

9-1-1991

Issue 6

North Carolina Law Review

Follow this and additional works at: <http://scholarship.law.unc.edu/nclr>



Part of the [Law Commons](#)

Recommended Citation

North Carolina Law Review, *Issue 6*, 69 N.C. L. REV. 1533 (1991).

Available at: <http://scholarship.law.unc.edu/nclr/vol69/iss6/1>

This Article is brought to you for free and open access by Carolina Law Scholarship Repository. It has been accepted for inclusion in North Carolina Law Review by an authorized administrator of Carolina Law Scholarship Repository. For more information, please contact law_repository@unc.edu.

COMMENT

Equal Protection in Jury Selection? The Implementation of *Batson v. Kentucky* in North Carolina

*We look to the jury box as to a sacred shrine, the place where human justice holds the scales to measure out the dues of man.*¹

For most of the twentieth century, all states have permitted prosecutors to exercise peremptory challenges²—the removal of prospective jurors from the venire “without cause, without explanation and without judicial scrutiny.”³ Since 1868, however, the fourteenth amendment⁴ has forbidden states to discriminate against racial and other protected groups without constitutionally sufficient reasons for treating those groups differently.⁵ The conflict between these two features of the American justice system is inescapable.⁶ By exercising peremptory challenges, the prosecutor can exclude entire groups of individuals from the jury without having to provide any, let alone constitutionally sufficient, justifications. The equal protection clause of the fourteenth amendment, to the contrary, requires that government officials adequately explain actions directed against

1. Brief of Defendant-Appellant at 19, *Byers v. Boice Hardwood Co.*, 201 N.C. 75, 159 S.E. 3 (1931) (No. 651).

2. Although the peremptory challenge is centuries old and embedded in the Anglo-American judicial process, the government's right to exercise peremptory challenges was not firmly established until the beginning of the twentieth century. See J. VAN DYKE, *JURY SELECTION PROCEDURES: OUR UNCERTAIN COMMITMENT TO REPRESENTATIVE PANELS* 150 (1977); Brown, McGuire & Winters, *The Peremptory Challenge as a Manipulative Device in Criminal Trials: Traditional Use or Abuse*, 14 NEW ENG. L. REV. 192, 195 (1978); *infra* notes 26-45 and accompanying text.

3. *Swain v. Alabama*, 380 U.S. 202, 212 (1965).

4. The equal protection clause of the fourteenth amendment provides that “[n]o State shall . . . deny to any person within its jurisdiction the equal protection of the laws.” U.S. CONST. amend. XIV, § 1.

5. See, e.g., J. NOWAK, R. ROTUNDA & J. YOUNG, *CONSTITUTIONAL LAW* § 14.2, at 525 (3d ed. 1986) (equal protection clause ensures that government does not make arbitrary classifications); Tussman & tenBroek, *The Equal Protection of the Laws*, 37 CALIF. L. REV. 341, 365 (1949) (classifications made by government must be reasonably related to the purpose of the law). The equal protection clause originally was intended to protect black persons from discrimination. *Strauder v. West Virginia*, 100 U.S. 303, 306 (1879). Since then, the Supreme Court has held that the fourteenth amendment also applies to nonracial classifications, although the level of its protection varies with the classification. See, e.g., *Cleburne v. Cleburne Living Center, Inc.*, 473 U.S. 432, 446 (1985) (mental capacity); *Craig v. Boren*, 429 U.S. 190, 204 (1976) (gender); *Massachusetts Bd. of Retirement v. Murgia*, 427 U.S. 307, 312 (1976) (age); *Dandridge v. Williams*, 397 U.S. 471, 483-87 (1970) (number of children); *Williamson v. Lee Optical Co.*, 348 U.S. 483, 488-89 (1955) (profession).

6. Jurists and commentators also have argued that the peremptory challenge conflicts with other constitutional provisions. Some have contended that the prosecutor's peremptory removal of specific groups from the jury violates the defendant's sixth amendment right to a jury drawn from a fair cross-section of the community. See, e.g., *McCray v. Abrams*, 750 F.2d 1113, 1131 (2d Cir. 1984); Massaro, *Peremptories or Peers?—Rethinking Sixth Amendment Doctrine, Images, and Procedures*, 64 N.C.L. REV. 501, 541-63 (1986). In 1990, however, the Supreme Court rejected this contention. See *Holland v. Illinois*, 110 S. Ct. 803, 807-11 (1990). At least one commentator has argued that because peremptory challenges are exercised arbitrarily and capriciously, prosecutorial peremptory challenges deprive defendants of their fifth and fourteenth amendment rights to due process of law. See Note, *Due Process Limits on Prosecutorial Peremptory Challenges*, 102 HARV. L. REV. 1013, 1024-33 (1989).

members of protected groups. "The Equal Protection Clause says in essence, 'When the government treats people differently, it has to have a reason.' The peremptory challenge says in essence, 'No, it doesn't.'"⁷

Nevertheless, both the equal protection clause and the peremptory challenge exist to secure fundamental rights. The equal protection clause aims to protect the fundamental human right to freedom from invidious discrimination. The peremptory challenge seeks to ensure the right to an impartial jury.⁸ Consequently, in spite of the ever-present tension between the equal protection clause and the peremptory challenge, the United States Supreme Court has searched for ways to guarantee equal protection to protected groups while preserving the State's right to exercise peremptory challenges. In the landmark case of *Batson v. Kentucky*,⁹ decided in 1986, the Court embarked on its most recent attempt to reconcile the arbitrary nature of the peremptory challenge with the dictates of the equal protection clause. Five years later, the question arises whether lower courts charged with implementing *Batson* have made progress toward the Court's Solomonic goal of ending the discriminatory use of peremptory challenges without abolishing the peremptory challenge entirely. This Comment explores that question as it applies to the appellate courts of North Carolina.

The clash between the peremptory challenge and the equal protection clause is the latest battle in a conflict that racial and ethnic minorities and the states have been waging since the ratification of the fourteenth amendment: the fight over how far the equal protection clause reaches to protect minorities against unconstitutionally motivated exclusion from jury service by the State.¹⁰ At first, states attempted to preclude minorities from being called to jury service altogether.¹¹ After minorities won the right to be included on jury lists, prosecutors began to use peremptory challenges to bar minority persons from sitting on juries.¹² In 1986 the *Batson* Court recognized that the problem of racially motivated peremptory challenges had become widespread.¹³ The Court, however, refused to address the problem by abolishing the peremptory challenge altogether.¹⁴ Instead, the Court attempted to provide defendants with an effective means of proving discrimination. Under *Batson*, the prosecutor must come forward with race-neutral reasons for her peremptory challenges once the de-

7. Alschuler, *The Supreme Court and the Jury: Voir Dire, Peremptory Challenges, and the Review of Jury Verdicts*, 56 U. CHI. L. REV. 153, 203 (1989).

8. The sixth amendment provides, in part, that "[i]n all criminal proceedings, the accused shall enjoy the right to a . . . trial, by an impartial jury." U.S. CONST. amend. VI.

9. 476 U.S. 79 (1986).

10. See, e.g., *Carter v. Jury Comm'n*, 396 U.S. 320, 327 (1970) (substantial underrepresentation of minorities on jury lists); *Avery v. Georgia*, 345 U.S. 559, 561 (1953) (discriminatory formation of jury lists); *Neal v. Delaware*, 103 U.S. 370, 387-88 (1880) (jury service limited to qualified voters; black persons not qualified to vote); *Strauder v. West Virginia*, 100 U.S. 303, 305 (1879) (per se statutory prohibition of jury service by black persons).

11. *Neal*, 103 U.S. at 387; *Strauder*, 100 U.S. at 305.

12. J. VAN DYKE, *supra* note 2, at 150; Kuhn, *Jury Discrimination: The Next Phase*, 41 S. CAL. L. REV. 235, 283 (1968).

13. *Batson*, 476 U.S. at 101 (White, J., concurring).

14. *Id.* at 99 n.22.

fendant makes a threshold showing of an inference of discrimination.¹⁵ The Court left to state and lower federal courts the task of fleshing out the evidentiary framework—determining what evidence gives rise to an inference of discrimination and what proffered reasons rebut such an inference. In effect, having prescribed the relevant constitutional rules, the *Batson* majority challenged the lower courts to take racial prejudice out of jury selection.

Several critics, most notably Justice Marshall, have charged that under *Batson* the lower courts will be unable to end prosecutorial discrimination. Only by prohibiting peremptory challenges entirely, they claim, can minorities be assured freedom from discrimination in jury selection.¹⁶ Though these critics ultimately may be correct that *Batson* cannot achieve its own goals and that a more drastic solution is appropriate, such a conclusion is impossible to reach at this point. Rather than use the decision as a blueprint for eliminating discriminatory jury selection practices, some lower courts have tried to minimize the decision's impact on the operation of the peremptory challenge. As a result, these courts have required defendants to meet unduly high standards of prima facie proof.¹⁷ When defendants have been able to satisfy these standards, the courts routinely have accepted prosecutors' explanations as sufficient to rebut the prima facie case.¹⁸ In sum, lower courts have failed to implement fully the spirit, and in some cases the letter, of *Batson*. Until they do, the debate will continue whether aggressive application of *Batson* might solve the problem of jury selection discrimination without resort to abolition of the peremptory challenge.

Illustrative of the incomplete implementation of *Batson* is the North Carolina experience. Neither the North Carolina Supreme Court nor the North Carolina Court of Appeals ever has held for a defendant on the merits of a *Batson* claim.¹⁹ In part, this result is attributable to the courts' misapprehension of the operation of the *Batson* prima facie evidentiary system.²⁰ In part, it also arises from the courts' undue deference to the findings of trial courts.²¹ As a result of these misapplications of *Batson*, North Carolina has yet to contribute meaningfully to the debate over the decision's continuing validity, much less to achieve its nondiscrimination goal.

This Comment begins with a brief review of the common-law origin and theoretical functions of the peremptory challenge, its introduction into the American judicial system, and its reception by the legislature and courts of

15. *Id.* at 96-97.

16. *E.g., id.* at 102-05 (Marshall, J., concurring); *Chew v. State*, 71 Md. App. 681, 717, 527 A.2d 332, 350 (1987); *Commonwealth v. Hardcastle*, 519 Pa. 236, 261-64, 546 A.2d 1101, 1113-14 (1988) (Nix, C.J., dissenting), *cert. denied*, 110 S. Ct. 1169 (1990); Note, *Batson v. Kentucky: A Half Step in the Right Direction (Racial Discrimination and Peremptory Challenges Under the Heavier Confines of Equal Protection)*, 72 CORNELL L. REV. 1026, 1039 (1987).

17. *See infra* notes 159-201 and accompanying text.

18. *See infra* notes 235-56 and accompanying text.

19. In one case, the supreme court found a procedural error and remanded the proceedings to the trial court for an evidentiary hearing on *Batson* issues. *See State v. Green*, 324 N.C. 238, 240-41, 376 S.E.2d 727, 728 (1989).

20. *See infra* notes 159-89 and accompanying text.

21. *See infra* notes 276-83 and accompanying text.

North Carolina.²² Next, the Comment discusses the constitutional background of *Batson* and reviews the *Batson* decision, paying particular attention to the Court's goals and to the issues it left to state and lower federal courts.²³ The Comment then examines the North Carolina appellate courts' implementation of *Batson* and shows that the courts have been true to neither the letter nor the spirit of the decision. The Comment suggests doctrinal revisions that are better suited to achieving *Batson*'s goals.²⁴ Many of these recommended procedures and standards derive from the implementation of *Batson* in other states. The Comment concludes that the North Carolina courts' unduly constrictive view of *Batson* has rendered the decision ineffective in North Carolina. Therefore, revisions are necessary to guarantee North Carolina criminal defendants and potential jurors equal protection of the laws.²⁵

I. HISTORICAL BACKGROUND: THE ORIGIN AND DEVELOPMENT OF THE PEREMPTORY CHALLENGE

The peremptory challenge has long been a part of the Anglo-American judicial process, although the degree to which it has been approved and employed has varied.²⁶ It developed during the twelfth century, when the English jury was transformed "from a body of fact knowers to one of fact finders."²⁷ The new fact-finding system required jurors to determine the facts based not on personal favoritism, but on evidence presented in court.²⁸ The peremptory challenge evolved as one safeguard²⁹ of jury impartiality.³⁰

In theory, peremptory challenges promote impartiality in three ways. First, attorneys may use them to remove prospective jurors with suspected, but unprovable, prejudice or bias.³¹ Second, peremptory challenges facilitate the exer-

22. See *infra* notes 26-54 and accompanying text.

23. See *infra* notes 55-144 and accompanying text.

24. See *infra* notes 202-18, 261-74, 292, 313-16 and accompanying texts.

25. See *infra* text accompanying note 340.

26. Brown, McGuire & Winters, *supra* note 2, at 193; see Swain v. Alabama, 380 U.S. 202, 212-13 (1965).

27. Gobert, *The Peremptory Challenge—An Obituary*, 1989 CRIM. L. REV. 528, 528; see J. VAN DYKE, *supra* note 2, at 2. Originally, jurors were chosen because they had some knowledge of the facts and the parties. Gobert, *supra*, at 528. During the reign of Henry II (1154-89), the Crown began to impanel men who had knowledge of the case to decide whether a suspect should be charged. J. VAN DYKE, *supra* note 2, at 2. This system, analogous to the modern grand jury, employed the first juries of fact-finders and provided the foundation for the modern jury. *Id.* at 2-3.

28. Gobert, *supra* note 27, at 528.

29. In addition to the peremptory challenge, two practices designed to promote impartial juries developed. First, jurors were selected only from among those citizens thought capable of deciding cases fairly and objectively. *Id.* Professor Alschuler has noted that these persons were an elite group of propertied men. Alschuler, *supra* note 7, at 164-65. Second, the challenge for cause evolved whereby the parties could remove from the panel an unlimited number of jurors with demonstrable bias. Gobert, *supra* note 27, at 528. To this day, to exercise a challenge for cause attorneys must "assign cause" for the challenge; they must explain their reasons for believing the juror is partial and the judge must agree. J. VAN DYKE, *supra* note 2, at 139-41.

30. Gobert, *supra* note 27, at 528.

31. See Swain v. Alabama, 380 U.S. 202, 219 (1965); J. VAN DYKE, *supra* note 2, at 146; Babcock, *Voir Dire: Preserving "Its Wonderful Power"*, 27 STAN. L. REV. 545, 551 (1975). Moreover, the peremptory challenge permits counsel to remove these jurors without embarrassing them

cise of challenges for cause, which are themselves designed to remove biased potential jurors.³² Third, the challenges foster the appearance of impartiality, since the parties may dismiss without explanation jurors they perceive to be biased, whether or not the potential jurors are in fact biased.³³

Originally the Crown enjoyed an unlimited number of peremptory challenges.³⁴ In 1305, recognizing that this practice produced juries that were biased in favor of the prosecution, Parliament abolished the right of prosecutors to exercise peremptory challenges.³⁵ It still, however, viewed peremptory challenges as necessary for defendants; accordingly, defendants retained the right to remove jurors peremptorily.³⁶ Despite Parliament's command that prosecutors challenge potential jurors only for cause, the English courts soon construed the 1305 statute to reinstate, in effect, the prosecutor's right to exercise peremptory challenges.³⁷ The courts permitted the prosecutor to "stand aside" prospective jurors and postpone assigning cause for his challenges.³⁸ The prosecutor had to assign cause only if too few jurors remained to constitute a jury after the defendant exercised his challenges.³⁹ Since this contingency rarely occurred, the Crown effectively enjoyed the right to exercise peremptory challenges.⁴⁰

Although early American courts and legislatures readily accepted the defendant's right to exercise peremptory challenges as part of the common law, the prosecutor's right was more uncertain.⁴¹ Some states permitted prosecutorial peremptory challenges, but severely limited them in number.⁴² Other states did not allow prosecutors to exercise peremptory challenges at all.⁴³ It was not until

with public accusations of bias or with public revelation of their prejudices. *Babcock*, *supra*, at 553-54; *see* Gobert, *supra* note 27, at 529-30.

32. An attorney often can uncover the prejudices of a prospective juror necessary to justify a challenge for cause only by asking the juror probing and personal questions; the availability of peremptory challenges alleviates the attorney's fear of incurring a juror's hostility during this questioning, because the attorney knows that if he alienates a juror, he may strike the juror peremptorily. Thus, the availability of the peremptory challenge prevents a chilling effect on the voir dire questioning that facilitates the challenge for cause. *Swain*, 380 U.S. at 219-20; *Babcock*, *supra* note 31, at 554-55.

33. This satisfies Justice Frankfurter's maxim that "justice must satisfy the appearance of justice." *Offutt v. United States*, 348 U.S. 11, 14 (1954). By furthering the litigants' belief in the system's impartiality, peremptory challenges also promote confidence in and respect for the criminal justice system. *See* *Babcock*, *supra* note 31, at 552; Gobert, *supra* note 27, at 529.

34. *J. VAN DYKE*, *supra* note 2, at 147; Brown, McGuire & Winters, *supra* note 2, at 194; Massaro, *supra* note 6, at 525. The defendant was permitted thirty-five peremptory strikes. 4 W. BLACKSTONE, COMMENTARIES *354.

35. The Ordinance of Inquests, 33 Edw., ch. 4 (1305). Furthermore, the unlimited number of prosecutorial peremptory challenges subjected trials to substantial delays. *See* COKE ON LITTLETON 156 (14th ed. 1791), *quoted in* *Swain*, 380 U.S. at 213.

36. *See* *J. VAN DYKE*, *supra* note 2, at 147; Brown, McGuire & Winters, *supra* note 2, at 194.

37. *See* *Swain*, 380 U.S. at 213; *J. VAN DYKE*, *supra* note 2, at 148; Brown, McGuire & Winters, *supra* note 2, at 194.

38. *Swain*, 380 U.S. at 213; *J. VAN DYKE*, *supra* note 2, at 148.

39. *Swain*, 380 U.S. at 213; *J. VAN DYKE*, *supra* note 2, at 148.

40. *J. VAN DYKE*, *supra* note 2, at 148.

41. *Id.*; Brown, McGuire & Winters, *supra* note 2, at 194; Massaro, *supra* note 6, at 525.

42. *See* *J. VAN DYKE*, *supra* note 2, at 149. As late as 1856, the Supreme Court held that prosecutors in federal court had no common-law right to engage in the practice of "standing aside." *United States v. Shackelford*, 59 U.S. (18 How.) 588, 590 (1856).

43. The two most populous original states, New York and Virginia, did not allow prosecutors

1870 that peremptory challenges for the prosecution became "the rule rather than the exception."⁴⁴ By 1900 "the government's right to exercise peremptory challenges was firmly established."⁴⁵

North Carolina was among the first American jurisdictions officially to recognize the defendant's right to exercise peremptory challenges, the prosecutor's right to "stand aside" prospective jurors, and the prosecutor's right to exercise peremptory challenges.⁴⁶ In 1777 the North Carolina General Assembly codified common-law practice by granting capital defendants thirty-five peremptory challenges.⁴⁷ Then, in 1829, the North Carolina Supreme Court formally approved the practice of "standing aside."⁴⁸ Most dramatically, in 1827 the North Carolina General Assembly broke with five centuries of English practice and became one of the first state legislatures to allow prosecutors to exercise peremptory challenges.⁴⁹

At first the North Carolina prosecutor's privilege existed only in capital cases and was limited to four peremptory strikes,⁵⁰ in sharp contrast to the defendant's thirty-five peremptory challenges.⁵¹ Throughout the twentieth century, the North Carolina General Assembly periodically revised the number of peremptory challenges exercisable by defendants and by the State, narrowing the gap between the two sides.⁵² Finally, in 1977 the general assembly granted the

to exercise peremptory challenges in felony cases until 1858 and 1919 respectively. See Act of April 17, 1858, ch. 332, § 1, 1858 N.Y. Laws 557 (current version at N.Y. CRIM. PROC. LAW § 270.25 (McKinney 1982)); VA. CODE ANN. § 4898 (1919) (statutory revision pursuant to Act of March 8, 1918, ch. 108, 1918 Va. Acts 211) (current version at VA. CODE ANN. § 19.2-262 (1990)).

44. J. VAN DYKE, *supra* note 2, at 150; see Brown, McGuire & Winters, *supra* note 2, at 195.

45. J. VAN DYKE, *supra* note 2, at 150; see Brown, McGuire & Winters, *supra* note 2, at 195. Several developments account for the eventual acceptance of prosecutorial peremptory challenges. By the mid-nineteenth century, the mistrust of government that characterized the Revolutionary period gave way to greater acceptance of state power. J. VAN DYKE, *supra* note 2, at 150. Moreover, in light of the growing heterogeneity in many American cities, the Supreme Court determined that the government had a legitimate interest in exercising peremptory challenges to keep certain "elements" off juries. See *Hayes v. Missouri*, 120 U.S. 68, 70-71 (1887). Eventually the Court found the right to exercise peremptory challenges to be inherent in the right to jury trial. See *Lewis v. United States*, 146 U.S. 370, 376 (1892); Babcock, *supra* note 31, at 556.

46. See *State v. Benton*, 19 N.C. (2 Dev. & Bat.) 196, 204 (1836) (practice of standing aside "has . . . prevailed in the courts of this state"); J. VAN DYKE, *supra* note 2, at 148-49, 171 n.47.

47. Act of Nov. 15, 1777, ch. 2, § 94, J. IREDELL, LAWS OF THE STATE OF NORTH CAROLINA 317 (1791) (current version at N.C. GEN. STAT. § 15A-1217(a)(1) (1988)); see also *State v. Arthur*, 13 N.C. (2 Dev.) 217, 220 (1829) (purpose of peremptory challenge is to ensure that the prisoner has a jury "free from all objection"). The general assembly later extended the right to exercise peremptory challenges to defendants not on trial for their lives. Act of Nov. 16, 1801, ch. 592, § 1, 2 H. POTTER, LAWS OF THE STATE OF NORTH CAROLINA 953 (1821).

