality products to promote consumer confusion." (citations omitted)).

is an embarrassment to the multifactor test, and not simply because tarnishment should have no relevance to a finding of fact as to the likelihood of consumer confusion, nor because similarity in quality should already be addressed under the proximity factor, but because the factor is so utterly pliable.

V

Toward a National Standard

There is a special irony in an empirical study of the multifactor test for the likelihood of consumer confusion. The test itself is essentially a substitute for empirical work. Ideally, a court would determine the likelihood of consumer confusion by taking testimony from every consumer who has been or will be exposed to the plaintiff's and defendant's marks. The court would then establish what proportion of this population of consumers is or will be confused, and decide whether that proportion is sufficiently high to justify a ruling in favor of the plaintiff. The court, in other words, would conduct a survey. But because a court lacks the time, resources, and capacity to do so,²⁴⁹ it must instead consider a variety of factors designed to help it estimate the results of that ideal survey. These are the factors of the multifactor test.

When the multifactor tests of the various circuits are held up against the standard of the ideal survey,²⁵⁰ most, if not all of them, are found wanting. Certain of their factors as well as their overall design often distract from their ultimate purpose: to estimate what is actually occurring or will occur in the marketplace. Clearly, considerations such as the comparative quality of the parties' goods or the inherent distinctiveness of the plaintiff's mark rarely aid in this inquiry. In the case of intent, the precise wording of the factor appears to affect courts' analyses. More generally, multifactor tests of ten or even eight factors appear to ask too much of the judge's ability simultaneously to weigh competing concerns and may simply result in

249. *But see* Triangle Publ'ns, Inc. v. Rohrlich, 167 F.2d 969, 974 (2d Cir. 1948) (Frank, J., dissenting). In *Triangle Publications*, involving the trademark SEVENTEEN for the plaintiff's magazine and for the defendant's girdles, Judge Frank took it upon himself to conduct his own survey:

Like the trial judge's, our surmise must here rest on "judicial notice." As neither the trial judge nor any member of this court is (or resembles) a teen-age girl or the mother or sister of such a girl, our judicial notice apparatus will not work well unless we feed it with information directly obtained from "teen-agers" or from their female relatives accustomed to shop for them. Competently to inform ourselves, we should have a staff of investigators like those supplied to administrative agencies. As we have no such staff, I have questioned some adolescent girls and their mothers and sisters, persons I have chosen at random. I have been told uniformly by my questionees that no one could reasonably believe that any relation existed between plaintiff's magazine and defendants' girdles.

Id. at 976 (footnote omitted).

250. *Cf.* Carl Karcher Enters., Inc. v. Stars Rests. Corp., 35 U.S.P.Q.2d 1125, 1132 (T.T.A.B. 1995) ("[T]here is no such thing as a perfect survey.").

the stampeding of less significant factors.²⁵¹ Finally, as this Article has sought to demonstrate, the diversity of tests has made judicial analysis under them less uniform and less predictable.

This Article thus recommends adopting a new national multifactor test, one whose sole purpose should be to aid the judge in estimating the results of an ideal survey of the relevant consumer population. Of course, knowledge of the empirical data will only take us so far in designing such a revised test. Nevertheless, the data do recommend a few general principles. First and most importantly, the basic test should not seek to be exhaustive in its list of possible considerations. Rather, as social science work recommends, the list should consist of a limited number of core factors, ideally no more than three or four. This list may be set forth on an illustrative rather than limitative basis. However, if the history of the multifactor test for trademark infringement teaches us anything, it is that judges, especially circuit court judges, should be wary of adding factors to the list that are only occasionally relevant lest those factors end up being institutionalized. If unlisted factors must be considered in light of the facts of a given case, judges should emphasize the narrow context in which these new factors apply. Second, as the data on the intent factor demonstrated, the precise wording of the factors can be important. The adopted wording should emphasize the empirical rather than formal nature of the inquiry. In trademark law, the question is always of consumer perception in the marketplace rather than judicial perception in the courtroom. Third, the order in which the factors are listed should reflect as much as possible the weight that should be given to them. As such, threshold factors should be listed first. Fourth and finally, if factors are introduced into the test that do not address the empirical question that the test seeks to answer, then that should be stated explicitly. For example, if judges believe that factors such as the defendant's intent or the comparative quality of the goods are worth considering even though they do not go directly to the question of the actual likelihood of consumer confusion, then judges should make that clear.