48. *Arthur*, 13 N.C. (2 Dev.) at 219.

49. Act of Nov. 19, 1827, ch. 10, 1827 N.C. Sess. Laws 15 (current version at N.C. GEN. STAT. § 15A-1217 (1988)). Only four states authorized their governments to exercise peremptory challenges earlier than North Carolina: Delaware (1782), Pennsylvania (1813), Tennessee (1821), and Georgia (1822). J. VAN DYKE, *supra* note 2, at 171 n.57.

50. 1827 N.C. Sess. Laws 15.

51. *Id.* In addition, the prosecutor had to exercise all of his peremptory challenges before he tendered the prospective jurors to the defendant. *Id.*

52. See, e.g., Act of Feb. 26, 1907, ch. 415, § 1, 1907 N.C. Sess. Laws 608 (current version at N.C. GEN. STAT. § 15A-1217(b)(2) (1988)) (allowing the State two peremptory challenges per defendant in noncapital cases); Act of March 1, 1913, ch. 31, §§ 3, 4, 1913 N.C. Sess. Laws 55, 56 (current version at N.C. GEN. STAT. § 15A-1217 (1988)) (reducing defendants' peremptory challenges to twelve in capital cases and four in noncapital cases); Act of May 11, 1935, ch. 475, §§ 2, 3,

State the same number of peremptory challenges as the defendant.⁵³ In equalizing the number of peremptory challenges allowed, the general assembly brought North Carolina practice in line with that of most states.⁵⁴ Unfortunately, during the same period in which North Carolina prosecutors received this increased power to exercise peremptory challenges, prosecutors around the country began to use their peremptory challenges to deny racial minorities the opportunity to serve on juries.

II. CONSTITUTIONAL BACKGROUND: FROM *STRAUDER* TO *SWAIN*

In the wake of the Civil War, the American people ratified three constitutional amendments whose common purpose was to secure equal rights for newly freed black citizens.⁵⁵ The second of these, the fourteenth amendment, prohibits the states from denying any person "equal protection of the laws."⁵⁶ One of the earliest United States Supreme Court decisions construing the equal protection clause did so in the context of jury selection.

In *Strauder v. West Virginia*⁵⁷ a black man had been charged with murder.⁵⁸ At the time, West Virginia law authorized only white males to serve as jurors.⁵⁹ A state court rejected Strauder's claim that the statute denied him equal protection of the laws; he was convicted by an all-white jury and his ap-

1935 N.C. Sess. Laws 834, 835 (current version at N.C. GEN. STAT. § 15A-1217 (1988)) (increasing defendants' peremptory challenges from 12 to 14 in capital cases and from four to six in noncapital cases; increasing State's peremptory challenges from four to six in capital cases and from two to four for each defendant in noncapital cases); Act of March 11, 1971, ch. 75, § 1, 1971 N.C. Sess. Laws 56 (codified as amended at N.C. GEN. STAT. § 9-21(b) (1971)) (current version at N.C. GEN. STAT. § 15A-1217(a)(2) (1988)) (increasing State's peremptory challenges to nine in capital cases). The general assembly abolished the practice of "standing aside" in 1913. Act of March 1, 1913, ch. 31, § 4, 1913 N.C. Sess. Laws 55, 56 (current version at N.C. GEN. STAT. § 15A-1217 (1988)); see *State v. Ashburn*, 187 N.C. 717, 720-21, 122 S.E. 833, 834 (1924).

53. Act of June 23, 1977, ch. 711, § 1, 1977 N.C. Sess. Laws 853, 858 (codified at N.C. GEN. STAT. § 15A-1217 (1988)). In capital cases today each defendant is entitled to 14 peremptory challenges and the State is entitled to 14 challenges per defendant; in noncapital cases, each defendant may exercise six peremptory challenges and the State may exercise six challenges per defendant. In selecting alternate jurors, each party may use any unused challenges plus one peremptory challenge for each alternate juror to be selected. *Id.*

54. See N.C. GEN. STAT. § 15A-1217 official commentary (1988). One possible explanation for North Carolina's early acceptance of peremptory challenges for both parties in criminal cases is the state's apparent tradition of allowing attorneys freedom in selecting juries. North Carolina trial judges generally allow attorneys wide latitude in questioning jurors during the jury selection process. R. PRICE, NORTH CAROLINA CRIMINAL TRIAL PRACTICE § 18-11, at 289 (1980); see also 2 G. WILSON, NORTH CAROLINA CIVIL PROCEDURE § 47-3, at 121 (1989) ("trial judge must grant sufficient leeway . . . so that counsel can develop meaningful information"). Prior to 1977, attorneys often were able to see the jury list, thus enabling them to shape the jury selection by altering their strategy appropriately. DeMent, *Jury Selection and the Criminal Jury Trial in Superior Court*, in NORTH CAROLINA ACADEMY OF TRIAL LAWYERS, PERSUASION AND THE ART OF ADVOCACY & CRIMINAL PROCEDURE—TRIAL AND APPEAL: NEW G.S. 15A, § 5, at 1 (1978). Indeed, in 1978, seminar material from the North Carolina Academy of Trial Lawyers claimed that "control of jury selection by the litigants is the rule." *Id.* at 4.

55. See *Strauder v. West Virginia*, 100 U.S. 303, 306 (1879); *Slaughter-House Cases*, 83 U.S. (16 Wall.) 36, 71 (1872).

56. The equal protection clause provides in part that "[n]o State shall . . . deny to any person within its jurisdiction the equal protection of the laws." U.S. CONST. amend. XIV, § 1.

57. 100 U.S. 303 (1879).

58. *Id.* at 304.

59. *Id.* at 305.

peal eventually reached the Supreme Court.⁶⁰

The Supreme Court agreed with *Strauder* that the West Virginia statute contravened the equal protection clause. The Court explained that the purpose of the equal protection clause is to end racial discrimination by state governments.⁶¹ It found that the West Virginia statute discriminated against black persons in two ways. First, the statute denied black defendants equal protection by providing that they were to be tried by juries from which members of their race had been purposely excluded, while entitling whites to juries selected from persons of their own race.⁶² The Court held that although no defendant has a right to a jury consisting of members of his race,⁶³ a state may not summarily deprive anyone of that possibility.⁶⁴ Second, the statute denied black prospective jurors equal protection by refusing them the privilege of participating equally in the administration of justice.⁶⁵ *Strauder* thus became a powerful precedent proscribing racial discrimination in jury selection.

Strauder was the first shot fired in a more than one-hundred-year war over the extent of the equal protection clause's reach in regulating jury selection practices. Throughout the century following *Strauder*, the Court consistently reaffirmed the case's fundamental pronouncement outlawing jury selection procedures that discriminate against racial minorities.⁶⁶ Moreover, the Court extended *Strauder*, which had involved a facially discriminatory statute, by forbidding the states from applying facially neutral statutes in a discriminatory manner⁶⁷ and by prohibiting purposeful, substantial underrepresentation of mi-

60. *Id.* at 304.

61. *Id.* at 306-07, 310. The Court wrote:

What is [the meaning of the fourteenth amendment] but declaring that the law in the States shall be the same for the black as for the white; that all persons, whether [black] or white, shall stand equal before the laws of the States, and, in regard to the [black] race, for whose protection the amendment was primarily designed, that no discrimination shall be made against them by law because of their color?

Id. at 307. During the past 100 years, the Court has expanded this interpretation so that the equal protection clause now protects against governmental discrimination on grounds other than race, although the protection exists to different extents depending upon the identity of the protected group. See J. NOWAK, R. ROTUNDA & J. YOUNG, *supra* note 5, at §§ 14.11-25; L. TRIBE, *AMERICAN CONSTITUTIONAL LAW* §§ 16-23 to -31 (2d ed. 1988).

62. *Strauder*, 100 U.S. at 309.

63. *Id.*; accord *Virginia v. Rives*, 100 U.S. 313, 322-23 (1879) (companion case to *Strauder*; mere fact that black defendant was convicted by all-white jury did not warrant reversal when defendant made no showing of racial discrimination); cf. *Taylor v. Louisiana*, 419 U.S. 522, 538 (1975) (defendant not entitled to a jury of any particular composition).

64. See *Strauder*, 100 U.S. at 309.

65. See *id.* at 308. The Court explained:

The very fact that [black] people are singled out and expressly denied by a statute all right to participation in the administration of the law, as jurors, because of their color, though they are citizens, and may be in other respects fully qualified, is practically a brand upon them, affixed by the law, an assertion of their inferiority, and a stimulant to that race prejudice which is an impediment to securing to individuals of the race that equal justice which the law aims to secure to all others.

Id.

66. See, e.g., *Castaneda v. Partida*, 430 U.S. 482, 492-93 (1977); *Carter v. Jury Comm'n*, 396 U.S. 320, 329 (1970); *Whitus v. Georgia*, 385 U.S. 545, 549-50 (1967); *Avery v. Georgia*, 345 U.S. 559, 561 (1953); *Norris v. Alabama*, 294 U.S. 587, 589 (1935).

67. *Hernandez v. Texas*, 347 U.S. 475, 478-79 (1954); see *Avery*, 345 U.S. at 562-63 (although

norities on jury lists.⁶⁸

Prior to 1965 the Court's equal protection-jury selection cases were limited to claims of discrimination in the formation of jury pools.⁶⁹ As the Court struck down states' discriminatory practices in this area and black persons finally were included on jury lists, however, prosecutors found other ways to prevent black persons from sitting on juries.⁷⁰ One method prosecutors used was exercising peremptory challenges at trial to strike black potential jurors from the venire.⁷¹ The Court addressed the constitutionality of this practice for the first time in *Swain v. Alabama*.⁷²

Swain, a black man, was charged with rape.⁷³ At trial he moved to strike the venire and to declare void the all-white jury, contending that both had been chosen in a discriminatory manner. Specifically, Swain claimed that the prosecutor had exercised his peremptory challenges unlawfully against all of the black persons on the venire.⁷⁴ The trial court denied Swain's motions and the Alabama Supreme Court affirmed his conviction and death sentence.⁷⁵

The United States Supreme Court first dismissed Swain's argument that the State intentionally excluded black persons from the jury pool.⁷⁶ The Court then considered his claim that the prosecutor had exercised peremptory challenges purposely to exclude all black persons from the jury in violation of the equal protection clause. In its analysis, the Court recognized the inherent tension between peremptory challenges and the fourteenth amendment. It acknowledged, on the one hand, that settled constitutional principles prohibiting racially discriminatory state jury selection practices apply not only to discrimination in selecting persons for jury service, but also to discrimination in selecting the jurors in a particular case.⁷⁷ On the other hand, the Court noted that peremptory challenges historically have been thought to facilitate the selection of fair and impartial juries—both in fact and as perceived by the parties.⁷⁸ The Court con-

statute was facially neutral, State used different color record cards to identify black persons and to make it easier for jury commissioners to avoid selecting blacks for jury service); *Norris*, 294 U.S. at 589-91 (despite statute permitting black persons to serve as jurors, no black person ever was called to jury service); *Neal v. Delaware*, 103 U.S. 370, 371-73, 394 (1881) (finding equal protection violation when statute authorized only qualified voters to be jurors and state law prohibited black persons from voting).

68. *Castaneda*, 430 U.S. at 493; see *Whitus*, 385 U.S. at 550-51.

69. See, e.g., *Hernandez*, 347 U.S. at 478-79; *Avery*, 345 U.S. at 561; *Patton v. Mississippi*, 332 U.S. 463, 464 (1947); *Norris*, 294 U.S. at 590-93.

70. *J. VAN DYKE*, *supra* note 2, at 150; Kuhn, *supra* note 12, at 283.

71. *J. VAN DYKE*, *supra* note 2, at 150; Kuhn, *supra* note 12, at 283.

72. 380 U.S. 202 (1965).

73. *Id.* at 203.

74. See *id.* at 203, 205-06, 209-10.

75. *Id.* at 203.

76. *Id.* at 209. The Court noted that eight black persons had been on the venire and that Swain failed otherwise to prove purposeful discrimination. *Id.* at 205-09.

77. See *id.* at 224 (holding that the peremptory challenge may not be used to deny black persons the right and opportunity to participate in the administration of justice that white persons enjoy).

78. *Id.* at 212-20. The sixth amendment guarantees that "[i]n all criminal prosecutions, the accused shall enjoy the right to a . . . trial, by an impartial jury." U.S. CONST. amend. VI. For a discussion of the history of the peremptory challenge and its adoption into the American legal sys-

cluded that challenging black prospective jurors on account of race does not necessarily violate the equal protection clause.⁷⁹ Peremptory challenges, the Court reasoned, are necessarily discretionary. To remove all jurors they believe are biased, litigants must be permitted to exercise peremptory challenges on the basis of prospective jurors' looks, gestures, habits, and group affiliations.⁸⁰ This principle holds true particularly in light of the limited knowledge litigants normally have about prospective jurors.⁸¹ Moreover, the Court explained, since all persons are equally subject to being challenged peremptorily, black prospective jurors are disadvantaged by peremptory challenges no more than white prospective jurors.⁸² The Court presumed that prosecutors use peremptory challenges to obtain impartial juries, not because of racial animus.⁸³ Accordingly, it held that a defendant can never prove an equal protection violation solely from the prosecutor's peremptory challenges in his case.⁸⁴ To give meaning to *Strauder* and its progeny, however, the Court added that a defendant may prove an equal protection violation by showing that the "prosecutor . . . in case after case, whatever the circumstances, whatever the crime and whoever the defendant or victim may be, is responsible for" peremptorily striking all qualified black persons.⁸⁵ Such evidence would show that the prosecutor challenged members of the defendant's race not to ensure an impartial jury, but rather to deny them categorically and unconstitutionally the opportunity to serve as jurors.⁸⁶

Swain's standard for proving a fourteenth amendment violation—systematic discrimination by the prosecutor over many cases—soon proved almost impossible for defendants to satisfy. In the twenty years following *Swain*, virtually no defendants successfully challenged prosecutors' uses of peremptory challenges.⁸⁷ The continuing practice of prosecutors peremptorily striking all or most blacks caused the Supreme Court in 1986 to re-examine, in *Batson*, the *Swain* standard of proof.

III. *BATSON V. KENTUCKY*

A. *The Batson Opinions*

James Kirkland Batson, a black man, was charged with second-degree burglary and receipt of stolen goods.⁸⁸ On the first day of his trial in a Kentucky circuit court, the judge excused a number of jurors for cause and then permitted

tem, see *supra* notes 26-54 and accompanying text. For a discussion of how, in theory, the peremptory challenge facilitates the selection of impartial jurors, see *supra* notes 31-33 and accompanying text.

79. *Swain*, 380 U.S. at 221.

80. See *id.* at 220-21.

81. *Id.*

82. See *id.* at 221.

83. *Id.* at 222.

84. See *id.*

85. *Id.* at 223.

86. *Id.* at 224.

87. See *McCray v. Abrams*, 750 F.2d 1113, 1118 (2d Cir. 1984) ("*Swain* has led most courts to reject all constitutional challenges to the prosecutor's alleged discriminatory use of peremptories.>").

88. *Batson*, 476 U.S. at 82.

the attorneys to exercise peremptory challenges.⁸⁹ The prosecutor used his peremptory challenges to strike all four black persons on the venire.⁹⁰ Batson's counsel objected to the prosecutor's actions and moved to discharge the jury.⁹¹ He argued that striking the black persons violated Batson's sixth and fourteenth amendment rights to a jury drawn from a fair cross-section of the community⁹² and his fourteenth amendment right to equal protection of the laws.⁹³ The trial court denied the motion, and Batson was convicted by the all-white jury.⁹⁴ The Kentucky Supreme Court affirmed.⁹⁵ Relying on *Swain*, the court found that Batson had neither alleged nor proved the prosecutor's systematic exclusion of black persons.⁹⁶ The United States Supreme Court granted certiorari and reversed Batson's conviction.⁹⁷

Justice Powell wrote for a seven-justice majority. Once again, the Court reaffirmed the basic constitutional principle that the State may not engage in racial discrimination when selecting juries.⁹⁸ "The harm from discriminatory jury selection," Justice Powell remarked, "extends . . . to touch the entire community" by undermining public confidence in the fairness of the judicial system.⁹⁹ The Court then overruled *Swain's* holding that a defendant can prove unconstitutional discrimination only by showing systematic discrimination over time.¹⁰⁰ The Court explained that the equal protection clause proscribes not only those peremptory challenges that are motivated by racial animus, but also those that are based on the patently erroneous assumption that members of the defendant's race as a group are biased.¹⁰¹ This holding was a stark departure from *Swain*, which had held expressly that group affiliations are constitutionally proper bases for peremptory challenges.¹⁰² Moreover, the *Batson* Court recog-

89. *Id.* at 82-83.

90. *Id.* at 83.

91. *Id.*

92. The sixth amendment provides in pertinent part: "In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed . . ." U.S. CONST. amend. VI. It is applicable to the states through the fourteenth amendment. See *Duncan v. Louisiana*, 391 U.S. 145, 149 (1968). The Supreme Court has construed this clause to guarantee the defendant a jury drawn from a fair cross-section of the community. *Taylor v. Louisiana*, 419 U.S. 522, 528 (1975).

93. *Batson*, 476 U.S. at 83.

94. *Id.*

95. *Id.* at 84.

96. *Id.* Batson did not press his equal protection claim in the Kentucky Supreme Court, apparently conceding that he could not meet the *Swain* standard of systematic exclusion of blacks over time. In analyzing Batson's sixth amendment claim, however, the Kentucky Supreme Court held that the *Swain* standard also applied when the defendant alleged a fair cross-section violation as a result of the prosecutor's peremptory challenges. *Id.* at 83-84.

97. *Id.* at 84.

98. *Id.* Batson never raised an equal protection claim in the Supreme Court. Rather, he relied solely on the sixth amendment right to a jury drawn from a fair cross-section of the community. Nevertheless, the majority declined to address the sixth amendment claim and reversed Batson's conviction exclusively on equal protection grounds. This action spurred a vehement dissent from Chief Justice Burger. See *id.* at 112-18 (Burger, C.J., dissenting).

99. *Id.* at 87.

100. *Id.* at 92-93.

101. See *id.* at 86.

102. See *Swain v. Alabama*, 380 U.S. 202, 221 (1965); *supra* notes 79-83 and accompanying text.

nized that *Swain's* requirement of systematic discrimination over time saddled defendants with a "crippling burden of proof"¹⁰³ while leaving prosecutors' peremptory challenges "largely immune from constitutional scrutiny."¹⁰⁴ The Court held that defendants may prove equal protection violations from the peremptory challenges exercised in their cases alone.¹⁰⁵

Justice Powell then announced the particular manner in which defendants may prove discrimination.¹⁰⁶ Initially, the defendant must establish a *prima facie* case of purposeful discrimination.¹⁰⁷ The *prima facie* case has three elements. First, the defendant must be a member of a cognizable racial group and the prosecutor must have exercised peremptory challenges against members of that group.¹⁰⁸ Next, the defendant is entitled to rely on the fact that peremptory challenges "permit[] 'those to discriminate who are of a mind to discriminate.'"¹⁰⁹ Last, the defendant must show that these facts and any other relevant circumstances raise an inference that the prosecutor used peremptory challenges to remove prospective jurors on account of race.¹¹⁰ Once the defendant establishes a *prima facie* case, the burden shifts to the prosecutor to come forward with "clear and reasonably specific" race-neutral reasons for the peremptory challenges under attack.¹¹¹ This explanation need not rise to a level that would justify a challenge for cause, but it must be more than an assertion of good faith.¹¹² The defendant bears the ultimate burden of persuasion on the issue of the prosecutor's purposeful discrimination.¹¹³

The *Batson* Court instructed that this burden-shifting evidentiary system would operate in the same manner as the evidentiary system used in "disparate treatment" employment discrimination cases under Title VII¹¹⁴ of the Civil Rights Act of 1964.¹¹⁵ Justice Powell already had explained the purpose of such

103. *Batson*, 476 U.S. at 92.

104. *Id.* at 93-94.

105. *Id.* at 95-98; *see id.* at 100-01 (White, J., concurring). Because the Court believed that the peremptory challenge makes an important contribution to the justice system, however, it expressly declined to abolish the challenge. *Id.* at 98-99 & n.22.

106. In formulating the standard of proof, the Court was guided by the general principle that governmental action alleged to be discriminatory must be traced to a racially discriminatory purpose. This requirement emerged from the Court's landmark equal protection cases of the 1970s. *See id.* at 93 (quoting *Washington v. Davis*, 426 U.S. 229, 240 (1976)).

107. *See id.*

108. *Id.* at 96 (citing *Castaneda v. Partida*, 430 U.S. 482, 494 (1977)). Since 1986, the Court has retreated from the position that the defendant and the challenged juror must be of the same race. In 1991, the Court held that a white defendant has standing to allocate peremptory challenges exercised against members of other, constitutionally recognizable racial groups. *Powers v. Ohio*, 111 S. Ct. 1364, 1370, 1373 (1991).

109. *Batson*, 476 U.S. at 96 (quoting *Avery v. Georgia*, 345 U.S. 559, 562 (1953)).

110. *Id.*

111. *Id.* at 97-98 & n.20 (quoting *Texas Dep't of Community Affairs v. Burdine*, 450 U.S. 248, 258 (1981)).

112. *Id.* at 97-98. A prosecutor may not claim simply that in her intuitive judgment the challenged jurors would be partial to the defendant because of their shared race. *Id.* at 97.

113. *Id.* at 94 n.18. Since the trial court's findings will turn largely on credibility evaluations, the Court held, appellate courts should give those findings great deference. *Id.* at 98 n.21; *see infra* notes 277-83 and accompanying text.

114. Civil Rights Act of 1964, tit. VII, 42 U.S.C. §§ 2000e to 2000e-17 (1988).

115. *Batson*, 476 U.S. at 94 n.18 (citing *United States Postal Serv. Bd. of Governors v. Aikens*,

a system in *Texas Department of Community Affairs v. Burdine*:¹¹⁶ it "serves to bring the litigants and the court expeditiously and fairly to [the] ultimate question"¹¹⁷ whether the party in question engaged in illegal discrimination. At issue in discrimination cases is alleged improper intent or motivation. A person's intent is difficult to prove by objective evidence.¹¹⁸ Furthermore, the alleged discriminating party naturally has superior access to proof of his own reasons for the questioned conduct. By shifting the burden of coming forward to the party whose actions are at issue once the complaining party provides threshold prima facie evidence, the burden-shifting system provides persons alleging discrimination with "the kind of detailed discovery that would make it possible for them to prove illicit intent."¹¹⁹ To effectuate this discovery purpose, *Burdine* explicitly held that "[t]he burden of establishing a prima facie case . . . is not onerous."¹²⁰ Likewise, therefore, the prima facie burden under *Batson* is not burdensome.¹²¹

Justice White, the author of *Swain*, concurred in the decision to overrule *Swain*.¹²² That case, he explained, had been a warning that removing black persons on the assumption that they could not judge black defendants fairly would contravene the equal protection clause.¹²³ Since discriminatory peremptory challenges remained widespread despite this warning, Justice White argued that defendants, in appropriate cases, should have an opportunity to inquire into prosecutors' reasons for challenging black potential jurors.¹²⁴

460 U.S. 711, 714-15 (1983); see *Texas Dep't of Community Affairs v. Burdine*, 450 U.S. 248, 252-56 (1981); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802-03 (1973)). Numerous lower courts have interpreted *Batson* as relying on the Court's Title VII cases to explain the operation of the burden-shifting rules. See, e.g., *People v. Harris*, 129 Ill. 2d 123, 177, 184, 544 N.E.2d 357, 381, 384 (1989), cert. denied, 110 S. Ct. 1323 (1990); *Stanley v. State*, 313 Md. 50, 60, 542 A.2d 1267, 1271-72 (1988); *Chisolm v. State*, 529 So. 2d 635, 637 (Miss. 1988); *State v. Antwine*, 743 S.W.2d 51, 63 (Mo. 1987) (en banc), cert. denied, 486 U.S. 1017 (1988); *State v. Porter*, 326 N.C. 489, 497-98, 391 S.E.2d 144, 150 (1990).

116. 450 U.S. 248 (1981).

117. *Id.* at 253.

118. See *Trans World Airlines v. Thurston*, 469 U.S. 111, 121 (1985) (shifting burdens of proof under Title VII designed to assure that "plaintiff [has] his day in court despite the unavailability of direct evidence" of discriminatory intent (quoting *Loeb v. Textron, Inc.*, 600 F.2d 1003, 1014 (1st Cir. 1979))); 2 A. LARSON & L. LARSON, EMPLOYMENT DISCRIMINATION § 50.10, at 10-6 (1990) (employers are "too sophisticated to profess their prejudices on paper or before witnesses"); Blumoff & Lewis, *The Reagan Court and Title VII: A Common-Law Outlook on a Statutory Task*, 69 N.C.L. REV. 1, 9 (1990) (direct evidence of discriminatory intent is often unavailable).

119. Bartholet, *Proof of Discriminatory Intent Under Title VII*: United States Postal Service Board of Governors v. Aikens, 70 CALIF. L. REV. 1201, 1206 (1982); see also 2 A. LARSON & L. LARSON, *supra* note 118, § 50.10, at 10-5 (prima facie case may be established from objective evidence); E. CLEARY, MCCORMICK ON EVIDENCE § 343, at 968 (3d ed. 1984) (one reason for having burden-shifting presumptions is simple fairness, including situations in which opposing party has superior means of access to proof).

120. *Burdine*, 450 U.S. at 253; accord, *Patterson v. McLean Credit Union*, 491 U.S. 164, 186 (1989).

121. *Stanley v. State*, 313 Md. App. 50, 71, 542 A.2d 1267, 1277 (1988), cert. denied, 322 Md. 240, 587 A.2d 247 (1991); Blume, *Racial Discrimination in the State's Use of Peremptory Challenges: The Application of the United States Supreme Court's Decision in Batson v. Kentucky in South Carolina*, 40 S.C.L. REV. 299, 330 (1989).

122. *Batson*, 476 U.S. at 101 (White, J., concurring).

123. *Id.* (White, J., concurring).

124. See *id.* (White, J., concurring).

Justice Marshall's concurrence was not nearly as approving of the majority opinion as Justice White's. Although he did characterize the Court's opinion as an "historic step toward eliminating the shameful practice of racial discrimination in the selection of juries,"¹²⁵ Justice Marshall opined that *Batson's* new standard of proof would not bring about the end of discrimination.¹²⁶ He feared that prosecutors easily could evade *Batson* by proffering pretextual, facially non-racial explanations that the courts could not second-guess.¹²⁷ Stressing that the peremptory challenge is not of constitutional magnitude, Justice Marshall argued that the peremptory challenge should be banned entirely.¹²⁸

In a dissenting opinion joined by Justice Rehnquist, Chief Justice Burger also predicted that *Batson* would not eradicate discrimination in jury selection.¹²⁹ The Chief Justice came to the opposite conclusion from Justice Marshall, however, asserting that the peremptory challenge is essential to assure the appearance of justice.¹³⁰ He argued that since all groups are subject to the peremptory challenge in a given case, striking jurors because of group affiliation does not violate the equal protection clause in the particular case.¹³¹ The Chief Justice thus would have reaffirmed *Swain* and upheld *Batson's* conviction.¹³²

125. *Id.* at 102 (Marshall, J., concurring).

126. *Id.* at 102-05 (Marshall, J., concurring) (predicting that all but the most flagrant violations of the rule would remain unassailable).

127. *Id.* at 105-06 (Marshall, J., concurring).

128. *Id.* at 108 (Marshall, J., concurring). In addition to the concurring opinions by Justices White and Marshall, Justice O'Connor wrote to argue that the decision should not be applied retroactively, *see id.* at 111 (O'Connor, J., concurring), and Justice Stevens wrote to justify the Court's decision to resolve the case on equal protection grounds, even though *Batson* had not argued that issue before the Court, *see id.* at 108-11 (Stevens, J., concurring).

129. The Chief Justice had several other criticisms of the majority's analysis and holding. He doubted that prosecutors exercising peremptory challenges could meet the burden of providing neutral explanations for their action. *Batson*, 476 U.S. at 129 (Burger, C.J., dissenting). He also was skeptical that prosecutors could offer explanations somewhere between that required for a peremptory challenge (no explanation) and that required for a challenge for cause; permitting any inquiry into the basis for a peremptory challenge would "force 'the peremptory challenge [to] collapse into the challenge for cause.'" *Id.* at 127 (Burger, C.J., dissenting) (quoting *United States v. Clark*, 737 F.2d 679, 682 (7th Cir. 1984)).

130. *Id.* at 120 (Burger, C.J., dissenting).

131. *Id.* at 122-23 (Burger, C.J., dissenting).

132. *Id.* at 118-31 (Burger, C.J., dissenting). In a separate dissenting opinion joined by Chief Justice Burger, Justice Rehnquist asserted the importance of the peremptory challenge as a part of the system of trial by jury. He added that "[t]he use of group affiliations, such as age, race, or occupation . . . based on the assumption or belief that members of one group are more likely to favor defendants who belong to the same group, has long been accepted as a legitimate basis for the State's exercise of peremptory challenges." *Id.* at 138 (Rehnquist, J., dissenting). Justice Rehnquist's argument that members of the defendant's race may be assumed to be biased in favor of the defendant, however, even if true, cuts two ways. Consider, for example, the common case with a black defendant and a white victim and prosecutor. Justice Rehnquist's theory states that black members of the jury may be biased for the defendant. Consequently, the prosecutor should be permitted to remove those black jurors peremptorily. Justice Rehnquist's argument necessarily implies, however, that white persons may be assumed to be biased in favor of whites. It follows that the white jurors may be assumed to be biased in favor of the victim and prosecutor, and against the defendant. Since in most cases more white persons will be on the venire than black persons, the defendant will not be able to remove as great a percentage of the whites as the prosecutor will be able to remove of the blacks, and the result will not be an impartial jury, but rather one in which the "subtle group biases of the majority . . . operate, while those of the minority [are] silenced." *Commonwealth v. Soares*, 377 Mass. 461, 488, 387 N.E.2d 499, 516, *cert. denied*, 444 U.S. 881 (1979); Kuhn, *supra* note 12, at

B. Undecided Issues

Batson struck a new balance in the ongoing conflict between the equal protection clause and the peremptory challenge. It made clear that, despite normally being exercised in an arbitrary and capricious manner, the peremptory challenge is subject to the strictures of the fourteenth amendment.¹³³ As a federal appeals court later remarked, the case's core principle is that "a defendant [has] the right to be tried by a jury whose members are selected pursuant to nondiscriminatory criteria."¹³⁴ More specifically, *Batson* sought to provide a more effective means than *Swain* of ensuring that members of racial minorities are not excluded from juries on account of race.¹³⁵

The Court expressly refused, however, to specify the implementation of the *Batson* evidentiary scheme.¹³⁶ Instead, it assigned to state and lower federal

290-91. Even if Justice Rehnquist's assumption were true, therefore, it would operate in favor of restrictions on the right to exercise peremptory challenges on the basis of race.

133. In a footnote, Justice Powell wrote: "The standard we adopt . . . is designed to ensure that a State does not use peremptory challenges to strike any black juror because of . . . race." *Batson*, 476 U.S. at 99 n.22. At least one commentator has criticized the *Batson* Court for seeking "to manifest its symbolic opposition to racial discrimination while doing as little as possible to alter the peremptory challenge." Alschuler, *supra* note 7, at 199. While it is possible that the subjective intention of the *Batson* majority was simply to oppose racial discrimination symbolically, courts must take the Supreme Court at its word when implementing the decision.

134. *United States v. David*, 803 F.2d 1567, 1570 (11th Cir. 1986) (quoting *Batson*, 476 U.S. at 85-86). The court noted that *Batson*'s command is to eliminate, not merely minimize, racial discrimination in jury selection. *Id.* at 1571.

135. *Commonwealth v. McCormick*, 359 Pa. Super. 461, 474, 519 A.2d 442, 449 (1986).

136. *Batson*, 476 U.S. at 99 & n.24. In addition to implementation issues, *Batson* left unanswered a panoply of questions regarding the scope of its application. For example, the decision did not address whether the use of peremptory challenges against minorities may be attacked as a violation of the sixth amendment right to a jury drawn from a fair cross-section of the community. *Batson*, 476 U.S. at 84 n.4. Four years later, however, the Court held that the sixth amendment does not forbid prosecutors from striking jurors on account of race. *Holland v. Illinois*, 110 S. Ct. 803, 811 (1990).

The Court also left unanswered what constitutes a "cognizable group" entitled to *Batson* protection. Lower courts have held that Hispanics are one such group. *See, e.g., State v. Sandoval*, 105 N.M. 696, 700, 736 P.2d 501, 505 (Ct. App. 1987); *State v. Ramos*, 574 A.2d 1213, 1216 (R.I. 1990); *State v. Cantu*, 750 P.2d 591, 596 (Utah 1988). Native Americans are another. *See, e.g., United States v. Roan Eagle*, 867 F.2d 436, 441 (8th Cir.), *cert. denied*, 490 U.S. 1028 (1989); *United States v. Chalan*, 812 F.2d 1302, 1314 (10th Cir. 1987); *State v. Porter*, 326 N.C. 489, 499, 391 S.E.2d 144, 151 (1990). The courts have split, however, on whether women constitute a cognizable *Batson* group. *Compare United States v. De Gross*, 913 F.2d 1417, 1422 (9th Cir. 1990) (prohibiting peremptory challenges on the basis of gender) with *State v. Oliveira*, 534 A.2d 867, 870 (R.I. 1987) (women not a cognizable group). *See also People v. Blunt*, 162 A.D.2d 86, 88-89, 561 N.Y.S.2d 90, 92 (1990) (state constitution's equal protection clause prohibits peremptory challenges on basis of gender); *Sagawa, Batson v. Kentucky: Will It Keep Women on the Jury?*, 3 BERKELEY WOMEN'S L.J. 14, 37-44 (1987-88) (arguing that *Batson* proscribes exclusion of women through peremptory challenges).

The question whether white persons may assert *Batson* claims has arisen in two contexts. One issue is whether white defendants have standing to challenge the striking of black prospective jurors. *Batson*'s language suggests that white persons have no such standing, since one of the elements of the *Batson* prima facie case is that members of the defendant's race were struck. *Batson*, 476 U.S. at 96. In 1991, however, the Court held that the defendant's race is irrelevant to the issue of standing to raise an equal protection claim. *Powers v. Ohio*, 111 S. Ct. 1364, 1370, 1373 (1991). The second issue regarding the standing of white defendants is whether they may attack peremptory challenges exercised against white jurors on account of race. The few courts that have addressed this question have answered it in the affirmative. *See Government of Virgin Islands v. Forte*, 365 F.2d 59, 64 (3d Cir. 1989); *State v. Smith*, 515 So. 2d 149, 150 (Ala. Crim. App. 1987) (dictum).

A hotly debated issue left unanswered by *Batson* is whether the defendant is precluded by the

courts¹³⁷ the task of determining what evidence gives rise to a prima facie inference of discrimination and what prosecutorial explanations rebut a prima facie case.¹³⁸ In effect, having laid down the relevant constitutional guidelines, the Court challenged the lower courts to eradicate discriminatory jury selection practices.

One of the major side effects of *Batson*'s evidentiary standard is that it fundamentally changes the prosecutorial peremptory challenge. By definition, peremptory challenges require no explanation. Thus, when prosecutors must give reasons for their peremptory challenges, those challenges are no longer peremptory. Lower courts implementing *Batson* can handle this consequence in several ways. They can require a low threshold showing by the defendant to raise a prima facie inference of discrimination, in which case prosecutors will have to explain their challenges more often. They also can scrutinize prosecutors' rebuttal explanations closely, causing more prosecutorial challenges to be found unconstitutional. Although the attack on discrimination might be successful, the effect would be to sacrifice the peremptory challenge's peremptory nature. Alternatively, the courts can require a high level of prima facie proof and not scrutinize prosecutors' proffered reasons so closely. Although this approach would protect the peremptory challenge from serious intrusion, it might render *Batson* no more effective than *Swain* in rooting out discrimination. One answer to this apparent dilemma lies in the purpose of *Batson*: the Court specifically stated that it intended to ensure that no state strikes any juror because of race.¹³⁹ Arguably, therefore, lower courts should construe any doubt in favor of combating discrimination, even at the expense of the peremptory challenge. By technically permitting the implementing courts to operate within this range of options, *Batson*'s language gives these courts considerable power to define the extent to which the discriminatory use of peremptory challenges and the peremptory chal-

Constitution from exercising peremptory challenges on the basis of race. In his *Batson* dissent, Chief Justice Burger contended that "[b]etween [the defendant] and the state the scales are to be evenly held." 476 U.S. at 126 (Burger, C.J., dissenting) (quoting *Hayes v. Missouri*, 120 U.S. 68, 70 (1887)). Thus, *Batson* also should apply to the defendant's discriminatory use of peremptory challenges. *Id.* at 125-26 (Burger, C.J., dissenting); see also *Alschuler, supra* note 7, at 197-98 (defendants' peremptory challenges are "state actions" subject to the strictures of the fourteenth amendment); Note, *Discrimination by the Defense: Peremptory Challenges After Batson v. Kentucky*, 88 COLUM. L. REV. 355, 365-68 (1988) (prosecutors should have third-party standing to assert rights of jurors excluded by defendants). In 1990 New York became the first state to prohibit racially motivated peremptory challenges by the defendant. See *People v. Kern*, 75 N.Y.2d 638, 650, 554 N.E.2d 1235, 1241, 555 N.Y.S.2d 647, 653, cert. denied, 111 S. Ct. 77 (1990). Some commentators have argued, however, that the defendant's use of peremptory challenges is not a "state action" subject to the strictures of the fourteenth amendment. Goldwasser, *Limiting a Criminal Defendant's Use of Peremptory Challenges: On Symmetry and the Jury in a Criminal Trial*, 102 HARV. L. REV. 808, 838 (1989); see Note, *Defendant's Discriminatory Use of the Peremptory Challenge After Batson v. Kentucky*, 62 ST. JOHN'S L. REV. 46, 57 (1987).

137. Although the equal protection clause by its terms applies only to actions of state governments, the right to equal protection, and thus *Batson*, applies to actions of the federal government through the fifth amendment. *United States v. Horsley*, 864 F.2d 1543, 1544 n.2 (11th Cir. 1989) (citing *Johnson v. Robison*, 415 U.S. 361, 364 n.4 (1974)).

138. *Batson*, 476 U.S. at 99 & n.24. The Court cited the "variety of jury selection practices" in state and federal trial courts as its reason for not dictating how to implement the decision. *Id.* at 99 n.24.

139. *Id.* at 97.

lence itself survive. In short, the lower courts have become laboratories experimenting with whether and how the equal protection clause and the peremptory challenge can coexist in the American judicial system.

Several critics have echoed Justice Marshall's opinion that the peremptory challenge and the equal protection clause cannot coexist and have called for the abolition of peremptory challenges as the only effective way of ending discriminatory jury selection.¹⁴⁰ They have noted that "a prosecutor can easily assert facially neutral reasons for striking a juror," thus rendering *Batson's* constitutional protections illusory.¹⁴¹ Before a sufficient number of lower courts apply *Batson* rigorously, however, it would be premature to toss out the decision and, with it, the eight-hundred-year old peremptory challenge.¹⁴² As is evident from North Carolina's experience, some states have failed to implement *Batson* aggressively, instead adopting an unduly constricted view of the decision.¹⁴³ As a result, the jury remains out on *Batson*, while defendants and prospective jurors continue to suffer the invidious effects of racially discriminatory jury selection practices.¹⁴⁴

IV. THE IMPLEMENTATION OF *BATSON* IN NORTH CAROLINA

In his concurring opinion, Justice White predicted that "[m]uch litigation [would] be required to spell out the contours of the Court's equal protection holding."¹⁴⁵ In the five years since *Batson*, state and federal courts have heard literally hundreds of *Batson* claims. Nevertheless, no consensus has emerged on how best to implement the decision. As a result, *Batson's* impact has varied from state to state. In North Carolina, neither the supreme court nor the court of appeals ever has found a prosecutor guilty of violating *Batson*.¹⁴⁶ This section examines how North Carolina has implemented the decision and considers why no North Carolina defendant has won a *Batson* claim.

140. See *Chew v. State*, 71 Md. App. 681, 717, 527 A.2d 332, 350 (1987); *Commonwealth v. Hardcastle*, 519 Pa. 236, 261-64, 546 A.2d 1101, 1113-14 (1988) (Nix, C.J., dissenting), cert. denied, 110 S. Ct. 1169 (1990); Note, *supra* note 16, at 1039; see also Brown, McGuire & Winters, *supra* note 2, at 234 (prosecutor's peremptory challenges should be abolished); Massaro, *supra* note 6, at 560-63 (same); Note, *The Use of Peremptory Challenges to Exclude Blacks from Petit Juries in Civil Actions: The Case for Striking Peremptory Strikes*, 4 REV. OF LITIGATION 175, 212-13, 215 (1984) [hereinafter Note, *The Case for Striking Peremptory Strikes*] (peremptory challenges should be abolished in civil cases).

141. *Batson*, 476 U.S. at 106 (Marshall, J., concurring); see *Hardcastle*, 519 Pa. at 261, 546 A.2d at 1113 (Nix, C.J., dissenting); Note, *supra* note 16, at 1036-38.

142. See Note, *Batson v. Kentucky and the Prosecutorial Peremptory Challenge: Arbitrary and Capricious Equal Protection?*, 74 VA. L. REV. 811, 818 (1988).

143. See Blume, *supra* note 121, at 300 (unduly restrictive view of *Batson* adopted in South Carolina).

144. In his concurring opinion, Justice Marshall scolded that "[m]isuse of the peremptory challenge to exclude black jurors has become both common and flagrant." *Batson*, 476 U.S. at 103 (Marshall, J., concurring).

145. *Id.* at 102 (White, J., concurring).

146. See *supra* note 19 and accompanying text.

A. *The Prima Facie Case*

Batson set forth the three elements of the prima facie case. The prosecutor must have peremptorily challenged members of a cognizable racial group. Second, the defendant may rely on the fact that the peremptory challenge is a device susceptible to being used in a discriminatory manner. Third, the defendant must show that these facts and "other relevant circumstances" raise an inference of discrimination.¹⁴⁷ The North Carolina Supreme Court and the North Carolina Court of Appeals have dealt almost exclusively with the third element, and thus this question: What facts and circumstances raise an inference of purposeful discrimination?¹⁴⁸

The North Carolina appellate courts have identified several circumstances that are relevant to this prima facie inquiry. Some of these circumstances tend to support an inference of discrimination. They include a pattern of peremptory challenges against black persons, use of a disproportionate number of challenges against black persons, questions and remarks by the prosecutor during voir dire that suggest racial animus, and the prominence of racial issues in the case.¹⁴⁹ Other circumstances the courts have recognized, in contrast, tend to refute an allegation of discrimination. Those circumstances include the acceptance rate of minority jurors by the State¹⁵⁰ and the ultimate racial composition of the jury.¹⁵¹

Only once, however, has a North Carolina appellate court found the circumstances of a case sufficient to raise a prima facie inference of discrimina-

147. *Batson*, 476 U.S. at 96; see *supra* notes 107-10 and accompanying text.