In light of these principles, a starting point for reform might consist of the following statutory language:

In determining whether a mark is likely to cause confusion, or to cause mistake, or to deceive, the court may consider all relevant factors, including the following: (i) the degree of similarity of the marks as perceived by the relevant consumer population; (ii) the degree of proximity of the goods as perceived by the relevant consumer population, including the degree of proximity of marketing methods and channels of distribution and sale; (iii) evidence of actual confusion, mistake, or deception, including

251. Cf. *Ty, Inc. v. Publ'ns Int'l Ltd.*, 292 F.3d 512, 522 (7th Cir. 2002) (Posner, J.) ("More can be less, even in law.").

survey evidence; (iv) the marketplace strength of the mark allegedly infringed; and (v) the purpose of the alleged infringer in adopting and using its mark and if the purpose is to cause confusion, or to cause mistake, or to deceive, then the likelihood that the alleged infringer will accomplish that purpose.

Admittedly, this list is far from ideal, and violates some of the principles set forth above. First, it proposes more than three or four factors. Nevertheless, the data suggest that, in practice, judges consider each of the five factors proposed to be important to the outcome of the multifactor test. Additionally, the list conflates factors in a way that may complicate the multifactor analysis. Specifically, it forces consumer sophistication considerations into both the similarity and proximity inquiries. The goal here is to emphasize that the multifactor inquiry is an empirical—rather than formal—inquiry that seeks to determine the likely perception of consumers in the marketplace. Finally, though the data suggest that courts place great weight on a finding of bad faith intent, the list proposes the intent factor as the fifth and final factor to be considered. It does so in an effort to limit the impact of this factor on what should remain a tightly-focused fact-finding inquiry into the likelihood of consumer confusion rather than into the commercial morality of the defendant.

Conclusion

Unlike American copyright law and patent law, American trademark has not gone through a postwar phase of reform and remains largely unrationalized.²⁵² This is nowhere more evident than in the basic test for trademark infringement, and it is here, at the doctrinal fulcrum of trademark law, that the reform process should begin. Indeed, it is remarkable that other arguably less centrally important, but more up-to-date areas of U.S. trademark law, such as fame²⁵³ or cybersquatting doctrine,²⁵⁴ enjoy statutorily-prescribed multifactor tests, but the likelihood of confusion test itself continues to be left to the circuit courts.

This Article has sought to present evidence in support of such reform. In the process, it has sought more generally to develop an array of theoretical approaches to the legal multifactor test. Obviously, much more work remains to be done in this regard, not only in developing our theoretical

252. The Lanham Act of 1946, which gave birth to modern American trademark law but failed to codify which factors, if any, should be considered in the likelihood of confusion analysis, was essentially written in the 1930s. See Pattishall, *supra* note, 215 at 196-98; Edward S. Rogers, *The Lanham Act and the Social Function of Trade-Marks*, 14 *Law & Contemp. Probs.* 173, 177-80 (1949).

253. See 15 U.S.C. § 1125(c)(1)(A)-(H) (2006) (prescribing a nonexhaustive eight-factor test to determine whether a mark is “distinctive and famous”).

254. See 15 U.S.C. § 1125(d)(1)(B)(i) (2006) (prescribing a nonexhaustive nine-factor test to determine whether a defendant has a “bad faith intent to profit” from the use of a domain name “identical or confusingly similar” to plaintiff’s trademark).

understanding of the legal multifactor test as applied, but in testing those theories against data on other legal multifactor tests in U.S. and foreign law. Certain questions are of particular importance going forward. First, to what extent do core factors and a core factor heuristic guide other multifactor tests? Specifically, can we predict the outcomes of other multifactor tests based on the outcomes of only one or two factors within the tests, and is a simple classification-tree approach viable? Are threshold factors prevalent among the wide variety of legal multifactor tests? Second, to what extent do other legal multifactor tests tend to stampede, and is stampeding more common or pronounced when judges make a strong decision that intervenes in the status quo? Relatedly, is it appropriate to model judges' application of the legal multifactor tests by means of regression analysis, particularly if judges are employing coherence-based reasoning? Third, as a historical matter, have other multifactor tests as applied decayed into mechanical formalism? Do they tend to accumulate irrelevant factors? Fourth and finally, does a judge's political ideology affect her use of the multifactor test?²⁵⁵ These are not merely academic questions. Good answers will enable us to improve our design and use of multifactor tests.

Though this Article has been critical of many aspects of the multifactor test for the likelihood of confusion, it has also sought to be ameliorist in orientation. One thing it has not done is question the viability or utility of the multifactor form of judicial analysis itself. Elsewhere in the world of legal multifactor tests, judges have expressed dissatisfaction with the multifactor heuristic. Judge Easterbrook has made clear his "reluctan[ce] to accept an approach that calls on the district judge to throw a heap of factors on a table and then slice and dice to taste."²⁵⁶ In criticizing the *Barker v. Wingo*²⁵⁷ criteria for evaluating speedy trial claims, Justice Thomas la-

255. With a view to shedding light on the degree to which political ideology might affect trademark infringement adjudication, the data set for this study included a number of judge-specific variables, such as the judge's gender, age at the time of the opinion, length of tenure at the time of the opinion, and a rough quantification of the judge's political ideology based on the Poole common space score of the judge's appointing president. Poole common space scores place presidents, senators, and representatives on a scale ranging from -1.000 (most liberal) to 1.000 (most conservative) that seeks to be consistent across time and institutions. Senators' and representatives' scores are based on their voting records. See Keith T. Poole & Howard Rosenthal, *Congress: A Political-Economic History of Roll Call Voting* (1997). See also Keith T. Poole, Data Download Front Page, <http://www.voteview.com/dwnl.htm> (last visited Feb. 12, 2006) (President Jimmy Carter's Poole common space score is -.540; President Ronald Reagan's is .525; President George W. Bush's is .538). See generally Gregory C. Sisk & Michael Heise, *Judges and Ideology: Public and Academic Debates About Statistical Measures*, 99 Nw. U. L. Rev. 743 (2005) (discussing debates concerning the utility of the Poole common space score). Logistic regression analysis of the outcome of the multifactor test was performed on these variables and the five core factors from the multifactor test. The results of this analysis suggest that none of the four judge-specific variables, including political ideology, have any significant effect on the overall outcome of the multifactor test.

256. *Reinsurance Co. of Am., Inc. v. Administratia Asigurarilor de Stat*, 902 F.2d 1275, 1283 (7th Cir. 1990) (Easterbrook, J., concurring).

257. 407 U.S. 514 (1972).

mented that the *Barker* “factors now appear to have taken on a life of their own.”²⁵⁸ Another justice has asserted that “multifactor balancing tests generally tend to produce negative answers.”²⁵⁹ Still, multifactor tests appear to be the least worst alternative, if not the only alternative, to a wide-open “totality of the circumstances”²⁶⁰ or “rule of reason”²⁶¹ type of analysis. This is not simply the case in fact-intensive inquiries such as the likelihood of confusion analysis, but also in broadly political inquiries such as those found in constitutional law. Our goal, then, should be to make the best of this situation. This means designing multifactor heuristics not so much in light of judges’ cognitive limitations, but, as the fast and frugal tradition might say, in light of their cognitive ingenuity, their ability, in short, to bring heuristics to heuristics.

Appendix A

The Selection and Coding of the Opinions

A. *Selection of District Court Opinions*

The potential population for the study was all district court opinions that made substantial use of a multifactor test for the likelihood of consumer confusion. In light of the difficulty of compiling and coding the entirety of the population, I chose a recent five-year period, specifically, from January 1, 2000 to December 31, 2004 inclusive, from which to create my sample. By 2000, if not by the mid-1990s, each of the circuits had settled on a standard multifactor test for use in its jurisdiction, which facilitated coding and analysis.

I compiled an initial sample of opinions from the Westlaw and Lexis databases by using search methods and terms designed to capture all opinions included in those databases that made *any* reference to a multifactor test for the likelihood of consumer confusion.²⁶² This approach yielded an

258. *Doggett v. United States*, 505 U.S. 647, 670 (1992) (Thomas, J., dissenting).

259. *Middlesex County Sewerage Auth. v. Nat’l Sea Clammers Ass’n*, 453 U.S. 1, 25 (1981) (Stevens, J., concurring in part and dissenting in part).

260. *See, e.g., Johnny Blastoff, Inc. v. L.A. Rams Football Co.*, 188 F.3d 427, 433 (7th Cir. 1999) (“The determination of whether a party has established protectable rights in a trademark is made on a case by case basis, considering the totality of the circumstances.”). *See also United States v. Mead Corp.*, 533 U.S. 218, 241 (2001) (Scalia, J., dissenting) (“The Court has largely replaced *Chevron*, in other words, with that test most beloved by a court unwilling to be held to rules (and most feared by litigants who want to know what to expect): th’ ol’ ‘totality of the circumstances’ test.”).