148. The North Carolina courts have not focused on the first element of the prima facie case because only a few cases have considered *Batson* when the defendant was not black. See *State v. Porter*, 326 N.C. 489, 499, 391 S.E.2d 144, 151 (1990) (Native Americans are racial group cognizable for *Batson* purposes); *State v. Agudelo*, 89 N.C. App. 640, 647-48, 366 S.E.2d 921, 925-26 (denying Hispanic defendant's *Batson* claim without deciding whether Hispanics are cognizable group), *disc. rev. denied*, 323 N.C. 176, 373 S.E.2d 115 (1988). The courts have not addressed the second "element" of the prima facie case because it is merely a proposition upon which the defendant may rely when attempting to show discrimination. It is not a fact to be established, and thus has generated no litigation.

149. See *State v. Smith*, 328 N.C. 99, 121, 400 S.E.2d 712, 724 (1991); *State v. Porter*, 326 N.C. 489, 498, 391 S.E.2d 144, 150-51 (1990); *State v. Crandell*, 322 N.C. 487, 502, 369 S.E.2d 579, 588 (1988); *State v. Jackson*, 322 N.C. 251, 255, 368 S.E.2d 838, 840 (1988), *cert. denied*, 490 U.S. 1110 (1989); *State v. Robbins*, 319 N.C. 465, 490-91, 356 S.E.2d 279, 294-95, *cert. denied*, 484 U.S. 918 (1987).

150. *Smith*, 328 N.C. at 121, 400 S.E.2d at 724. Prior to *Smith* the North Carolina Supreme Court had held, in effect, that any time the State's acceptance rate of black potential jurors was 40% or greater, then no prima facie inference of discrimination arose. See *State v. Allen*, 323 N.C. 208, 219, 372 S.E.2d 855, 862 (1988) (prosecutor accepted 41% of black prospective jurors), *vacated on other grounds*, 110 S. Ct. 1463 (1990); *Crandell*, 322 N.C. at 502, 369 S.E.2d at 588 (prosecutor accepted 50% of black prospective jurors); *State v. Abbott*, 320 N.C. 475, 481-82, 358 S.E.2d 365, 369-70 (1987) (prosecutor accepted 40% of black prospective jurors); *State v. Belton*, 318 N.C. 141, 159, 347 S.E.2d 755, 766 (1986) (prosecutor accepted 50% of black prospective jurors). This rule was inconsistent with *Batson* because it permitted prosecutors to strike peremptorily up to 60% of the minorities tendered virtually without fear of *Batson* consequences. Thus, by recognizing that the acceptance rate of minorities by the State is relevant to, but not dispositive of, the prima facie inquiry, the *Smith* decision brought North Carolina practice in line with the dictates of the *Batson* Court.

151. *Porter*, 326 N.C. at 500, 391 S.E.2d at 152; *Belton*, 318 N.C. at 159, 347 S.E.2d at 766; see *infra* note 159.

tion.¹⁵² In *State v. Smith*,¹⁵³ decided in 1991, the prosecutor struck twelve out of the twenty-one black prospective jurors tendered to him.¹⁵⁴ He used his first three peremptory challenges, six of his first seven, and twelve out of a total of fifteen exercised, to remove black persons.¹⁵⁵ In addition, the case, which involved an interracial killing, was so polluted with racial emotions that its venue had been changed to another county.¹⁵⁶ Most significantly, the prosecutor's remarks during jury selection suggested strongly that his peremptory challenges were racially motivated. The district attorney complained that the defendant had struck several white prospective jurors and then stated:

I submit to the Court that the State, the victim in this case is also entitled to a fair representation of those jurors who are seated there. The victim is white, they ought to have a fair representation as to the number of black/white jurors that are on there, and at the rate that we're going, we'll have—if it's any wish apparently of the defendant, we'll have nine—nine/three or worse.¹⁵⁷

The court held that the pattern of discrimination, the exercise of a disproportionate percentage of the State's challenges against black persons, the racially-charged nature of the case, and the prosecutor's race-conscious remarks together established a prima facie inference of discrimination.¹⁵⁸

The North Carolina appellate courts have found only this one blatant prima facie case to have been established in the five years since *Batson*, primarily because they have misapplied the *Batson* rule by taking into account the voir dire responses of challenged jurors when evaluating the strength of the prima facie inference.¹⁵⁹ *State v. Robbins*¹⁶⁰ illustrates this misapplication. In *Robbins* a

152. In some cases, trial courts have found that the defendant made out a prima facie case. See *infra* note 224.

153. 328 N.C. 99, 400 S.E.2d 712 (1991).

154. *Id.* at 121, 400 S.E.2d at 724.

155. *Id.* at 123, 400 S.E.2d at 725.

156. *Id.* at 122, 400 S.E.2d at 725.

157. *Id.*

158. *Id.* at 123, 400 S.E.2d at 725. The court then found that the prosecutor adequately rebutted the prima facie case, and thus was not guilty of a *Batson* violation. *Id.* at 126-27, 400 S.E.2d at 727-28.

159. Another way in which the North Carolina Supreme Court has misapplied the *Batson* prima facie case is by holding that the racial composition of the impaneled jury mirroring the racial composition of the county in which the trial took place is relevant evidence refuting an allegation of discrimination. See *State v. Belton*, 318 N.C. 141, 159, 347 S.E.2d 755, 766 (1986); see also *Smith*, 328 N.C. at 124, 400 S.E.2d at 726 (ultimate racial makeup of jury relevant to determination that *Batson* not violated); *State v. Porter*, 326 N.C. 489, 500, 391 S.E.2d 144, 152 (1990) (fact that jury mirrored racial composition of county is relevant to determination that *Batson* not violated). The relationship between the racial composition of the jury and the racial composition of the county is relevant only to the question whether the defendant was tried by a jury made up of a cross-section of the community. The right to a jury drawn from a fair cross-section of the community is guaranteed by the sixth amendment. See *Taylor v. Louisiana*, 419 U.S. 522, 528 (1975). *Batson*, however, is grounded solely on the fourteenth amendment equal protection clause, which prohibits any discriminatory peremptory challenge regardless of the racial composition of the eventual jury. Indeed, the Supreme Court has held that the racial composition of the jury is irrelevant to *Batson*. See *Alvarado v. United States*, 110 S. Ct. 2995, 2996 (1990) (per curiam) (reversing decision of court of appeals, which had denied *Batson* claim on ground that the jury chosen satisfied the sixth amendment fair cross-section concept).

160. 319 N.C. 465, 356 S.E.2d 279, cert. denied, 484 U.S. 918 (1987).

black man was charged with first degree murder, robbery with a firearm, and kidnapping.¹⁶¹ Of the seventy-six potential jurors questioned during voir dire, twenty-one were black.¹⁶² The state challenged ten of the black persons for cause because of their unequivocal opposition to the death penalty,¹⁶³ and the trial court excused two blacks on its own motion.¹⁶⁴ The prosecutor then exercised peremptory challenges against seven of the nine remaining black prospective jurors.¹⁶⁵ An all-white jury convicted Robbins and sentenced him to death.¹⁶⁶

On appeal, the North Carolina Supreme Court rejected Robbins's contention that the prosecutor's striking seven black potential jurors established a prima facie inference of discrimination. The court first emphasized that judges must consider "all relevant circumstances" when determining whether prima facie cases exist.¹⁶⁷ The fact that the jury was composed entirely of white persons, the court held, did not necessarily imply discrimination.¹⁶⁸ Moreover, the court noted that the prosecutor had examined all potential jurors in the same manner; nothing in his questions or statements indicated a discriminatory motive.¹⁶⁹ Most significantly, however, the court relied on facts that the challenged jurors themselves had revealed in response to voir dire questioning: three of them had reservations about their ability to impose the death penalty; one was related to some of the defense witnesses; and two had been exposed to some pretrial publicity.¹⁷⁰ The court found that these facts dispelled any possible inference that the prosecutor struck the prospective jurors because of their race.¹⁷¹

Considering the challenged jurors' voir dire responses when deciding whether a prima facie inference of discrimination exists has been a common practice of the North Carolina appellate courts. In every such case, the courts

161. *Id.* at 481, 356 S.E.2d at 289.

162. *Id.* at 491-92, 356 S.E.2d at 295.

163. The racial implications of jury selection are magnified in capital cases for two reasons. First, studies have shown that race is an important factor in the imposition of the death penalty; black defendants who kill white victims have the greatest likelihood of being sentenced to death. *See McCleskey v. Kemp*, 481 U.S. 279, 286-87 (1987). Second, statistically, black persons oppose capital punishment more often than do whites. *See THE DEATH PENALTY IN AMERICA* 85 (H. Bedau 3d ed. 1982); Goldberg, *Toward Expansion of Witherspoon: Capital Scruples, Jury Bias, and Use of Psychological Data to Raise Presumptions in the Law*, 5 HARV. C.R.-C.L. L. REV. 53, 62 (1970). As a result, blacks are at a greater risk of being removed for cause on those grounds. *See Darden v. Wainwright*, 477 U.S. 168, 175 (1986) (absolute moral opposition to death penalty valid grounds for challenge for cause in capital case); *Lockhart v. McCree*, 476 U.S. 162, 173 (1986) (same).

164. *Robbins*, 319 N.C. at 492, 356 S.E.2d at 295.

165. *Id.* The defendant peremptorily challenged one black potential juror; the other black person sat as an alternate juror in the case. *Id.*

166. *Id.* at 491, 356 S.E.2d at 295.

167. *Id.* at 489, 356 S.E.2d at 293 (quoting *Batson*, 476 U.S. at 96-97).

168. *Id.* at 494-95, 356 S.E.2d at 297.

169. *Id.* at 493-94, 356 S.E.2d at 296.

170. *See id.* The courts cited similar voir dire responses in rejecting prima facie cases in *State v. Davis*, 325 N.C. 607, 619, 386 S.E.2d 418, 424 (1989) (potential juror held reservations about imposing death penalty), *cert. denied*, 110 S. Ct. 2587 (1990); *State v. Aytche*, 98 N.C. App. 358, 365, 391 S.E.2d 43, 47 (1990) (potential juror previously convicted of manslaughter); and *State v. Batts*, 93 N.C. App. 404, 409, 378 S.E.2d 211, 213 (1989) (potential juror knew members of defendant's family and attended school with defendant's brother).

171. *Robbins*, 319 N.C. at 494, 356 S.E.2d at 296.

have held that the facts disclosed by the jurors refuted any possible inference of prosecutorial discrimination. Thus, for example, no inference of discrimination has arisen when voir dire questioning revealed that the challenged black jurors previously had been convicted for nonsupport and writing bad checks;¹⁷² that the potential juror had served on a jury within the previous four years;¹⁷³ that she had worked with youth who had drug and alcohol problems;¹⁷⁴ that she had a "hard look on her face";¹⁷⁵ and, in a case in which the defendant was twenty years old, that the juror had three grown children.¹⁷⁶

By factoring in the challenged jurors' voir dire responses, North Carolina courts have confused facts that are relevant only to the prosecutor's rebuttal with facts that are relevant to the prima facie case. For example, in determining whether an inference of discrimination arose in *Robbins*, the court took into account statements by one challenged juror who had said she was related to some of the defense witnesses, two who had said they had been exposed to pre-trial publicity, and one who had expressed reservations about imposing the death penalty.¹⁷⁷ These are not circumstances surrounding the disputed peremptory challenges, but are merely possible reasons why the prosecutor might have challenged the jurors. The distinction between the circumstances surrounding the prosecutor's peremptory challenges and the prosecutor's possible reasons for those challenges is crucial because, according to *Batson*, courts are supposed to consider only the circumstances surrounding the challenges at the prima facie stage; the prosecutor is supposed to proffer her reasons for the challenges once the defendant has established a prima facie case.¹⁷⁸

The circumstances surrounding the prosecutor's peremptory challenges consist of the prosecutor's handling of the jury selection, and the conditions that characterize the jury selection and the case itself. Specifically, they include the manner in which the prosecutor questioned prospective jurors, the racial composition of the jury, the pattern of the prosecutor's peremptory challenges, and the

172. *Batts*, 93 N.C. App. at 409, 378 S.E.2d at 213.

173. *Davis*, 325 N.C. at 619, 386 S.E.2d at 424.

174. *Id.*

175. *State v. Melvin*, 99 N.C. App. 16, 30, 392 S.E.2d 740, 748 (1990).

176. *Davis*, 325 N.C. at 619, 386 S.E.2d at 424.

177. *State v. Robbins*, 319 N.C. 465, 492, 356 S.E.2d 279, 295, *cert. denied*, 484 U.S. 918 (1987); see *supra* notes 170-71 and accompanying text.

178. In *Tolbert v. State*, 315 Md. 13, 553 A.2d 228 (1989), the Maryland Court of Appeals stated: "What reasons a prosecutor may advance for his challenges are not relevant to a prima facie showing *vel non*. It is the 'circumstances' concerning the prosecutor's use of peremptory challenges which may create a prima facie case of discrimination against black jurors . . ." *Id.* at 18, 533 A.2d at 230. See also *People v. Trevino*, 39 Cal. 3d 667, 692 n.26, 704 P.2d 719, 733 n.26, 217 Cal. Rptr. 652, 666 n.26 (1985) (en banc) ("It is neither the function nor the duty of the trial courts, or the appellate courts on review, to speculate as to prosecutorial motivation . . ."); *People v. Harris*, 129 Ill. 2d 123, 184, 544 N.E.2d 357, 384 (1989) ("court should not presume, or infer from the facts of the case, that an unarticulated neutral explanation exists"), *cert. denied*, 110 S. Ct. 1323 (1990); Raphael, *Discriminatory Jury Selection: Lower Court Implementation of Batson v. Kentucky*, 25 WILLAMETTE L. REV. 293, 315 (1989) ("Most courts consider the prosecutor's reasons for challenges only after a prima facie case has been established."). This is not to say that prosecutors' possible reasons for exercising the questioned peremptory challenges are irrelevant to whether the prosecutors violated *Batson*; indeed, *Batson* violations ultimately turn on the prosecutors' motivation for striking the prospective jurors. Rather, the argument is merely that such reasons are irrelevant to determining whether a prima facie case of *Batson* discrimination exists.

racial overtones of the case.¹⁷⁹ The potential reasons for the prosecutor's challenges, in contrast, are the substantive statements that prospective jurors reveal during voir dire, and any characteristics of the prospective jurors that might cause the prosecutor to strike the jurors.¹⁸⁰ They might include, for example, the fact that the juror previously had been a criminal defendant or that the juror is related to the defendant.

A hypothetical situation illustrates further the distinction between "circumstances" and "reasons." Suppose ten black prospective jurors are on the venire. Assume further that nothing in the prosecutor's questions or statements during jury selection suggests racial animus; the prosecutor, however, exercises peremptory challenges against all ten black jurors. These events and conditions are the circumstances surrounding the prosecutor's peremptory challenges. The prosecutor's apparently evenhanded conduct during the voir dire is a circumstance weighing against an inference of discrimination; contrarily, the overwhelming pattern of challenges against black jurors is a circumstance weighing in favor of an inference of discrimination. If the trial judge finds an inference of discrimination, under *Batson* the prosecutor then must show that the peremptory challenges were not racially motivated. To do that, the prosecutor might explain that she removed the jurors not because of their race, but because, in response to voir dire questioning, the jurors indicated that they were related to the defendant, had reservations about the death penalty, had been exposed to pretrial publicity, or the like. These responses would be the prosecutor's reasons for the challenges.

Taking into account at the prima facie stage what would normally be the prosecutor's rebuttal reasons is contrary to the letter of *Batson* and seriously compromises *Batson*'s ability to provide defendants with an effective means of proving discrimination. The language of *Batson* specifically commands that prosecutors provide legitimate reasons for their peremptory challenges once defendants establish a prima facie case.¹⁸¹ If the lower courts consider prosecutors' potential reasons before the prosecutors state those reasons themselves, this command becomes meaningless.

More importantly, taking prosecutors' anticipated reasons into account at the prima facie stage prevents the proper functioning of the *Batson* prima facie case.¹⁸² As noted above, *Batson*'s burden-shifting evidentiary system requires

179. The "relevant circumstances" that the *Robbins* court listed as examples all fit this description. See *Robbins*, 319 N.C. at 490-91, 356 S.E.2d at 294-95 (pattern of strikes against blacks; striking of disproportionate number of blacks; questions and remarks by prosecutor; fact that victim and defendant are of different races; racial issues bound up with the conduct of the trial). For a listing of other "circumstances," see *infra* notes 190-201 and accompanying text.

180. See *infra* notes 235-51 and accompanying text. Indeed, the purpose of questioning jurors is to alert the attorneys to facts that might cause the attorneys to challenge the juror.

181. *Batson*, 476 U.S. at 97.

182. The term "prima facie case" has several connotations. In one sense, it is the burden on the party asserting a claim or affirmative defense to produce enough evidence to permit the jury to decide the issue. 9 J. WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 2494, at 379 (Chadbourn rev. 1981). The standard of proof for this type of prima facie case is evidence sufficient to permit a reasonable juror to find the material facts asserted. E. CLEARY, *supra* note 119, § 338, at 953. In another context, "prima facie case" denotes the establishment of a rebuttable presumption

that prosecutors explain their allegedly discriminatory actions upon a mere threshold showing by the defendant. It does not require the defendant to establish her prima facie case with evidence of actual discriminatory intent on the prosecutor's part.¹⁸³ The purpose of this system is to provide the defendant with the hard-to-obtain information she needs to prove a discriminatory prosecutor's illicit intent.¹⁸⁴ To effectuate this purpose, the prima facie burden is not onerous.¹⁸⁵ In fact, a Pennsylvania court has held expressly that "in a close case it is prudent for a court to err on the side of finding a prima facie case and requiring a neutral explanation."¹⁸⁶ The North Carolina courts' practice of considering prosecutors' potential rebuttal explanations at the prima facie stage is inconsistent with this type of prima facie case in several respects. Most obviously, it makes establishing prima facie cases much more difficult than if the court does not take those facts into account. The potential reasons for prosecutors' peremptory challenges are facts that necessarily weigh against an inference of discrimination. Accordingly, when courts consider those facts at the prima facie stage, defendants must produce more affirmative evidence of discrimination to overcome the suggestion of nondiscrimination that the possible reasons evoke.¹⁸⁷ Furthermore, *Batson* proscribes all peremptory challenges that are actually racially motivated—not merely those for which no possible legitimate explanation exists. When a court rejects a prima facie case on the basis of the prosecutor's potential reasons for the disputed challenges, however, the court upholds the prosecutor's peremptory challenges based not on what actually motivated the prosecutor, but on what the court believes the prosecutor reasonably could have (and indeed would have) asserted as the reasons for the challenges, had he been asked. No one knows whether the prosecutor would have offered those, or any, credible reasons for the challenges in question.¹⁸⁸ Finally, even if

that requires the opposing party to come forward with some answer to the prima facie case. 9 J. WIGMORE, *supra*, § 2494, at 379. "Prima facie case" in the *Batson* context is the latter type. *Cf.* Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248, 254 n.7 (1981) ("prima facie case" in Title VII cases denotes rebuttable presumption).

183. See *supra* notes 107-21 and accompanying text.

184. See *supra* notes 116-19 and accompanying text.

185. See, e.g., *United States v. Chalan*, 812 F.2d 1302, 1314 (10th Cir. 1987) (striking single juror may be sufficient); *State v. Slappy*, 522 So. 2d 18, 22 (Fla.) (any doubt whether defendant has met his initial burden should be resolved in his favor), *cert. denied*, 487 U.S. 1219 (1988); *Stanley v. State*, 313 Md. App. 50, 71, 542 A.2d 1267, 1277 (1988) (*Batson* prima facie showing threshold is "not an extremely high one—not an onerous burden to establish"), *cert. denied*, 322 Md. 240, 587 A.2d 247 (1991); see also *Blume, supra* note 121, at 330 (prima facie case may be established on minimal evidence in order to allow reasonable inquiry into defendant's claims). The lower prima facie threshold does not materially disadvantage the prosecutor, because the weaker prima facie case presumably is easier to rebut. See *Gamble v. State*, 257 Ga. 325, 327, 357 S.E.2d 792, 795 (1987).

186. *Commonwealth v. Jackson*, 386 Pa. Super. 29, 52, 562 A.2d 338, 349 (1989), *appeal denied*, 525 Pa. 631, 578 A.2d 926 (1990).

187. This result is antithetical to the burden-shifting evidentiary system employed in *Batson*, which specifically recognizes and attempts to compensate for objective evidence of discriminatory intentions being difficult to obtain. See 2 A. LARSON & L. LARSON, *supra* note 118, § 50.10, at 10-6 ("Employers are . . . too sophisticated to profess their prejudices on paper or before witnesses."); E. CLEARY, *supra* note 119, § 343, at 968 (burden-shifting presumptions may be used when opposing party has superior access to the proof).

188. For example, it is possible that, if required to explain their peremptory challenges, prosecutors who strike jurors on account of race would not say that the juror had been exposed to pretrial publicity or was connected somehow to the defendant. Rather, the prosecutor might offer some

the prosecutor would have proffered the facts revealed during the voir dire as the reasons for the peremptory challenges, considering those "reasons" without requiring the prosecutor to state them deprives the trial court of the opportunity to evaluate the prosecutor's credibility.¹⁸⁹ Thus, courts cannot possibly evaluate whether the facts revealed by the challenged jurors genuinely motivated the prosecutor's peremptory challenges or whether the prosecutor would have asserted them merely as pretexts for discrimination.