261. *See, e.g., Cal. Dental Ass’n v. FTC*, 526 U.S. 756 (1999) (rejecting abbreviated rule-of-reason analysis by FTC of allegedly anticompetitive advertising restrictions).

262. For the district courts of each circuit, the following search terms and connectors were used in Westlaw: (“trademark infringement” & confus!) or ((trademark mark) & “likelihood of confusion”) & da(aft 12/31/1999 & bef 01/01/2005). This search yielded a total of 1236 opinions across all circuits. For the district courts of each circuit, the following search terms and connectors were used in LEXIS: (“trademark infringement” & confus!) or ((trademark mark) & “likelihood of confusion”) and date(geq (01/01/2000) and leq (12/31/2004)). This yielded a total of 1135 opinions across the circuits. Combining the lists yielded a data set of 1252 opinions. For a discussion of the procedure Westlaw uses

initial sample of 1,252 opinions from the five-year period. I then reviewed each of these opinions to determine whether it made *substantial use* of the multifactor test. I defined substantial use liberally as any use beyond the mere citation without analysis of the test. Out of the initial sample of 1,252 opinions, 364 met this criterion.

From this population of 364 opinions, I excluded a small minority of fact patterns that led courts to apply the multifactor test in ways that could skew the results of the study. In most counterfeiting opinions, for example, the likelihood of confusion is very clear and the factors tend to weigh overwhelmingly in favor of the plaintiff.²⁶³ The same is true of opinions involving an alleged breach of a franchising,²⁶⁴ licensing,²⁶⁵ or distribution²⁶⁶ agreement. These opinions were thus excluded from the sample. For similar reasons, I also excluded opinions on motions to dismiss or on motions where the non-moving party failed to appear.²⁶⁷ I retained and noted opinions involving claims of reverse confusion, and fact patterns in which the defendant repackaged plaintiff's goods.

This resulted in a sample of 337 opinions. I excluded the six opinions in which the outcome of the multifactor test was reversed, which yielded a final sample of 331 opinions.

B. Coding of the Opinions

The data set was coded entirely by the author.²⁶⁸ I conducted two rounds of coding. In the first round, I began with the Second, Seventh, and Ninth Circuits, and then proceeded through the remainder of the circuits in numerical order. In the second round, I worked through the circuits in numerical order, both checking my initial round of coding and adding various information that I had neglected to record in the first round. The coding was done directly into an author-designed Microsoft Access form. It was

to select opinions for inclusion in its database, see Kimberly D. Krawiec & Kathryn Zeller, *Common-Law Disclosure Duties and the Sin of Omission: Testing the Meta-Theories*, 91 Va. L. Rev. 1795, 1884 (2005). Of the final sample of 331 opinions, 187 were published in the Federal Supplement. Each of the 331 opinions comprising the final sample was available from both the Westlaw and Lexis databases.

263. See, e.g., *Rolex Watch U.S.A., Inc. v. Jones*, No. 99 Civ. 2359(DLC), 2000 WL 1528263, at *2 (S.D.N.Y. Oct. 13, 2000). See also *Nike, Inc. v. Top Brand Co. Ltd.*, No. 00 Civ.8179 KMW RLE, 2005 WL 1654859, at *6 (S.D.N.Y. July 13, 2005); *Gucci Am., Inc. v. Duty Free Apparel, Ltd.*, 286 F. Supp. 2d 284, 287 (S.D.N.Y. 2003); *Microsoft Corp. v. V3 Solutions, Inc.*, No. 01 C 4693, 2003 WL 22038593, at *12 (N.D. Ill. Aug. 28, 2003); *Calvin Klein Jeanswear Co. v. Tunnel Trading*, No. 98 Civ. 5408(THK), 2001 WL 1456577, at *10 (S.D.N.Y. Nov. 16, 2001).

264. See, e.g., *Ramada Franchise Sys., Inc. v. Boychuk*, 283 F. Supp. 2d 777, 788 (N.D.N.Y. 2003).

265. See, e.g., *Sonista, Inc. v. Hsieh*, 348 F. Supp. 2d 1089, 1094-95 (N.D. Cal. 2004).

266. See, e.g., *Ryan v. Volpone Stamp Co.*, 107 F. Supp. 2d 369, 401-02 (S.D.N.Y. 2000).

267. See, e.g., *Hearts on Fire Co., LLC v. L C Int'l Corp.*, No. 04Civ.2536LTSMH, 2004 WL 1724932, at *3 (S.D.N.Y. July 30, 2004).

268. This is obviously a significant limitation of the data set. See generally Lee Epstein & Gary King, *The Rules of Inference*, 69 U. Chi. L. Rev. 1 (2002).

then converted into an Excel spreadsheet and finally into Stata format. All statistical analysis was done using Intercooled Stata 9.2.

I recorded the caption of the opinion, its citation, its district, the judge who authored the opinion, the date the opinion was filed, and its posture. The posture of the opinion was coded as: (1) preliminary injunction; (2) summary judgment motion by plaintiff; (3) summary judgment motion by defendant; (4) cross-motions for summary judgment; or (5) bench trial.²⁶⁹ I noted those cases in which the plaintiff sought a declaratory judgment of no likelihood of confusion and coded the plaintiff as the defendant and vice-versa. I further recorded whether the opinion involved a word/image mark and/or trade dress, and whether the defendant formulated a defense of parody.

For each factor of the multifactor test, I used dummy variables (a series of binary 1/0 variables) to code the factor as being: (1) found to favor a likelihood of confusion or otherwise not to favor no likelihood of confusion; (2) found to favor no likelihood of confusion or otherwise not to favor a likelihood of confusion; (3) found to be neutral; (4) found to be irrelevant; (5) found to be a fact issue or otherwise premature; (6) not addressed by the court; or (7) unclear.

For certain factors, I coded additional information relevant to that factor. Under the similarity of the marks factor, I noted those cases in which the marks were found to be “identical,” the “same,” or “substantially” the same. I followed the same process under the similarity of the goods factor. Under the actual confusion factor, I coded whether the plaintiff presented survey evidence and whether it was credited by the court, and coded the same for defendant. I further coded whether the opinion explicitly stated that it was too early to expect the plaintiff to produce survey evidence of actual confusion and whether the court explicitly drew an adverse inference from plaintiff’s lack of anecdotal and/or survey evidence of actual confusion. Under the intent factor, in cases in which the court found the factor to favor confusion, I coded whether the court made its determination based on direct evidence of defendant’s intent, inferred intent from the identity of the marks and defendant’s knowledge of plaintiff’s mark, or was unclear in its determination. Under the quality factor, I coded whether the theory underlying the court’s determination was that similar quality facilitated confusion or that the defendant’s lower quality harmed plaintiff or both.

Under the strength factor, I noted where in the *Abercrombie* spectrum the mark was placed (and whether the spectrum was even employed at all), if a “dual test” was explicitly used to determine strength, or if the court

269. I encountered one opinion that addressed a motion for a temporary restraining order that had not been converted into a motion for a preliminary injunction. I coded this opinion as a preliminary injunction opinion. See *Qwest Commc’ns Int’l v. Qwest Commc’ns Corp.*, NO. CO2-829R, 2002 U.S. Dist. LEXIS 25469 (W.D. Wash. 2002).

provided no analysis of trademark strength. Further, with respect to inherent distinctiveness, I noted whether the court found the mark to be inherently strong or weak, found inherent strength to be a fact issue, or did not address inherent strength. I coded the same categories with respect to acquired distinctiveness. Finally, I noted whether the court found third party uses of the mark to disfavor strength, and whether the court found the incontestability of the mark to favor strength.

For purposes of assessing any effect of certain judge-specific factors on the uses and outcome of the test, the data set also included the birth date and gender of the judge, the date the judge received his or her commission, the appointing president, and the Poole common space score²⁷⁰ of the appointing president.

Appendix B

The Unreliability of the Federal Court Cases: Integrated Data Base of the Administrative Office of the U.S. Courts

Researchers have made extensive use of federal court data assembled by the Administrative Office of the United States Courts (AO) and the Federal Judicial Center (FJC).²⁷¹ The AO data set currently contains information about every case filed in federal district court from 1970 to 2004, as well as every appeal filed in the twelve non-specialized federal appellate courts. Legal scholars have long been suspicious of the accuracy of the AO data. AO bankruptcy data in particular have been judged “error ridden”²⁷² and “utterly inadequate for policy purposes.”²⁷³ Legal scholars working in other fields have defended the AO data as at least serviceable depending on the research question and the subtlety of the statistical techniques used. For example, Theodore Eisenberg and Margo Schlanger compared AO data on tort and inmate cases to district court docket sheets.²⁷⁴ They explain that the “implications of our findings depend in part on whether researchers are interested in assessing win rates or award levels.”²⁷⁵ AO data on the former,