Numerous circumstances are relevant to the question whether an inference of discrimination exists, but do not constitute possible reasons for disputed peremptory challenges. The North Carolina Supreme Court already has recognized many of them.¹⁹⁰ These circumstances include the disproportionate removal of minorities;¹⁹¹ the nature of the crime;¹⁹² the presence of racial issues in the case;¹⁹³ disparate treatment of prospective jurors who are similar in all relevant respects except race;¹⁹⁴ and whether the prosecutor failed to question minority jurors, questioned them only perfunctorily, or questioned them differently from nonminorities.¹⁹⁵ Not all of the circumstances that are relevant to the inference of discrimination tend to prove discrimination, however. The manner in which the prosecutor conducts jury selection, for instance, is a circumstance that may weigh against an inference of discrimination.¹⁹⁶

Illustrative of circumstances that tend to negate an inference of discrimination is *State v. Davis*.¹⁹⁷ In *Davis* the potential jurors entered the courtroom separately for voir dire and the attorneys examined each one individually.¹⁹⁸ As a result, the prosecutor did not know how many minority persons were on the

other explanation that more plainly would be a pretext for discrimination. Under North Carolina practice this situation is not likely to get past the prima facie stage; the trial judge would be permitted to assume incorrectly that the prosecutor struck the juror because of the exposure to pretrial publicity or the connection to the defendant and not because of race.

189. This practice is inconsistent with *Batson*, which noted that "the trial judge's findings . . . largely will turn on evaluation of credibility." *Batson*, 476 U.S. at 98 n.21.

190. See *supra* note 149 and accompanying text.

191. See, e.g., *Ex parte Branch*, 526 So. 2d 609, 622 (Ala. 1987); *Commonwealth v. Soares*, 377 Mass. 461, 490, 387 N.E.2d 499, 517, *cert. denied*, 444 U.S. 881 (1979); *State v. Robbins*, 319 N.C. 465, 490-91, 356 S.E.2d 279, 294, *cert. denied*, 484 U.S. 918 (1987).

192. *United States v. Clemons*, 843 F.2d 741, 748 (3d Cir.), *cert. denied*, 488 U.S. 835 (1988).

193. *Id.* (race of the defendant and the victim are relevant circumstances).

194. *United States v. Horsley*, 864 F.2d 1543, 1546 (11th Cir. 1989); *Branch*, 526 So. 2d at 623.

195. *Branch*, 526 So. 2d at 623; *People v. Wheeler*, 22 Cal. 3d 258, 282, 583 P.2d 748, 764, 148 Cal. Rptr. 890, 906 (1978); *Slappy v. State*, 503 So. 2d 350, 355 (Fla. Dist. Ct. App. 1987), *aff'd*, 522 So. 2d 18 (Fla.), *cert. denied*, 487 U.S. 1219 (1988).

Other examples of circumstances relevant to the prima facie case include the following: the fact that the only characteristic the challenged jurors shared was race, see *Wheeler*, 22 Cal. 3d at 280, 583 P.2d at 764, 148 Cal. Rptr. at 905; *Jackson v. Commonwealth*, 8 Va. App. 176, 184, 380 S.E.2d 1, 5, *aff'd on rehearing*, 9 Va. App. 169, 384 S.E.2d 343 (1989) (en banc); Note, *supra* note 142, at 824; and the fact that the challenged jurors had characteristics or a background ordinarily thought favorable to the prosecution, such as having been a crime victim or a police officer, see *People v. Turner*, 42 Cal. 3d 711, 719, 726 P.2d 102, 106, 230 Cal. Rptr. 656, 660 (1986).

196. See *supra* note 150 and accompanying text.

197. 325 N.C. 607, 386 S.E.2d 418 (1989), *cert. denied*, 110 S. Ct. 2587 (1990).

198. *Id.* at 620, 386 S.E.2d at 424. Under North Carolina jury selection procedures, the trial judge in a capital case may permit prospective jurors to be sequestered before and after selection. N.C. GEN. STAT. § 15A-1214(j) (1988); R. PRICE, *supra* note 54, § 18-4, at 283.

venire, or whether the next persons called would be minorities.¹⁹⁹ Three of the first four jurors seated were black.²⁰⁰ Because the prosecutor accepted the three black jurors at the beginning of the selection process when he did not know how many blacks remained in the jury pool, the North Carolina Supreme Court found that the circumstances suggested that the black persons the prosecutor did remove peremptorily were not struck on account of their race.²⁰¹ The court properly found no prima facie case, because the circumstances surrounding the prosecutor's peremptory challenges substantially refuted the defendant's claim of discrimination. By considering only prima facie stage circumstances such as these, and not possible explanations for the prosecutor's peremptory challenges, the courts can adhere to *Batson's* letter and subserve, rather than subvert, the purposes of the *Batson* prima facie case.

Given that the North Carolina courts should not consider the prosecutor's potential reasons for disputed peremptory challenges at the prima facie stage, the question remains: What circumstances are sufficient to raise a prima facie inference of discrimination? Ordinarily, courts will weigh the totality of the relevant circumstances to determine if a prima facie inference exists. The North Carolina courts, however, reasonably can promote *Batson's* goal of easing the evidentiary burden on defendants trying to prove discrimination by focusing on one question in particular: whether prosecutors exercised peremptory challenges disproportionately against minority prospective jurors. In *Smith* and *Robbins* the supreme court acknowledged that disproportionate strikes against minorities were relevant to a prima facie inference of discrimination.²⁰² Like most appellate courts, however, the North Carolina Supreme Court has not indicated how trial courts should determine disproportionate effect, and how much weight the trial courts should give it when evaluating the prima facie case.

An excellent illustration of the application of a disproportionate effects test is the Massachusetts case of *Commonwealth v. Soares*.²⁰³ In *Soares* the prosecutor exercised peremptory challenges against thirty-two white prospective jurors, thirty-four percent of the white persons available.²⁰⁴ The prosecutor challenged

199. *Davis*, 325 N.C. at 620, 386 S.E.2d at 424.

200. *Id.*

201. *Id.*

202. *State v. Robbins*, 319 N.C. 465, 490, 356 S.E.2d 279, 294, cert. denied, 484 U.S. 918 (1987); *accord* *United States v. Johnson*, 873 F.2d 1137, 1140 (8th Cir. 1989), cert. denied, 111 S. Ct. 304 (1990); *Ex parte Branch*, 526 So. 2d 609, 623 (Ala. 1987); *Ward v. State*, 293 Ark. 88, 92, 733 S.W.2d 728, 730 (1987); *Commonwealth v. Jackson*, 386 Pa. Super. 29, 49 n.5, 562 A.2d 338, 348 n.5 (1989), appeal denied, 525 Pa. 631, 578 A.2d 926 (1990); *Williams v. State*, 712 S.W.2d 835, 841 (Tex. Crim. App. 1986).

203. 377 Mass. 461, 387 N.E.2d 499, cert. denied, 444 U.S. 881 (1979). Massachusetts was one of a handful of states that, prior to *Batson*, had prohibited the discriminatory use of peremptory challenges on state constitutional grounds. The evidentiary methodology these states employed to prove state constitutional violations was substantially the same as that employed in *Batson*. See *id.* at 486-88, 387 N.E.2d at 516-17; *People v. Wheeler*, 22 Cal. 3d 258, 276-78, 583 P.2d 748, 761-63, 148 Cal. Rptr. 890, 903-04 (1978); *State v. Neil*, 457 So. 2d 481, 486 (Fla. 1984); *State v. Crespin*, 94 N.M. 486, 488, 612 P.2d 716, 718 (Ct. App. 1980); *People v. Thompson*, 79 A.D.2d 87, 106-10, 435 N.Y.S.2d 739, 752-55 (1981).

204. *Soares*, 377 Mass. at 473 & n.7, 387 N.E.2d at 508 & n.7.

only twelve black prospective jurors.²⁰⁵ Because only thirteen black persons were available on the venire, however, this number amounted to ninety-two percent of the available blacks.²⁰⁶ Thus, although the prosecutor struck more than two-and-one-half times as many whites as blacks (thirty-two as compared to twelve), the impact of the peremptory challenges was two-and-one-half times as great upon the black prospective jurors as upon the whites (ninety-two percent removed as compared to thirty-four percent removed). The court held that the prosecutor's use of peremptory challenges had a disproportionate impact on the black potential jurors.²⁰⁷ It concluded that the disparity suggested a possible discriminatory motivation and thus justified an inquiry into the prosecutor's reasons for his challenges.²⁰⁸

The *Soares* test, comparing the percentage of minorities removed, with the percentage of nonminorities removed, is useful for many reasons. First, it is highly relevant; whether the prosecutor uses peremptory challenges to remove a significantly greater percentage of minorities than nonminorities has a direct logical bearing on the strength of an inference that race motivated the use of those challenges. Second, it possesses a built-in control against the danger of pure statistics carrying too much weight in the *prima facie* determination. If the disparity between the percentages of blacks and whites challenged is striking, then statistics will play a greater role; if the disparity is small, statistics will have little probative value on the issue of discrimination.²⁰⁹ Third, the test can be applied in every case. Consequently, it reduces the possibility of arbitrary decisions, since the supreme court routinely may impose it as a limitation on the trial court's otherwise nearly unfettered discretion to decide what circumstances are relevant to the *prima facie* case. Fourth, it is an objective test. As a result, it is amenable to appellate review.²¹⁰ Fifth, it helps guard against discriminatory prosecutors who leave a few minority jurors on the jury to insulate their other, race-based strikes.²¹¹ Most important, the disproportionate effects test comports with the purpose and function of the *Batson* *prima facie* case. The test is not onerous. It permits the defendant and the court to draw preliminary inferences of discrimination from the information available to the defendant. It thus facilitates the exposure of the prosecutor's illicit motives by providing enough evidence to justify requiring the prosecutor to explain her reasons for the dis-

205. *Id.* at 473, 387 N.E.2d at 508.

206. *Id.*

207. *Id.* at 490, 387 N.E.2d at 517.

208. *Id.*; see also *Gamble v. State*, 257 Ga. 325, 326, 357 S.E.2d 792, 794 (1987) (disparity of 23.8 percentage points between proportion of blacks on jury panel and proportion of blacks on jury sufficient to show discriminatory effect).

209. Some courts have noted the significance of statistics in jury selection discrimination cases. See *Aldridge v. State*, 258 Ga. 75, 78-79, 365 S.E.2d 111, 114 (1988); cf. *McCleskey v. Kemp*, 481 U.S. 279, 294-95 (1987) (dictum) (statistics have greater role in jury selection discrimination cases than in cases regarding the discriminatory impact of the death penalty).

210. See *infra* note 292 and accompanying text.

211. Racially motivated prosecutors who strike a number of minority jurors but leave one or two on the jury will not necessarily be insulated from *Batson* because the percentage of minority jurors struck will be high.

puted peremptory challenges.²¹²

The most difficult aspect of the disproportionate effects test is determining what degree of disparity gives rise to a prima facie inference of discrimination. *Batson* implied that disproportionate effect, like any other relevant circumstance, would be just one factor in the prima facie inquiry; the required degree of disparity, therefore, ordinarily should depend on the other circumstances surrounding the peremptory challenges in dispute. To bring about *Batson*'s goals more effectively, however, the North Carolina Supreme Court should determine a threshold disparity above which a prima facie case automatically is established. A threshold figure recognizes that when prosecutors strike minorities at a certain substantially greater rate than they strike whites, the risk of an unconstitutional motive is substantial enough to further the inquiry by having the prosecutors explain their challenges, regardless of the absence of other evidence of discrimination.

Support for a per se disproportionate effects criterion comes from three states—Connecticut, Missouri, and South Carolina. Each of these states has adopted *Batson* procedures whereby the prosecutor must explain her peremptory challenges any time the defendant raises a *Batson* claim and demonstrates that he belongs to a cognizable group from which persons were challenged peremptorily.²¹³ These procedures avoid difficult case-by-case evaluations of the circumstances and ensure consistency.²¹⁴ More importantly, they advance the inquiry into the prosecutor's motivations.²¹⁵ Although these procedures are not required by *Batson*,²¹⁶ they further the purposes of the *Batson* prima facie case

212. See *supra* notes 114-21 and accompanying text; see also Note, *supra* note 142, at 823 ("[B]oth the prosecution and the defendant are likely to have better access to evidence on the issue of intent" if the *Batson* claim is heard at trial rather than after appellate review.)

213. See *State v. Holloway*, 209 Conn. 636, 645-46, 553 A.2d 166, 171-72, *cert. denied*, 490 U.S. 1071 (1989); *State v. Antwine*, 743 S.W.2d 51, 64 (Mo. 1987) (en banc), *cert. denied*, 486 U.S. 1017 (1988); *State v. Jones*, 293 S.C. 54, 57-58, 358 S.E.2d 701, 703 (1987). But see *United States v. Montgomery*, 819 F.2d 847, 851 (8th Cir. 1987) (*Batson* does not require specific mathematical formula); *State v. Slappy*, 522 So. 2d 18, 21 (Fla.) (rejecting bright-line test), *cert. denied*, 487 U.S. 1219 (1988); *Jackson v. Commonwealth*, 8 Va. App. 176, 183, 380 S.E.2d 1, 4 (same), *aff'd on rehearing*, 9 Va. App. 169, 384 S.E.2d 343 (1989) (en banc). In *Holloway* and *Jones* the courts simply held that to promote the goals of *Batson* more effectively trial courts should require the prosecutor to explain her peremptory strikes whenever the defendant is a member of a cognizable racial group and the prosecutor exercises peremptory challenges to remove members of that racial group. See *Holloway*, 209 Conn. at 646, 553 A.2d at 171-72; *Jones*, 293 S.C. at 57, 358 S.E.2d at 703. In *Antwine* the court noted that, as a practical matter, determining whether an inference of discrimination exists requires the court to consider the prosecutor's explanation of the manner in which she employed her challenges. Therefore, the court directed trial judges to consider the prosecutor's explanation as part of the process of determining whether a defendant has established a prima facie case. *Antwine*, 743 S.W.2d at 64. Although this method confuses the separate questions of the prima facie case and the prosecutor's rebuttal, it has the effect of requiring a prosecutorial explanation whenever the defendant raises a *Batson* issue.

214. *Jones*, 293 S.C. at 57, 358 S.E.2d at 703.

215. As the Connecticut Supreme Court wrote: "[B]ecause [the *Batson*] issue is of such vital importance to our real and perceived adherence to the rule of law . . . in all . . . cases in which the defendant asserts a *Batson* claim, we deem it appropriate for the state to provide . . . a . . . response consistent with the explanatory mandate of *Batson*." *Holloway*, 209 Conn. at 645-46, 553 A.2d at 171-72.

216. *Batson* mentioned that a "pattern" of strikes against black jurors was merely illustrative of the circumstances that courts could consider when evaluating the prima facie case. *Batson*, 476 U.S. at 97.

by providing defendants with a practical way to get to the next stage of the evidentiary scheme. While these procedures suffer from the drawback of permitting defendants to hunt for *Batson* claims without providing any evidence of discrimination other than the fact that the prosecutor struck members of their race,²¹⁷ the North Carolina courts can modify them to avoid this problem. By requiring that there be a specified degree of disproportionate effect before a prima facie case arises, the North Carolina courts can take advantage of the benefits of these procedures while avoiding their flaws. The self-executing nature of this approach would advance the inquiry into the prosecutor's motives and promote consistency. The requirement of a significant disparity, however, would make it difficult for the defendant to fish for *Batson* claims. Therefore, the North Carolina courts should hold that a prima facie case of *Batson* discrimination is established whenever the prosecutor strikes minorities at a rate a specific number of times greater than the rate at which he strikes whites.²¹⁸ When the prosecutor strikes minorities at a rate lower than the threshold comparison figure, the percentage of minorities struck would be just one relevant circumstance considered.

B. The Prosecutor's Rebuttal

Once the defendant raises a prima facie inference of discrimination, the burden shifts to the prosecutor to come forward with race-neutral explanations for the disputed peremptory challenges.²¹⁹ Assessing prosecutors' explanations is arguably the most difficult aspect of *Batson* for the courts,²²⁰ given the ease with which prosecutors can offer facially neutral reasons for their peremptory challenges.²²¹ Because attorneys normally may exercise peremptory challenges for any reason, the peremptory challenge is "uniquely suited to masking discriminatory motives."²²² *Batson*'s success at this stage, therefore, depends largely on courts adequately scrutinizing prosecutors' proffered reasons to determine whether those reasons are genuine or merely pretexts for discrimination.²²³

The North Carolina Supreme Court has defined the role of the trial courts in evaluating prosecutors' rebuttals.²²⁴ The trial court must "satisfy itself that

217. See Note, *supra* note 142, at 823.

218. See *Williams v. State*, 712 S.W.2d 835, 841 (Tex. Crim. App. 1986) (striking disproportionate number of minority persons so as to render minority representation on the jury impotent can be enough to make out a prima facie showing).

219. *Batson*, 476 U.S. at 97; *State v. Porter*, 326 N.C. 489, 497, 391 S.E.2d 144, 150 (1990).

220. See *Batson*, 476 U.S. at 105 (Marshall, J., concurring).

221. *Id.* at 106 (Marshall, J., concurring).

222. *State v. Slappy*, 522 So. 2d 18, 20 (Fla.), *cert. denied*, 487 U.S. 1219 (1988); see *Batson*, 476 U.S. at 96 (peremptory challenge permits those who want to discriminate to do so).

223. See Raphael, *supra* note 178, at 318 (" 'Rubber stamp' approval of all nonracial explanations . . . would cripple *Batson*'s commitment . . . ") (quoting *State v. Butler*, 731 S.W.2d 265, 268 (Mo. Ct. App. 1987); *Gamble v. State*, 257 Ga. 325, 327, 357 S.E.2d 792, 794 (1987)).

224. The North Carolina appellate courts have reached the issue of the prosecutor's rebuttal in several ways. In some cases, the trial court found that the defendant established a prima facie case. See *State v. Jackson*, 322 N.C. 251, 253, 368 S.E.2d 838, 839 (1988), *cert. denied*, 490 U.S. 1110 (1989); *State v. Sanders*, 95 N.C. App. 494, 498, 383 S.E.2d 409, 412, *disc. rev. denied*, 325 N.C. 712, 388 S.E.2d 470 (1989); *State v. Cannon*, 92 N.C. App. 246, 252, 374 S.E.2d 604, 608 (1988), *rev'd on other grounds*, 326 N.C. 37, 387 S.E.2d 450 (1990). In one case the prosecutor waived argument on

the explanation is genuine' " by undertaking a " 'sincere and reasoned attempt to evaluate [it].' " ²²⁵ This inquiry requires appraising both the race-neutrality of the explanation and the prosecutor's credibility in proffering the explanation. The court "should take great care to assure that [the prosecutor's] reasons are bona fide and not simply 'sham excuses belatedly contrived.' " ²²⁶ Trial judges' determinations should reflect the circumstances of the case, their knowledge of trial techniques, and their observations of the way in which the prosecutor conducted jury selection. ²²⁷ Finally, the court should consider the offered explanation in light of the strength of the prima facie case. ²²⁸

The North Carolina courts also have identified factors specifically for assessing the genuineness of prosecutors' explanations. Trial courts should consider the susceptibility of the case to racial discrimination and should make note of the races of the defendant, the victim, and key witnesses. ²²⁹ The ultimate racial composition of the jury is also relevant, though not dispositive. ²³⁰ In addition, trial courts should factor in whether the prosecutor appeared to deliberate carefully before exercising the peremptory challenges in question. ²³¹ Last, trial courts should evaluate the given reasons themselves. ²³² These factors are similar to those recognized in other states. ²³³ The supreme court has made

whether a prima facie case existed and explained the reasons for his challenges. See *State v. Porter*, 326 N.C. 489, 499, 391 S.E.2d 144, 151 (1990). In another case the trial court assumed, without deciding, that the prima facie case was established, and then considered the prosecutor's explanation. See *State v. McNeill*, 326 N.C. 712, 719, 392 S.E.2d 78, 82 (1990). Finally, in a few cases the prosecutors explained their challenges, notwithstanding that the trial court did not find that the defendants had established prima facie cases. See *State v. McNeil*, 99 N.C. App. 235, 241, 393 S.E.2d 123, 126 (1990) (although not required, State articulated the reasons for its peremptory challenges); *State v. Aytche*, 98 N.C. App. 358, 365, 391 S.E.2d 43, 47 (1990) (prosecutor explained challenges when asked to do so by the defendant); *State v. Attmore*, 92 N.C. App. 385, 397, 374 S.E.2d 649, 657 (1988) (court requested, but did not require, that the prosecutor explain peremptory challenges because of vague possibility of future *Batson* claim), *disc. rev. denied*, 324 N.C. 248, 377 S.E.2d 757 (1989). Since in all of these cases the prosecutors proffered explanations for their peremptory strikes, the sole issue before the appellate courts was the sufficiency of those explanations. In addition, in one case the supreme court found that the defendant established a prima facie case, and thus proceeded to consider the prosecutor's reasons for the disputed peremptory challenges. *State v. Smith*, 328 N.C. 99, 124, 400 S.E.2d 712, 726 (1991).

225. *Sanders*, 95 N.C. App. at 499, 383 S.E.2d at 412-13 (quoting *People v. Hall*, 35 Cal. 3d 161, 167, 672 P.2d 854, 858, 197 Cal. Rptr. 71, 75 (1983)).

226. *Id.* at 500, 383 S.E.2d at 413 (quoting *Jackson*, 322 N.C. at 260, 368 S.E.2d at 843 (Frye, J., concurring)).

227. *Id.* at 499, 383 S.E.2d at 413; see *Porter*, 326 N.C. at 498, 391 S.E.2d at 151.

228. *Porter*, 326 N.C. at 498-99, 391 S.E.2d at 151.

229. *Id.* at 498, 391 S.E.2d at 150-51.

230. *State v. Smith*, 328 N.C. 99, 124, 400 S.E.2d 712, 726 (1991). For a criticism of the court's holding that the racial makeup of the defendant's jury is relevant to *Batson*, see *supra* note 159.

231. *Porter*, 326 N.C. at 498, 391 S.E.2d at 151. Deliberation may tend to show that the prosecutor did not have a predetermined intention to strike minorities, but instead made an individualized decision to strike the juror based on the juror's overall characteristics.

232. *Id.*

233. Other states have also recognized several other factors central to evaluation of the sufficiency of the prosecutor's rebuttal, including whether the prosecutor failed to strike white persons who had the same characteristics as the challenged minority jurors, see *People v. Hall*, 35 Cal. 3d 161, 168, 672 P.2d 854, 858, 197 Cal. Rptr. 71, 76 (1983) ("strongly suggestive of bias"); *Slappy v. State*, 503 So. 2d 350, 355 (Fla. Dist. Ct. App. 1987), *aff'd*, 522 So. 2d 18 (Fla.), *cert. denied*, 487 U.S. 1219 (1988), and whether the proffered reasons were vague and inherently subjective and the prosecutor failed to probe sufficiently to determine if the juror was actually biased. See *People v. Turner*, 42 Cal. 3d 711, 727, 726 P.2d 102, 111-12, 230 Cal. Rptr. 656, 665-66 (1986) (good faith of

clear, however, that no one determinant necessarily demonstrates pretext: "[r]arely will a single factor control the decision-making process."²³⁴

In applying these guidelines, the North Carolina appellate courts have always found the prosecutor's explanations adequate to rebut the prima facie case. Two cases, *State v. Jackson*²³⁵ and *State v. Porter*,²³⁶ aptly illustrate the variety of reasons that the courts have held sufficient. In *Jackson* the prosecutors had to justify four peremptory challenges exercised against black persons.²³⁷ They first explained generally that the State had certain criteria for selecting jurors: stability, pro-government orientation, steady employment, ties to the community, and "a mind-set . . . that would . . . pay more attention to the needs of law enforcement than the fine points of individual rights."²³⁸ They then revealed the reasons for striking each of the four black prospective jurors. Two of the jurors were unemployed.²³⁹ One of these jurors had "answered [the prosecutors] hesitantly and . . . appeared indifferent or hostile about . . . being a member of a jury or indifferent or hostile to [the prosecutors]."²⁴⁰ The other had been a student counselor at Shaw University; the prosecutors felt that her background and demeanor indicated that she was "too liberal."²⁴¹ The third black prospective juror was a law student at the University of North Carolina and had been taught by professors of "somewhat liberal views."²⁴² The fourth juror had a son of approximately the same age as the defendant; although the juror also had a daughter the same age as the victim, the prosecutors feared that she would identify with the defendant and not the prosecution.²⁴³ The supreme court held that the prosecutors' stated criteria for choosing a jury were legitimate and approved the prosecutor's explanations.²⁴⁴

In *Porter* the state peremptorily challenged ten Native Americans.²⁴⁵ The

prosecutor required follow-up question that quickly would have clarified the matter); *People v. Harris*, 129 Ill. 2d 123, 188, 544 N.E.2d 357, 386 (1989) (courts should assess the extent of State's efforts to discover the unknown information), *cert. denied*, 110 S. Ct. 1323 (1990).

234. *Porter*, 326 N.C. at 501, 391 S.E.2d at 152; see *State v. Jackson*, 322 N.C. 251, 257, 368 S.E.2d 838, 841 (1988) (recognizing that nonchallenged white persons had same characteristics as challenged black persons, but noting that additional factors distinguished them), *cert. denied*, 490 U.S. 1110 (1989).

235. 322 N.C. 251, 368 S.E.2d 838 (1988), *cert. denied*, 490 U.S. 1110 (1989).

236. 326 N.C. 489, 391 S.E.2d 144 (1990).

237. *Jackson*, 322 N.C. at 252-53, 368 S.E.2d at 839.

238. *Id.* at 253, 368 S.E.2d at 840. In a concurring opinion, however, Justice Frye warned against the danger of abuse of such stated criteria or "profiles." *Id.* at 260, 368 S.E.2d at 843 (Frye, J., concurring).

239. *Id.* at 253, 368 S.E.2d at 839; see also *State v. Sanders*, 95 N.C. App. 494, 501, 383 S.E.2d 409, 414 (juror had worked three jobs in previous ten months), *disc. rev. denied*, 325 N.C. 712, 388 S.E.2d 470 (1989).

240. *Jackson*, 322 N.C. at 253, 368 S.E.2d at 839; see also *State v. Smith*, 328 N.C. 99, 125, 400 S.E.2d 712, 727 (1991) (jurors appeared "nervous" and "uncertain"); *State v. McNeil*, 99 N.C. App. 235, 241, 393 S.E.2d 123, 126 (1990) (juror seemed "unsure of himself"); *Sanders*, 95 N.C. App. at 501, 383 S.E.2d at 414 (juror had "headstrong and overbearing personality").

241. *Jackson*, 322 N.C. at 253, 368 S.E.2d at 839.

242. *Id.*

243. *Id.*; see also *Smith*, 328 N.C. at 125, 400 S.E.2d at 726 (jurors had sons of approximately the defendant's age).

244. *Jackson*, 322 N.C. at 256-57, 368 S.E.2d at 841.

245. *State v. Porter*, 326 N.C. 489, 499, 391 S.E.2d 144, 151 (1990).

prosecutor proffered reasons for each challenge. Many of the challenged jurors knew one or both of the defense attorneys.²⁴⁶ Several of the prospective jurors previously had been prosecuted for driving while intoxicated.²⁴⁷ The employment histories of others reflected unemployment or irregular employment.²⁴⁸ Two seemed to believe that racism was present in the case.²⁴⁹ One prospective juror "made constant eye contact with defense counsel, had majored in sociology, and read Rolling Stone magazine."²⁵⁰ On appeal, the supreme court affirmed the trial court's acceptance of these reasons as legitimate nonracial criteria for challenging the Native Americans.²⁵¹

The wide range of explanations that the North Carolina courts have found sufficient to rebut a prima facie case illustrates the difficulties courts face in applying this part of the *Batson* evidentiary system. Prosecutors easily can proffer facially neutral reasons for their peremptory challenges. All of the reasons the prosecutors asserted in *Jackson* and *Porter* are facially race-neutral and thus technically proper under *Batson*.²⁵² Many of them, however, are completely subjective and difficult to disprove. The *Jackson* court, for instance, accepted the explanation that a challenged juror "appeared indifferent or hostile,"²⁵³ similarly, the *Porter* court found sufficient the explanation that the challenged juror "made constant eye contact with defense counsel."²⁵⁴ Furthermore, several of the reasons that the courts have accepted appear nonracial, but actually may be a proxy for race. In *Jackson* the prosecutor excused one prospective juror because she had been a student counselor at Shaw University—a black university. In *Porter* the court accepted the prosecutors' assertions that several of the black prospective jurors had histories of unemployment or unsteady employment; the unemployment rate for black persons in North Carolina is more than twice that

246. The challenged prospective jurors were acquainted with the defendant's lawyers as a result of prior representation, being schoolmates, being a student of one of the lawyers, being related by marriage, or being social acquaintances. *Id.* at 499, 391 S.E.2d at 151; see also *Smith*, 328 N.C. at 125, 400 S.E.2d at 727 (juror had "earlier association" with defense counsel); *State v. McNeill*, 326 N.C. 712, 719, 392 S.E.2d 78, 82 (1990) (juror knew defendant, though had not seen him in five years); *State v. McNeil*, 99 N.C. App. 235, 241, 393 S.E.2d 123, 126 (1990) (juror knew proposed defense witness); *State v. Cannon*, 92 N.C. App. 246, 252-53, 374 S.E.2d 604, 608 (1988) (five of six challenged jurors connected to defendant, defendant's family, or a state witness), *rev'd on other grounds*, 326 N.C. 37, 387 S.E.2d 450 (1990).

247. *Porter*, 326 N.C. at 499, 391 S.E.2d at 151; see also *Smith*, 328 N.C. at 125, 400 S.E.2d at 727 (State's victim-witness coordinator thought juror had a nephew "in trouble with drugs"); *McNeil*, 99 N.C. App. at 241, 393 S.E.2d at 126 (jurors had prior convictions for driving under the influence, had been falsely accused of crime, or had relatives on probation); *State v. Aytche*, 98 N.C. App. 358, 365, 391 S.E.2d 43, 47 (1990) (juror previously convicted of manslaughter); *Cannon*, 92 N.C. App. at 253, 374 S.E.2d at 608 (juror recently fined for traffic violation, although claimed to be innocent).

248. *Porter*, 326 N.C. at 499, 391 S.E.2d at 151.

249. *Id.* at 500, 391 S.E.2d at 152.

250. *Id.* This potential juror also previously had been a witness for the defendant's attorney in a different case. *Id.*

251. *Id.*

252. See *Batson*, 476 U.S. at 97.

253. *State v. Jackson*, 322 N.C. 251, 253, 368 S.E.2d 838, 839 (1988), *cert. denied*, 490 U.S. 1110 (1989).

254. *Porter*, 326 N.C. at 500, 391 S.E.2d at 152.

for white persons.²⁵⁵ As Justice Marshall warned: "If such easily generated explanations are sufficient to discharge the prosecutor's obligation to justify his strikes on nonracial grounds, then the protection erected by the Court . . . may be illusory,"²⁵⁶ because only the most flagrant and explicit *Batson* violations would be caught.

The courts implementing *Batson* thus face this question: Which party should bear the burden of overcoming the difficulty of evaluating the genuineness of prosecutors' explanations? The courts generally have two choices. On the one hand, the courts can put this burden on defendants and accept all facially nonracial explanations. On the other hand, they can put the burden on prosecutors and reject certain facially neutral reasons that are particularly susceptible to abuse as pretexts. Accepting all facially nonracial explanations is supported by the literal language of *Batson*, which requires only that prosecutors proffer "neutral explanation[s]."²⁵⁷ This approach also would minimize the intrusion on the prosecutor's right to exercise peremptory challenges, since all peremptory challenges for which the prosecutor can offer a neutral explanation—effectively most of them—would be acceptable. This tactic, however, would facilitate prosecutorial efforts to hide discriminatory motives and evade *Batson*'s constitutional proscription. Rejecting certain facially neutral reasons, alternatively, better effects *Batson*'s ultimate purpose of eradicating racially motivated peremptory challenges. It would snare more racially motivated peremptory challenges and have a greater deterrent effect on future discriminatory challenges. This alternative, however, would seriously intrude on the prosecutorial peremptory challenge. Courts nationwide have responded differently to this apparent quandary.²⁵⁸

Batson implicitly furnishes an answer to the dilemma: the State should bear the burden of *Batson*'s inherent weaknesses and courts should be slow to accept prosecutorial explanations for which there is a high risk of abuse, even though such an approach threatens the traditional, capricious nature of the peremptory challenge. First, by creating a system that requires prosecutors to explain their

255. According to recent statistics from the U.S. Department of Labor, the unemployment rate for blacks in North Carolina in 1989 was 6.3% while the unemployment rate for North Carolina whites was only 2.8%. U.S. DEPT. OF LABOR, BUREAU OF LABOR STATISTICS, NORTH CAROLINA UNEMPLOYMENT IN 1989 (Press Release, May 11, 1990).

256. *Batson*, 476 U.S. at 106 (Marshall, J., concurring).

257. *Id.* at 97.

258. Compare *United States v. Forbes*, 816 F.2d 1006, 1010-11 (5th Cir. 1987) (intuitive nonracial reasons are sufficient); *United States v. Allen*, 666 F. Supp. 847, 853 (E.D. Va. 1987) (same), *aff'd sub nom. United States v. Harrell*, 847 F.2d 138 (4th Cir.), *cert. denied*, 488 U.S. 944 (1988); *People v. Young*, 128 Ill. 2d 1, 20, 538 N.E.2d 453, 457 (1989) (juror's demeanor acceptable basis for peremptory challenge), *cert. denied*, 110 S. Ct. 3290 (1990); *State v. Manuel*, 517 So. 2d 374, 376 (La. Ct. App. 1987) (same) and *Porter*, 326 N.C. at 500, 391 S.E.2d at 152 (failure to make eye contact with prosecutor and excessive eye contact with defense counsel acceptable reasons for peremptory challenge) with *United States v. Clemons*, 843 F.2d 741, 745 (3d Cir.) ("intuitive judgment" not sufficient), *cert. denied*, 488 U.S. 835 (1988); *United States v. Chalan*, 812 F.2d 1302, 1314 (10th Cir. 1987) (explanation that State struck juror "based upon his background and other things in his questionnaire" insufficient); *State v. Slappy*, 522 So. 2d 18, 23-24 (Fla.) (must be support in the record for the reasons given), *cert. denied*, 487 U.S. 1219 (1988) and *Commonwealth v. Jackson*, 386 Pa. Super. 29, 55, 562 A.2d 338, 350 (1989) (abuse of discretion to accept explanation not supported by record), *appeal denied*, 525 Pa. 631, 578 A.2d 926 (1990).

peremptory challenges, *Batson* expressly sanctions intrusion on the traditional arbitrary use of peremptory challenges. Furthermore, the *Batson* decision was rooted in the equal protection clause of the fourteenth amendment; its primary goal of ending discriminatory jury selection practices, unlike the peremptory challenge, is thus of constitutional magnitude.²⁵⁹ Most importantly, the *Batson* Court specifically noted that the purpose of the evidentiary system it set forth is "to ensure that a state does not use peremptory challenges to strike *any* black juror because of his race."²⁶⁰ The Court itself thus stressed the predominance of the nondiscrimination purpose of the decision. In short, the purpose and constitutional basis of *Batson* require that lower courts do more than satisfy themselves that prosecutorial explanations are race neutral; rather, they must scrutinize the explanations closely to ensure in a practical way that the explanations are not pretexts. This is not because these explanations are technically improper under *Batson*, but because they are peculiarly susceptible to being abused by prosecutors. The North Carolina courts, which in the past have readily accepted even highly subjective prosecutorial explanations, therefore should revise their approach to conform with this view, which is more in accordance with *Batson*'s spirit and purpose.

A reasonable, practical way to implement this closer scrutiny is to require prosecutors to support their peremptory challenges with other on-the-record reasons whenever their stated explanations involve a substantial risk of being pretextual.²⁶¹ Several types of prosecutorial explanations are so susceptible to abuse that heightened scrutiny of this sort is warranted. One such type involves reasons that may be proxies for racial animus.²⁶² Professor Raphael has noted, for example, that because a substantial percentage of black persons live in areas of high crime or in segregated areas, explanations based upon the challenged juror's connection with such areas easily may mask underlying discrimination.²⁶³ Likewise, reasons that apply disproportionately to minorities, such as unemployment, or previous study at a predominately black university, may be facially neutral substitutes for race-based reasons.²⁶⁴ The second type of prosecutorial explanation justifying greater scrutiny includes reasons that apply equally to white prospective jurors whom the prosecutor did not strike.²⁶⁵ The North Carolina Supreme Court already has recognized that disparate treatment

259. See *Batson*, 476 U.S. at 108 (Marshall, J., concurring) (quoting *Frazier v. United States*, 335 U.S. 497, 505 n.11 (1948); *United States v. Wood*, 299 U.S. 123, 145 (1936); and *Stilson v. United States*, 250 U.S. 583, 586 (1919)).

260. *Batson*, 476 U.S. at 99 n.22 (emphasis added); see also *United States v. David*, 803 F.2d 1567, 1571 (11th Cir. 1986) (*Batson*'s command is "to eliminate, not merely to minimize, racial discrimination in jury selection.").

261. See Blume, *supra* note 121, at 332 ("[C]ourts should be extremely reluctant to accept general and vague reasons for striking a juror, such as the juror was 'sullen,' 'distant,' or 'looked mean.'").

262. See *State v. Jackson*, 322 N.C. 251, 260, 368 S.E.2d 838, 843 (1988) (Frye, J., concurring) (court must remain alert to colloquial euphemisms for prejudice), *cert. denied*, 490 U.S. 1110 (1989); Alschuler, *supra* note 7, at 175; Raphael, *supra* note 178, at 322.

263. Raphael, *supra* note 178, at 322.

264. See *supra* text accompanying notes 241, 248.

265. See, e.g., *People v. Hall*, 35 Cal. 3d 161, 168, 672 P.2d 854, 858, 197 Cal. Rptr. 71, 76 (1983) (disparate treatment "strongly suggestive of bias"); Raphael, *supra* note 178, at 323 ("Courts

of white and minority jurors is a factor for evaluating pretext.²⁶⁶ The third and most important type of prosecutorial explanation that involves a substantial danger of being used as a pretext is the vague and highly subjective explanation.²⁶⁷ This type of explanation is the most difficult for the trial judge and defendant to verify because the substance of the explanation does not appear in the record and thus is impossible to review on appeal.²⁶⁸ Hence, they are the explanations most susceptible to prosecutorial chicanery. Whenever prosecutors submit one of these types of explanations, the court should not find the explanation sufficient unless there are some other reasons for the peremptory challenge that are observable from the record.

Several courts, including those in Florida and Pennsylvania, require that at least some evidence appear in the record to support the prosecutor's explanations.²⁶⁹ Furthermore, some authority calls for requiring objective justifications for prosecutors' usually discretionary actions when a danger exists that the prosecutors acted unconstitutionally. Prosecutors normally have discretion to decide what charges to bring against a suspect. When a prosecutor adds charges against a defendant who has successfully appealed his conviction, however, a danger arises that the prosecutor does so in retaliation for the defendant having taken the appeal.²⁷⁰ Such conduct, if so motivated, would violate the defendant's right to due process of law.²⁷¹ Because the prosecutor's motives are com-

are most likely to reject a prosecutor's explanations . . . when the prosecutor failed to challenge other jurors who were not of defendant's race yet who shared the same characteristic . . .").

266. *State v. Porter*, 326 N.C. 489, 501, 391 S.E.2d 144, 152 (1990). The court qualified this acknowledgement, however, by noting that prosecutors rarely exercise peremptory challenges on the basis of one reason alone. *Id.* Of course, the court was correct in asserting that multiple reasons often underlie peremptory challenges. Nevertheless, disparate treatment is so suggestive of racial motivation that it warrants greater scrutiny.

267. *State v. Smith*, 328 N.C. 99, 125, 400 S.E.2d 712, 727 (1991); *State v. Jackson*, 322 N.C. 251, 253, 368 S.E.2d 838, 839 (1988), *cert. denied*, 490 U.S. 1110 (1989); see *supra* notes 252-56 and accompanying text.

268. In *Smith* the supreme court held that "nervousness or uncertainty in response to counsel's questions may be a proper basis for a peremptory challenge." *Smith*, 328 N.C. at 126, 400 S.E.2d at 727. The court then found that "the record supports the . . . conclusion that the reasons given by the district attorney were not pretextual." *Id.* The court thus suggested that the appellate courts can adequately evaluate the prosecutor's bona fides in proffering vague and subjective explanations. How the record can support a finding that a juror was "nervous" or "uncertain" absent a direct statement in the record to that effect, however, is unclear.

269. See *State v. Slappy*, 522 So. 2d 18, 23-24 (Fla.) (must be support in the record for the reasons given to ensure "procedural regularity and racial neutrality"), *cert. denied*, 487 U.S. 1219 (1988); *Commonwealth v. Jackson*, 386 Pa. Super. 29, 55, 562 A.2d 338, 350 (1989) (abuse of discretion to accept explanation not supported by the record), *appeal denied*, 525 Pa. 631, 578 A.2d 926 (1990). Similarly, several courts have found subjective explanations alone to be insufficient to rebut a prima facie case. See, e.g., *United States v. Horsley*, 864 F.2d 1543, 1546 (11th Cir. 1989) ("I just got a feeling about him" held insufficient); *United States v. Clemons*, 843 F.2d 741, 745 (3d Cir.) (prosecutor's intuitive judgment held insufficient), *cert. denied*, 488 U.S. 835 (1988); *Ward v. State*, 293 Ark. 88, 93-94, 733 S.W.2d 728, 730 (1987) (insufficient to claim that challenged juror seemed "noncommittal"); *Ex parte Branch*, 526 So. 2d 609, 623 (Ala. 1987) ("[I]ntuitive judgment or suspicion by the prosecutor is insufficient to rebut the presumption of discrimination."); *State v. Antwine*, 743 S.W.2d 51, 65 (Mo. 1987) (*en banc*) (facially neutral explanations alone held insufficient), *cert. denied*, 486 U.S. 1017 (1988); *State v. Tomlin*, 299 S.C. 294, 298-99, 384 S.E.2d 707, 710 (1989) (claim that challenged juror was "extremely sluggish" held insufficient).

270. *Blackledge v. Perry*, 417 U.S. 21, 28 (1974). In addition, defendants might be unconstitutionally deterred from exercising their right to appeal. *Id.*

271. *United States v. Goodwin*, 457 U.S. 368, 375-76 (1982).

plex and difficult to prove, the court in these circumstances may presume "prosecutorial vindictiveness."²⁷² When this presumption arises, the prosecutor must provide objective, on-the-record evidence supporting the decision to add the new charges.²⁷³ Although the decision to bring charges normally is discretionary, the danger of unconstitutionally motivated prosecutorial conduct justifies these limitations on the prosecutor's discretion.²⁷⁴ Similarly, when a *prima facie* inference of discrimination arises in the *Batson* context, the court should require the prosecutor to give at least some objective, verifiable reason for the questioned peremptory challenges. Although prosecutors ordinarily may exercise peremptory challenges for any reason, the danger that the prosecutor's explanations will be pretextual warrants limiting prosecutors' discretion.