270. For a discussion of the Poole common space score, see *supra* note 255.

271. See Inter-university Consortium for Pol. & Soc. Res., Federal Court Cases: Integrated Data Base, 2003, ICPSR Study No. 4026 (2004); Inter-university Consortium for Pol. & Soc. Res., Federal Court Cases: Integrated Data Base, 2002, ICPSR Study No. 4059 (2003); Inter-university Consortium for Pol. & Soc. Res., Federal Court Cases: Integrated Data Base, 2001, ICPSR Study No. 3415 (2002); Inter-university Consortium for Pol. & Soc. Res., Federal Court Cases: Integrated Data Base, 1970-2000, ICPSR Study No. 8429 (last updated 2005). See generally Theodore Eisenberg & Margo Schlanger, *The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis*, 78 Notre Dame L. Rev. 1455 (2003).

272. Jennifer Connors Frasier, *Caught in a Cycle of Neglect: The Accuracy of Bankruptcy Statistics*, 101 Com. L.J. 307, 308 (1996).

273. Teresa A. Sullivan, Elizabeth Warren & Jay Lawrence Westbrook, *The Use of Empirical Data in Formulating Bankruptcy Policy*, 50 Law & Contemp. Probs 195, 210 (1987).

274. See Eisenberg & Schlanger, *supra* note 271.

275. *Id.* at 1488.

they conclude, are “overwhelmingly accurate.”²⁷⁶ On the latter, AO data “substantially overstates mean awards.”²⁷⁷

The circuit-wide bench trial win rate reported in Table 2 diverges substantially from that reported by the AO data. The AO data set, which includes a far more comprehensive sample of trademark litigation than the one used in this study, yields a bench trial win rate in trademark cases for the years 2000 through 2003 of .591 (n=76), while this study’s data set yields a win rate of .487 (n=37) for bench trial opinions sampled from the same four-year period.²⁷⁸ To understand this discrepancy, I sought to locate the AO observations for each of the thirty-seven bench trial opinions I sampled from the four-year period from 2000 to 2003. By matching docket numbers—the publicly available AO data set drops the names of the parties, the judges, and certain other information—I found AO observations for thirty-two of these opinions. Comparing these cases’ opinions and dockets to the AO observations yielded quite troubling results, particularly since important empirical work in intellectual property relies so heavily on the AO data.²⁷⁹

Of the thirty-two cases that ended with a bench trial, only eleven of these cases were classified in the FJC database as being “disposed of” by “court trial.” The following table sets forth, for the thirty-two cases, the AO classification of “[t]he manner in which the case was disposed of” under the AO variable “Disposition”:

FJC “Disposition”	N	%
“Dismissals - want of prosecution”	1	3.1
“Judgment on - motion before trial”	4	12.5
“Judgment on - jury verdict”	2	6.3
“Judgment on - court trial”	11	34.4
“Dismissals - settled”	2	6.3
“Dismissals - other”	4	12.5
“Judgment on - other”	7	21.9
“Unknown”	1	3.1
Total	32	100.1

If the AO variable “Procedural Progress” is used instead of “Disposition,” the data are still shown to be unreliable. The following table sets forth, for the thirty-two cases, the AO classification of “[t]he point to

276. *Id.* at 1489.

277. *Id.*

278. The difference between the two samples in the proportions of bench trials won by plaintiffs is not statistically significant ($z=-1.044$, $p=.296$), but is nevertheless troubling. The difference may partly be explained by the exclusion of counterfeiting, licensing, and similar fact-patterns from my data set. However, nearly all of the opinions excluded on that basis were preliminary injunction opinions. The difference may also be explained if certain bench trials that resulted in judgments for the plaintiff tended not to produce written opinions reported on Westlaw or Lexis, but were nevertheless sampled by the AO data set.

279. See, e.g., Landes, *supra* note 76.

which the case had progressed when it was disposed of” under the AO variable “Procedural Progress”:

FJC “Procedural Progress”	N	%
“Before issue joined-order entered”	2	6.3
“After issued joined - no court action”	3	9.4
“After issue joined - judgment on motion”	3	9.4
“After issue joined - pretrial conference held”	6	18.8
“After issue joined - after court trial”	11	34.4
“After issue joined - after jury trial”	3	9.4
“After issue joined - other”	3	9.4
“Unknown”	1	3.1
Total	32	100.2

The AO data on the outcomes of the cases were also unreliable. The following table cross-tabulates, for the thirty-two cases, the AO variable “Nature of Judgment” with this study’s coding of the outcome of the trademark infringement claim in the reported opinion.