C. Procedural Issues

In addition to determining what evidence gives rise to a *prima facie* inference of discrimination and what prosecutorial explanations rebut such an inference, the *Batson* Court assigned to the lower courts the task of formulating procedures for litigating *Batson* claims.²⁷⁵ The North Carolina courts' holdings regarding procedural issues parallel their holdings involving the *prima facie* case and the prosecutor's rebuttal. In the area of harmless error, for example, as with the *prima facie* case, the North Carolina courts have misapprehended *Batson's* theoretical underpinnings and, accordingly, have imposed undue burdens on defendants. In the areas of appellate review and cross-examination of the prosecutor, as with the prosecutor's rebuttal, the courts have construed *Batson's* language so narrowly that they have failed to promote *Batson's* fundamental purpose. In sum, the courts have erected procedural barriers that conflict with *Batson's* goal of providing defendants with a more effective means of proving discrimination.

1. Appellate Review

When formulating a standard of appellate review of trial court findings in *Batson* cases, the courts face yet another contradiction inherent in *Batson* situations: the trust reposed in prosecutors and trial judges by appellate courts sometimes conflicts with the realities of racial politics. For example, while the *Batson* Court observed in a footnote that judges and prosecutors would not fail to perform their constitutional duties under *Batson*,²⁷⁶ the decision itself is a recognition that government officials sometimes engage in unconstitutional discrimination. As a result, the degree to which the appellate courts should yield to the findings of the trial courts is questionable.

272. *Id.* at 373.

273. *Id.* at 374.

274. *See id.* at 376; *Perry*, 417 U.S. at 28.

275. *See Batson*, 476 U.S. at 99 ("We decline . . . to formulate particular procedures to be followed upon a defendant's timely objection to a prosecutor's challenges.").

276. *Id.* at 99 n.22.

In *State v. Jackson*²⁷⁷ the North Carolina Supreme Court held that because a trial court's findings depend on credibility determinations, those findings must be given "great deference" on appeal.²⁷⁸ Trial judges must make specific findings of fact,²⁷⁹ and those findings are conclusive on appeal, provided they are supported by the evidence.²⁸⁰ In reviewing *Batson* cases, the supreme court and court of appeals have been true to this standard. In numerous cases, the courts explicitly have noted that deference to the trial court commanded their decision to deny defendants *Batson* relief.²⁸¹ The courts routinely have declined to strike down even highly subjective rebuttal explanations on appeal, upholding all facially nonracial explanations.²⁸² The courts have been reluctant to give much weight to objective indicators of discrimination, such as disparate treatment of minority and white prospective jurors. Instead, they have relied on the general proposition that prosecutors rarely exercise peremptory challenges for single reasons and have deferred to the trial courts' overall evaluations of the circumstances.²⁸³ This extremely deferential approach almost certainly has contributed to the fact that no North Carolina defendant has won a *Batson* claim on the merits in the appellate courts.

While the appellate courts must accord the trial court's findings some deference, excessive deference has two related undesirable effects. First, it effectively "insulates the trial court's determinations from meaningful appellate review."²⁸⁴ Second, it sends to prosecutors the message that they may exercise peremptory challenges on account of race without fear of exposure by the appellate courts.²⁸⁵ Because the second effect in particular is contrary to *Batson*'s fundamental purpose, meaningful appellate review is essential to *Batson*'s success.²⁸⁶

277. 322 N.C. 251, 368 S.E.2d 838 (1988), *cert. denied*, 490 U.S. 1110 (1989); *see supra* notes 235, 237-44 and accompanying text.

278. *Jackson*, 322 N.C. at 255, 368 S.E.2d at 840 (citing *Batson*, 476 U.S. at 98 n.21); *see State v. Porter*, 326 N.C. 489, 498, 391 S.E.2d 144, 150 (1990); *State v. Allen*, 323 N.C. 208, 219, 372 S.E.2d 855, 861-62 (1988). Other states have set forth similar standards of review. *See Gamble v. State*, 257 Ga. 325, 327, 357 S.E.2d 792, 794 (1987); *People v. Harris*, 129 Ill. 2d 123, 175, 544 N.E.2d 357, 380 (1989), *cert. denied*, 110 S. Ct. 1323 (1990); *State v. Antwine*, 743 S.W.2d 51, 66 (Mo. 1987) (en banc), *cert. denied*, 486 U.S. 1017 (1988); *Jackson v. Commonwealth*, 8 Va. App. 176, 184, 380 S.E.2d 1, 5, *aff'd on rehearing*, 9 Va. App. 169, 384 S.E.2d 343 (1989) (en banc).

279. *State v. Sanders*, 95 N.C. App. 494, 500, 383 S.E.2d 409, 413, *disc. rev. denied*, 325 N.C. 712, 388 S.E.2d 470 (1989).

280. *Id.*

281. *State v. Gray*, 322 N.C. 457, 459-60, 368 S.E.2d 627, 629 (1988); *State v. McNeil*, 99 N.C. App. 235, 241, 393 S.E.2d 123, 126 (1990); *State v. Robinson*, 97 N.C. App. 597, 600, 601, 389 S.E.2d 417, 419, 420, *dismissal allowed and review denied*, 326 N.C. 804, 393 S.E.2d 904 (1990); *State v. Batts*, 93 N.C. App. 404, 410, 378 S.E.2d 211, 214 (1989); *State v. Cannon*, 92 N.C. App. 246, 253, 374 S.E.2d 604, 608 (1988), *rev'd on other grounds*, 326 N.C. 37, 387 S.E.2d 450 (1990). In fact, the *Jackson* court noted: "We might not have reached the same result as the superior court but giving, as we must, deference to its findings, we hold it was not error to deny the defendant's motion for mistrial." *Jackson*, 322 N.C. at 257, 368 S.E.2d at 841.

282. *See supra* notes 235-56 and accompanying text.

283. *See, e.g., State v. Porter*, 326 N.C. 489, 501, 391 S.E.2d 144, 152-53 (1990).

284. Blume, *supra* note 121, at 328.

285. *Id.* at 329. Conversely, if prosecutors know that the appellate courts will review their actions carefully, they are less likely to engage in discriminatory conduct in the first place.

286. Curiously, in *Batson*, Justice Powell seemed to suggest that there is little danger of prosecutors exercising peremptory challenges for discriminatory reasons. In justifying his rejection of Justice Marshall's suggestion that the peremptory challenge be abolished, Justice Powell wrote that no

Several reasons justify more aggressive appellate review of *Batson* claims. First, because of their own unconscious racism, trial judges simply may not be aware of racial discrimination during jury selection.²⁸⁷ In addition, although the *Batson* Court expressed confidence in the ability of trial judges to recognize *Batson* claims,²⁸⁸ the ultimate responsibility within each state for vindicating rights under the equal protection clause lies with the state supreme court. Furthermore, as with prosecutors' rebuttal explanations,²⁸⁹ the ease with which prosecutors can put forth pretextual but facially neutral explanations necessitates appellate intervention. Most important, appellate courts have a greater duty to scrutinize trial court findings when constitutional rights are at stake.²⁹⁰ As the Maryland Court of Appeals held in the *Batson* context: "When a claim is based upon a violation of a constitutional right it is [the appellate court's] obligation to make an independent constitutional appraisal from the entire record."²⁹¹ Therefore, the North Carolina appellate courts should scrutinize *Batson* claims much more closely than in the past. To scrutinize these claims without merely substituting their judgments on credibility issues for those of the trial judges, the supreme court and court of appeals should place greater emphasis on objective

reason supports the belief that prosecutors will shirk their duty to exercise peremptory challenges legitimately. *Batson*, 476 U.S. at 99 n.22. If that were the case, there might be little need for deterrence. The fact that *Batson* was necessary, however, together with the large number of *Batson* challenges brought, is evidence that some prosecutors do exercise peremptories on account of race. Indeed, Justice White commented that the practice is widespread. *Id.* at 101 (White, J., concurring). Thus, aggressive appellate review to deter prosecutors' unconstitutional uses of peremptory challenges is appropriate.

287. *Id.* at 106 (Marshall, J., concurring) ("A judge's own conscious or unconscious racism may lead him to accept . . . an explanation as well supported."); Johnson, *Unconscious Racism and the Criminal Law*, 73 CORNELL L. REV. 1016, 1026-36 (1988). Professor Johnson asserts that many judicial actions and inactions are colored by judges' own unconscious racism. She contends that this undetected racism results in a "blindspot" that causes judges to fail to recognize their own or other persons' actions as racially connected. She cites three reasons for the fact that unconscious racism is at present ignored in the reasoning of court decisions involving race and criminal procedure: (1) the mistaken belief that racism is equivalent to white supremacism, (2) the fear that there would be no limiting principle, and (3) denial. *Id.* at 1027-31. See generally Lawrence, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 328-44 (1987) (discussing relationship between the unconscious and racially discriminatory practices).

288. See *Batson*, 476 U.S. at 99 n.22; *id.* at 101 (White, J., concurring) (Court puts "considerable trust" in the trial judge).

289. See *supra* notes 252-68 and accompanying text.

290. See *Bose Corp. v. Consumers Union, Inc.*, 466 U.S. 485, 513-14 (1984) (although questions of fact are normally entitled to deferential appellate review, "actual malice" in constitutional defamation case is subject to independent appellate review because its determination affects first amendment rights).

291. *Tolbert v. State*, 315 Md. 13, 24, 553 A.2d 228, 232 (1989) (quoting *Harris v. State*, 303 Md. 685, 697, 496 A.2d 1074, 1080 (1985)); see *Commonwealth v. Jackson*, 386 Pa. Super. 29, 55, 562 A.2d 338, 350 (1989) (trial court's discretion is not unlimited), *appeal denied*, 525 Pa. 631, 578 A.2d 926 (1990).

In addition, the role of the state appellate courts in protecting defendants' rights under *Batson* may be amplified in light of the Supreme Court's decision in *Wainwright v. Witt*, 469 U.S. 412 (1985). In *Wainwright* the Court addressed the question whether federal courts, on petitions for habeas corpus relief, may review state court findings on whether a juror was challenged properly for cause under *Witherspoon v. Illinois*, 391 U.S. 510 (1968), on account of bias stemming from the juror's opposition to the death penalty. The Court held that the propriety of the challenge for cause is a "factual issue" entitled to a presumption of correctness, because the determination depends largely on credibility findings. *Wainwright*, 469 U.S. at 428. Since credibility determinations also are central to *Batson* inquiries, the scope of federal habeas corpus review may be limited in this context as well. Consequently, the states must aggressively promote *Batson*'s goals.

criteria. If objective evidence²⁹² suggests discrimination, or if the prosecutor's reasons for challenging the jurors are completely subjective, the courts should find a *Batson* error on appeal. Just as with the prosecutor's rebuttal, this focus on observable, on-the-record criteria does not mean that less easily observable criteria are irrelevant to the question of whether the prosecutor has exercised peremptory challenges in a discriminatory manner. Rather, it is an attempt to promote the goals of *Batson* more effectively by providing a reasonable, practical way of reviewing *Batson* cases.

A threshold requirement for meaningful appellate review is an adequate record.²⁹³ This record must include the number of black and white persons on the venire, the race of each prospective juror examined, and a transcript of the voir dire.²⁹⁴ In North Carolina, defendants appealing from the denial of a *Batson* claim have the burden of providing an adequate record from which to determine whether the prosecutor challenged jurors improperly at trial.²⁹⁵

In the five years since *Batson*, North Carolina defendants consistently have failed to meet their burden of providing an adequate record for appeal. On many occasions, the appellate courts pointedly noted these failures in denying defendants' *Batson* claims.²⁹⁶ These flaws can be eliminated easily, however, because North Carolina law currently enables defendants to perfect adequate appellate records. The jury selection statute permits transcription of the jury selection proceedings upon the defendant's request.²⁹⁷ Thus, at the beginning of all cases in which *Batson* may be an issue, defendants can invoke their statutory right to have the voir dire transcribed. Slightly more problematic is the issue of preserving for the record the race of each prospective juror. The supreme court wisely noted in *State v. Mitchell*²⁹⁸ that the practice of simply having the court reporter record the race of each potential juror is prone to error because race is

292. Objective criteria include, for example, a disproportionate number of peremptory challenges exercised against minorities or disparate treatment of minority and white prospective jurors with similar pertinent characteristics.

293. See, e.g., *State v. Holloway*, 209 Conn. 636, 644, 553 A.2d 166, 172, cert. denied, 490 U.S. 1071 (1989); *Aldridge v. State*, 258 Ga. 75, 77, 365 S.E.2d 111, 113 (1988); *Jackson*, 386 Pa. Super. at 52, 562 A.2d at 349; Note, *supra* note 142, at 825 ("All of the possible evidence that may bear on the defendant's prima facie showing depends on the creation of an adequate record [for appeal].").

294. See *Stanley v. State*, 313 Md. 50, 70 n.11, 542 A.2d 1267, 1277 n.11 (1988). Although a record consisting solely of the race of the persons struck and of the persons who served on the jury could "marginally suffice[]" for review, a transcript is necessary as a practical matter to facilitate appellate courts' meaningful evaluations of the trial court's findings. *State v. Sanders*, 95 N.C. App. 494, 499, 383 S.E.2d 409, 412, disc. rev. denied, 325 N.C. 712, 388 S.E.2d 470 (1989).

295. See *State v. Mitchell*, 321 N.C. 650, 654, 365 S.E.2d 554, 556 (1988). The mere statements of counsel on appeal regarding what transpired at trial will not suffice. *Id.*

296. See *State v. Payne*, 327 N.C. 194, 199-200, 394 S.E.2d 158, 160-61 (1990), cert. denied, 111 S. Ct. 977 (1991); *State v. McNeill*, 326 N.C. 712, 718-19, 392 S.E.2d 78, 82 (1990); *State v. Aytche*, 98 N.C. App. 358, 364, 391 S.E.2d 43, 47 (1990); *State v. Robinson*, 97 N.C. App. 597, 601, 389 S.E.2d 417, 420, disc. rev. denied, 326 N.C. 804, 393 S.E.2d 904 (1990); *Sanders*, 95 N.C. App. at 499, 383 S.E.2d at 412; *State v. Cannon*, 92 N.C. App. 246, 251, 374 S.E.2d 604, 607 (1988), *rev'd on other grounds*, 326 N.C. 37, 387 S.E.2d 450 (1990).

297. In capital cases, a record of the voir dire is required by statute. N.C. GEN. STAT. § 15A-1241(a)(1) (1988). In noncapital cases, the jury selection must be recorded if defense counsel so requests. *Id.* § 15A-1241(b); R. PRICE, *supra* note 54, § 18-3.

298. 321 N.C. 650, 365 S.E.2d 554 (1988).

not always obvious.²⁹⁹ In response to the problem of how to preserve the record, however, *Mitchell* provides that a defendant who believes a prospective juror is of a particular race may so inform the trial court and ensure that the information is placed on the record.³⁰⁰ For defendants to have to question prospective jurors about their race during voir dire to determine officially the race of prospective jurors, however, is unfair. Such questions might insult the juror and convey to the jury the impression that the defendant intends to make race an issue in the trial. Hence, the defendant would have a Hobson's choice: preserving the record and risking alienating the jury, or not preserving the record at all. Upon the defendant's pretrial request, therefore, the trial judge should inquire about each prospective juror's race. Under this practice, the race of the potential juror becomes a routine piece of requested information no different from the juror's address; as such, it is less likely to insult the juror. Moreover, because the judge asks the questions, defense counsel will not antagonize the jury. The North Carolina Supreme Court has hinted that such a method would be permissible,³⁰¹ but the court should go further and formally adopt it.

2. Examination of the Prosecutor

Another procedural issue that has arisen in North Carolina *Batson* claims is whether defendants may cross-examine prosecutors regarding their rebuttal explanations.³⁰² In *State v. Jackson* the supreme court held that defendants have no such right.³⁰³ The court feared that the disruption to the trial would outweigh any good that could be achieved by the prosecutor's testimony.³⁰⁴ Moreover, the court was confident that trial judges would be able to pass on prosecutors' credibility without the aid of cross-examination.³⁰⁵ The *Jackson* court did hold, however, that once prosecutors advance their reasons for the peremptory challenges in question, defendants may offer additional evidence to strengthen the inference of purposeful discrimination or to expose the prosecu-

299. *Id.* at 655-56, 365 S.E.2d at 557. Moreover, the court reporter's guess for the race of the potential juror is essentially a racial judgment based solely on appearance, and thus is susceptible to the same stereotypes that underlie race-based peremptory challenges.

300. *Id.* at 656, 365 S.E.2d at 557. If there is any question about the juror's race, the trial court may question the juror to determine it. *Id.*; see also *State v. Payne*, 327 N.C. 194, 199-200, 394 S.E.2d 158, 160-61 (1990) (suggesting that prior to trial defendant ask that the trial judge ask prospective jurors to state their race for the record during initial questioning).

301. In *Payne*, the court wrote: "The trial court noted . . . that had the defendant made his motion prior to jury selection, the court would have had each prospective juror state his or her race during the court's initial questioning. This would have . . . preserved an adequate record for appellate review." 327 N.C. at 199-200, 394 S.E.2d at 160; see also *State v. Attmore*, 92 N.C. App. 385, 397, 374 S.E.2d 649, 657 (1988) (trial judge noted race and sex of each person examined during voir dire), *disc. rev. denied*, 324 N.C. 248, 377 S.E.2d 757 (1989).

302. See Note, *Defense Presence and Participation: A Procedural Minimum for Batson v. Kentucky Hearings*, 99 YALE L.J. 187, 205-06 (1989) [hereinafter Note, *Defense Presence and Participation*]; see also Note, *supra* note 142, at 832-36 (discussing advantages and disadvantages of adversarial *Batson* hearings).

303. *State v. Jackson*, 322 N.C. 251, 258, 368 S.E.2d 838, 840 (1988), *cert. denied*, 490 U.S. 1110 (1989).

304. *Id.*; see also Note, *Defense Presence and Participation*, *supra* note 302, at 205 (administrative burden would result from adversarial *Batson* hearings).

305. *Jackson*, 322 N.C. at 258, 368 S.E.2d at 842.

tors' proffered reasons as mere pretexts for discrimination.³⁰⁶

In light of the purposes of *Batson*, the court underestimated the importance of examining the prosecutor and thus unduly restricted the defendant's ability to expose unconstitutional discrimination.³⁰⁷ The trial judge's evaluation of credibility is the most important trial-level determination in any *Batson* inquiry.³⁰⁸ Prosecutors easily can assert nonracial reasons for striking any juror; thus, the primary issue for the court is whether those reasons are genuine or pretextual.³⁰⁹ Evaluating the credibility of prosecutors in *Batson* cases, however, is not a simple matter. Because of unconscious racism, trial judges may not recognize racial discrimination.³¹⁰ Furthermore, many of the reasons that prosecutors advance are subjective and vague.³¹¹ Some of these reasons involve subtle actions by the prospective juror; others involve no objective indicia at all.³¹² Trial judges are unlikely to observe these claimed grounds for the alleged discriminatory peremptory challenges unless they search for them. Therefore, the trial judge must have an adequate opportunity to evaluate the prosecutor's sincerity when the prosecutor explains her peremptory challenges.

For more than two centuries the belief that "no safeguard for testing the value of human statements is comparable to that furnished by cross-examination" has pervaded Anglo-American law.³¹³ In light of the centrality of the credibility issue to the *Batson* inquiry and the inherent difficulty of evaluating the prosecutor's credibility, the North Carolina Supreme Court should permit defendants to examine the prosecutors once the prosecutors have explained their

306. See *id.* (finding "no reason why the defendant could not have offered evidence to strengthen his case after the State had made its showing"); see also *State v. Porter*, 326 N.C. 489, 497, 391 S.E.2d 144, 150 (1990) (expressly holding that after State advances its reasons for the disputed challenges defendant has right of surrebuttal to show that those reasons are pretextual); *State v. Green*, 324 N.C. 238, 240-41, 376 S.E.2d 727, 728 (1989) (same).

307. The Illinois Supreme Court has stated that because the prosecutor, an officer of the court, is under a high professional obligation to speak truthfully, no need for cross-examination arises. *People v. Young*, 128 Ill. 2d 1, 24-25, 538 N.E.2d 453, 459 (1989), cert. denied, 110 S. Ct. 3290 (1990). This argument is unpersuasive. Prosecutors are also under a high obligation to obey the fourteenth amendment, to follow the dictates of the United States Supreme Court, and not to exercise peremptory challenges on account of race. The need for *Batson* shows that some prosecutors, like other lawyers, breach their professional obligations. Furthermore, some prosecutors simply are unaware of the racial underpinnings of their actions. See *Batson*, 476 U.S. at 106 (Marshall, J., concurring) (unconscious racism may color prosecutor's judgments); Johnson, *supra* note 287, at 1026-36 (same). Persons who might be denied their constitutional rights as a result of prosecutors' conscious or unconscious racism are entitled to adequate procedures to protect those rights. Thus, prosecutors' obligation to speak truthfully is no substitute for adequate procedures to evaluate their credibility.

308. See *Batson*, 476 U.S. at 98 n.21 (trial judge's findings will depend largely on evaluation of prosecutor's credibility); *Jackson*, 322 N.C. at 255, 368 S.E.2d at 840 (reviewing court should give trial court's findings great deference because it depends on credibility).

309. See *supra* notes 252-56 and accompanying text.

310. See *supra* note 287.

311. See *supra* text accompanying notes 253-54.

312. *Id.*

313. 5 J. WIGMORE, *supra* note 182, § 1367, at 32 (Chadbourne rev. 1974); see E. CLEARY, *supra* note 119, § 19, at 47. So essential is cross-examination to the accuracy and completeness of testimony, that it is a right and not merely a privilege in trials. E. CLEARY, *supra* note 119, § 19, at 47. Wigmore described cross-examination as "the greatest legal engine ever invented for the discovery of truth" and "the great and permanent contribution of the Anglo-American system of law to improved methods of trial procedure." 5 J. WIGMORE, *supra* note 182, § 1367, at 32 (Chadbourne rev. 1974).

peremptory challenges.³¹⁴ The *Jackson* court's concern about and emphasis on the disruption that might result from adversarial *Batson* hearings, while understandable at first glance, are unwarranted. Permitting examination of the prosecutor would not add substantially to the disruption of the trial, since *Batson* hearings already involve an interruption of the trial. By the time the defendant examines the prosecutor, the trial already would have been stopped for the defendant to argue the existence of a prima facie case and for the prosecutor to explain the disputed peremptory strikes. Furthermore, the defendant already has the right to introduce further evidence in surrebuttal; examination of the prosecutor would not take significantly more time. Trial judges can prevent undue disruption by exercising their traditional powers to preclude examination that is badgering, repetitive, or a mere fishing expedition. Moreover, according to Professor Raphael, most courts simply have not experienced undue disruptions as a result of adversarial *Batson* hearings.³¹⁵ Most important, a small, additional disruption in the trial cannot justify withholding from defendants the legal system's most effective tool for exposing the falsity of prosecutors' assertions when the constitutional rights of minority defendants and prospective jurors are at stake.

At the very least, no rationale justifies the North Carolina Supreme Court's absolute prohibition of examining the prosecutor. When a prosecutor must explain his peremptory challenges in a hearing on remand from an appellate court, disruption of the trial is not a concern. In this situation no reason exists for an absolute prohibition against cross-examining the prosecutor because the problem at the heart of the *Jackson* court's rationale for the prohibition is absent.³¹⁶ Moreover, the trial judge is in the best position to determine the value of examining the prosecutor and the disruption that might result from that examination. The trial judge, therefore, should at least have the discretion to permit examination of the prosecutor.

3. Harmless Error

The North Carolina Supreme Court also has addressed the procedural issue of harmless error under *Batson*. In *State v. Robbins*³¹⁷ the court rejected the defendant's *Batson* claim in part because the defendant failed to exhaust his allotted peremptory challenges, the theory apparently being that the defendant still could have removed jurors with whom he was dissatisfied and thus was not prejudiced by any *Batson* errors.³¹⁸ This holding once again misconstrues the

314. One commentator has suggested that defendants ordinarily should be permitted to examine the prosecutor, with exceptions made only if the prosecutor demonstrates that his trial strategy would be "compromised substantially" as a result of the examination. See Raphael, *supra* note 178, at 338.

315. Professor Raphael reported: "The experience of most courts . . . indicates that the *Jackson* court was in error in fearing that cross-examination of a prosecutor would be disruptive" Raphael, *supra* note 178, at 338.

316. See Note, *supra* note 142, at 205.

317. See *supra* notes 160-76 and accompanying text.

318. *State v. Robbins*, 319 N.C. 465, 495, 356 S.E.2d 279, 297 (citing *State v. Wilson*, 313 N.C. 516, 524, 330 S.E.2d 450, 457 (1985)), *cert. denied*, 484 U.S. 918 (1987); see also *State v. Davis*, 325

constitutional principles upon which *Batson* is based.

The North Carolina Supreme Court's requirement that parties exercise all of their peremptory challenges to show prejudice derives from cases in which defendants claimed the trial court improperly denied their challenges for cause. The court held in those cases that the defendants were not prejudiced by the trial judge's refusal to excuse the jurors because, if they truly wanted the jurors off the jury, the defendants could have removed them with their remaining peremptory challenges.³¹⁹ These cases are inapposite, however, to attacks by defendants on prosecutorial peremptory challenges under *Batson*. When defendants make *Batson* claims, they are not necessarily asserting particular dissatisfaction with the persons who actually served on the jury. Rather, the claim simply is that some persons were removed unconstitutionally from the jury. This claim is especially true because, even if the defendants are not harmed by prosecutors' discriminatory peremptory challenges, such challenges deny the excluded jurors equal protection of the laws and undermine public confidence in the legal system.³²⁰ "Prejudice" for the purposes of *Batson*, therefore, is present when any person is peremptorily challenged on account of race. The fact that defendants still had peremptory challenges remaining and could have used them to strike persons accepted by the prosecutor does not remove this prejudice. Indeed, the United States Supreme Court has noted that because jury selection goes to the very integrity of the legal system, harmless error analysis simply does not apply.³²¹ Thus, the North Carolina Supreme Court's reliance on its previous cases dealing with prejudice and the use of peremptory challenges is misplaced.

D. Justifications for Suggested Doctrinal Revisions

North Carolina's implementation of *Batson* has been marked by a constrictive view of the decision and a misapprehension of *Batson*'s operation and constitutional underpinnings. The courts have misconstrued the operation of the prima facie case by improperly considering challenged jurors' voir dire responses when determining whether an inference of discrimination exists. The courts

N.C. 607, 620, 386 S.E.2d 418, 424 (1989) (finding no prejudice when defendant had three peremptory challenges remaining), *cert. denied*, 110 S. Ct. 2587 (1990).

319. See, e.g., *State v. Quesinberry*, 319 N.C. 228, 235, 354 S.E.2d 446, 450 (1987), *vacated on other grounds*, 110 S. Ct. 1465 (1990); *State v. Avery*, 315 N.C. 1, 21, 337 S.E.2d 786, 797 (1985); see also *State v. Wilson*, 313 N.C. 516, 524-25, 330 S.E.2d 450, 457 (1985) (defendant dissatisfied with jurors recruited by sheriff during middle of jury selection). These cases applied the North Carolina jury selection procedures statute, which provides in part: "In order for a defendant to seek reversal of the case on appeal on the ground that the judge refused to allow a challenge made for cause, he must have:

(1) Exhausted the peremptory challenges available to him[.]" N.C. GEN. STAT. § 15A-1214(h) (1) (1988).

320. *Batson*, 476 U.S. at 87.

321. See *Gray v. Mississippi*, 481 U.S. 648, 668 (1987) (*Witherspoon* error resulting in improper exclusion of jurors for cause in death penalty case is never harmless); see also *Mitchell v. State*, 295 Ark. 341, 351, 750 S.W.2d 936, 941 (1988) (Despite evidence of overwhelming guilt, *Batson* error requires reversal, since "[w]e are concerned here with prejudice to the system of justice."); cf. *State v. Coffield*, 320 N.C. 297, 301, 303, 357 S.E.2d 622, 624-25, 626 (1987) (question is not whether racial discrimination in selection of grand jury foreperson affected the outcome of the proceedings, but rather whether there was racial discrimination at all).

readily have accepted any facially neutral reason put forth by prosecutors as sufficient to rebut a *prima facie* case, and have paid excessive deference to trial court findings of *Batson* issues.

The courts' narrow views of *Batson* may be explained by North Carolina's historical support for the peremptory challenge. North Carolina was one of the first states to permit the exercise of peremptory challenges and, more particularly, prosecutorial peremptory challenges.³²² As a result, the North Carolina courts have sought to give life to *Batson* without meaningfully changing the exercise of the peremptory challenge. This phenomenon has been most evident in the wide range of rebuttal reasons the courts have accepted; they have been unwilling to reject any facially nonracial explanation that, by definition, is ordinarily a proper basis for a peremptory challenge. The result, however, has been a scheme that has been ineffective in answering *Batson's* challenge to erradicate jury selection discrimination. The inherent conflict between the equal protection clause and the peremptory challenge means that equal protection cannot be guaranteed without intrusions on the peremptory challenge. Unless the North Carolina courts revise their approach to *Batson* and promote in practical ways the decision's fundamental goal of eliminating discrimination, *Batson* protection in North Carolina may become permanently illusory.

The revisions suggested in this Comment are justified for several reasons, despite their limitations on North Carolina prosecutors' traditional uses of peremptory challenges. Most important, the North Carolina Constitution provides special protections against discrimination in jury selection; indeed, they are stronger than those contained in the federal constitution. Article 1, section 26 of the North Carolina Constitution expressly provides that "[n]o person shall be excluded from jury service on account of sex, race, color, religion, or national origin."³²³ This provision is distinct from, and in addition to, the state equal protection clause.³²⁴ In *State v. Cofield*³²⁵ the supreme court construed section 26 in the context of grand jury foreperson selection. In powerful language, the court noted that by adopting this constitutional provision the people of North Carolina

have declared that they will not tolerate the corruption of their juries by racism, sexism and similar forms of irrational prejudice. They have recognized that the judicial system of a democratic society must operate evenhandedly if it is to command the respect and support of those subject to its jurisdiction.³²⁶

322. See *supra* notes 46-54 and accompanying text. In addition, the North Carolina trial tradition has been to promote freedom for its attorneys in selecting juries. See *supra* note 54.

323. N.C. CONST. art. I, § 26.

324. See *id.* § 19. This section provides in pertinent part: "No person shall be denied the equal protection of the laws; nor shall any person be subjected to discrimination by the State because of race, color, religion, or national origin." *Id.*

325. 320 N.C. 297, 357 S.E.2d 622 (1987) (*Cofield I*).

326. *Id.* at 302, 357 S.E.2d at 625. The supreme court found the need for racially neutral procedures so imperative that in *Cofield II* the court again reversed the defendant's conviction because of the grand jury foreperson selection process employed, even though the court was "satisfied that there was not the slightest hint of racial motivation" in the judge's selection of the foreperson. *State v. Cofield*, 324 N.C. 452, 459-60, 379 S.E.2d 834, 839 (1989) (*Cofield II*).

The court explicitly acknowledged that this policy applies not only to the selection of grand jury forepersons, but also to the selection of petit jurors.³²⁷ As Justice Mitchell explained in his concurring opinion: "[I]t is clear beyond any doubt that this section of our Constitution was intended as an absolute guarantee that all citizens of this State would participate fully in the honor and obligation of jury service in all forms; as petit jurors, grand jurors, and as foremen of the grand jury."³²⁸ The supreme court later discussed the specific relationship of peremptory challenges to section 26 in *Jackson v. Housing Authority*.³²⁹ The *Housing Authority* court observed: "Although long embedded in our common law, the use of peremptory challenges is based upon statutory authority and is not of [state] constitutional dimension. Therefore, the statutory authority to exercise peremptory challenges must yield to . . . constitutional mandate . . ."³³⁰ The mandate of the people of North Carolina thus requires that the North Carolina appellate courts rigorously protect defendants and prospective jurors from invidious discrimination, even if it means contraction of a prosecutor's use of peremptory challenges.

This Comment's recommended revisions also are justified because the peremptory challenge simply is not of federal constitutional magnitude,³³¹ and reasonable limitations on its use are essential to the protection of fourteenth amendment rights. Furthermore, proponents of the prosecutorial peremptory challenge overstate its historical significance. Professor Van Dyke has noted that peremptory challenges "have been subject to abuse from the time juries were first introduced in England."³³² This phenomenon was true despite peremptory challenges in England being much less susceptible to discriminatory use than those in the United States because of the homogeneity of English society; in the heyday of the English peremptory challenge, only propertied males could serve as jurors.³³³ As times changed and English jurors became more diverse, Parliament limited, and eventually abolished, the peremptory challenge.³³⁴ Similarly, the historical significance of peremptory challenges in North Carolina is less than overwhelming. Although the North Carolina General Assembly in 1827 granted prosecutors the right to exercise peremptory challenges, it was not until one hundred fifty years later that they gave prosecutors the same

327. *Coffield I*, 320 N.C. at 303, 357 S.E.2d at 626; see also *State v. Mitchell*, 321 N.C. 650, 653, 365 S.E.2d 554, 556 (1988) (recognizing that *Batson* and *Coffield I* stand for analogous propositions that potential jurors may not be excluded nor grand jury forepersons selected on discriminatory grounds).

328. *Coffield I*, 320 N.C. at 310, 357 S.E.2d at 630 (Mitchell, J., concurring).

329. 321 N.C. 584, 364 S.E.2d 416 (1988). *Jackson* prohibited the exercise of peremptory challenges on account of race in civil cases. *Id.* at 585, 364 S.E.2d at 417.

330. *Id.* As early as 1887, the North Carolina Supreme Court recognized that "if [the prosecutor's right to stand jurors aside] had been [abused] . . . such abuse would have warranted a recall of the permission." *State v. Sloan*, 97 N.C. 499, 502-03, 2 S.E. 666, 668 (1887).

331. See *Batson*, 476 U.S. at 108 (Marshall, J., concurring); *Stilson v. United States*, 250 U.S. 583, 586 (1919). The Court has held that the right to peremptory challenge may be withheld without impairing the constitutional guarantee of an impartial jury. *Frazier v. United States*, 335 U.S. 497, 505 n.11 (1948); *United States v. Wood*, 299 U.S. 123, 145 (1936).

332. J. VAN DYKE, *supra* note 2, at 147.

333. See Alschuler, *supra* note 7, at 165.

334. See Gobert, *supra* note 27, at 528-29.

number of peremptory challenges as defendants.³³⁵

The final justification for this Comment's suggested revisions to North Carolina's *Batson* law is that the practical value of the peremptory challenge is questionable.³³⁶ In a 1978 study, for example, the prospective jurors whom the prosecutor peremptorily challenged were as likely to favor convictions as were the jurors actually selected.³³⁷ It follows that the most important function of the peremptory challenge is not to facilitate the selection of juries that are actually impartial, but rather to foster the perception of impartiality and thus promote confidence in the criminal justice system.³³⁸ Both *Batson* and *Cofield* teach that when peremptory challenges are used for discriminatory reasons, "[t]he harm . . . extends . . . to touch the entire community" because unconstitutionally motivated challenges "undermine public confidence in the fairness of our system of justice."³³⁹ Limiting the use of the peremptory challenge to protect against discrimination, therefore, is not only justified, but is the only way that the peremptory challenge may perform its function of promoting confidence in the criminal justice system.

V. CONCLUSION

Five years after *Batson*, no North Carolina defendant has attacked prosecutorial peremptory challenges successfully on equal protection grounds. *Batson* has been rendered ineffective in North Carolina by the North Carolina appellate courts' efforts to minimize *Batson*'s impact on the peremptory challenge, a device historically employed in North Carolina. The American criminal justice system is not big enough for both the equal protection clause and the traditional, arbitrary and capricious peremptory challenge; the equal protection clause and the peremptory challenge are inherently contradictory. Consequently, the North Carolina courts must revise their approach to *Batson* and promote more effectively *Batson*'s primary goal of eradicating the discriminatory use of peremptory challenges. As Justice Frye of the North Carolina Supreme Court recognized: "[I]t is the province of the courts [of North Carolina] to ensure that [peremptory challenges] are used in such a manner not offensive to the constitutional rights of our citizens."³⁴⁰

PAUL H. SCHWARTZ

335. See *supra* note 53 and accompanying text.

336. See R. HASTIE, S. PENROD & N. PENNINGTON, *INSIDE THE JURY* 129 (1983) ("no hard evidence that current systematic jury selection methods are useful in typical felony cases"); Note, *The Case for Striking Peremptory Strikes*, *supra* note 140, at 212-13.

337. Zeisel & Diamond, *The Effect of Peremptory Challenges on Jury and Verdict: An Experiment in a Federal District Court*, 30 STAN. L. REV. 491, 513-18 (1978).

338. See *supra* note 33 and accompanying text.

339. *Batson*, 476 U.S. at 87.

340. *State v. Jackson*, 322 N.C. 251, 260, 368 S.E.2d 838, 843 (1988) (Frye, J., concurring), *cert. denied*, 490 U.S. 1110 (1989).

NOTES

North Carolina Air Toxics Regulations

The State of North Carolina recently adopted regulations for the control of toxic air pollutants (TAPs) within its borders.¹ The regulations are the latest step in an air toxics program the State began developing in 1985 in response to reports of increasing in-state emissions of toxic pollutants.² Although the federal government purported to control air toxics through section 112 of the Clean Air Act (CAA),³ federal regulation had proven to be virtually unworkable. Not only did the federal statutory scheme place unrealistic deadlines upon the Environmental Protection Agency (EPA),⁴ but it also required the Agency to regulate hazardous pollutants with an "ample margin of safety,"⁵ a particularly arduous task considering no conclusive scientific data exists for determining "safe" exposure levels for many pollutants.⁶ As a practical matter, individual states had to regulate air toxics on their own.⁷

In February 1990, the North Carolina Environmental Management Commission (EMC) adopted a program regulating 105 toxic air pollutants.⁸ Because the air toxic guidelines went into effect as of May 1990,⁹ there has been little opportunity to assess the effectiveness of the North Carolina measures or their impact on industry. Meanwhile, in November 1990, the federal government revitalized its regulatory scheme to address more effectively the hazardous air pollution problem.¹⁰ Title III of the 1990 CAA lists 189 TAPs that the EPA must regulate by source categories over the next ten years.¹¹ The Act delegates

1. NORTH CAROLINA ENVIRONMENTAL DEFENSE FUND, TOXIC AIR EMISSIONS IN NORTH CAROLINA: AN UPDATE FOR 1989 2 (2d ed. Oct. 1990) [hereinafter NCEDF]. The North Carolina Environmental Defense Fund analyzes air toxic emissions in North Carolina and prepares yearly data reports. *Id.* at 1.

2. *Id.* at 6. In 1988, North Carolina ranked twelfth nationally in total air toxics emissions by weight and exceeded the emissions of California and New Jersey. *Id.* at 22.

3. 42 U.S.C. § 7412 (1988), amended by Clean Air Act of 1990, Pub. L. No. 101-549, § 301, 104 Stat. 2399, 2531 (1990) (codified at 42 U.S.C.A. § 7412 (West Supp. 1991)).

4. See *infra* text accompanying notes 21-22.

5. 42 U.S.C. § 7412(b)(1)(B) (1988).

6. See *infra* text accompanying notes 25-26.

7. See 1 DIVISION OF ENVIRONMENTAL MANAGEMENT, DEPARTMENT OF ENVIRONMENT, HEALTH, AND NATURAL RESOURCES, REPORT OF PROCEEDINGS OF PUBLIC HEARINGS ON PROPOSED AMENDMENTS TO REGULATIONS 15A NORTH CAROLINA ADMINISTRATIVE CODE 2D.0535 AND .0902 AND 2H.0602, at I-1 (Aug. 1989) [hereinafter HEARINGS]. As of 1986 approximately 38 states had regulated or were in the process of regulating toxic air pollutants. 4 *id.* app. B at VI-60 (survey of state air toxics programs).

8. N.C. ADMIN. CODE tit. 15A, r. 2D.1104 (Aug. 1990). The air toxics regulations controlled emissions for 84 of the toxic substances as of May 1, 1990, and emissions for the remaining 21 beginning May 1, 1991. *Id.* Vigorous opposition by various industries as well as the requirement of a lengthy economic impact assessment delayed the promulgation for several years. NCEDF, *supra* note 1, at 2; see 1 HEARINGS, *supra* note 7, at I-2.

9. N.C. ADMIN. CODE tit. 15A, r. 2D.1104 (Aug. 1990).

10. See Clean Air Act of 1990, Pub. L. No. 101-549, § 301, 104 Stat. 2399, 2531-74 (1990) (codified at 42 U.S.C.A. § 7412 (West Supp. 1991)).

11. 42 U.S.C.A. § 7412(b)(1) (West Supp. 1991).

the task of developing standards of "maximum achievable control technology" (MACT) for various source categories to the EPA, with such standards to be developed on an industry-wide basis.¹² In light of the federal government's new air toxics regulations, the question arises whether the EMC will administer the North Carolina regulations in their present form or as a supplement to the federal standards, or whether the Commission will have to revise or abandon the North Carolina regulations to meet the new federal requirements.

This Note begins by surveying the development of the North Carolina air toxics regulations and summarizing the substantive provisions of both Title III of the 1990 CAA amendments and the North Carolina regulations, distinguishing the regulatory approaches of each.¹³ The Note then suggests that North Carolina retain its present regulations to remedy the existing problem of toxic air pollution, demonstrating that these regulations can coexist with the new federal system of air toxics regulations. The Note further contends that North Carolina should not rely on its regulatory scheme as merely a protective measure in the event of the EPA's default or failure of the federal air toxics program, but should begin immediate and complete implementation of its own toxic air regulations to ensure that its goals for cleaner air are attained. Unlike the federal regulations, North Carolina's air toxics guidelines may be better suited to address specific problem areas by regulating on a facility-by-facility basis and providing for additional attention where multiple sources are located. The Note concludes that, although the uniform regulatory approach used by the CAA may provide the best "fit" for a national solution to the air toxics problem, the federal program may overlook localized problem areas. North Carolina must retain its more flexible regulations, implementing them in conjunction with the CAA to ensure more comprehensive control of toxic air pollution.

I. EVOLUTION OF THE NORTH CAROLINA AIR TOXICS REGULATIONS

Congress designed former section 112 of the federal CAA to eliminate the TAP problem; unfortunately, the statute fell far short of this goal.¹⁴ Section 112 required that the EPA "list" certain toxic substances for which it intended to promulgate standards.¹⁵ Although there were dozens of potential candidates for regulation, the EPA listed only eight pollutants in the twenty years after section 112 was enacted.¹⁶

12. *Id.* § 7412(d),(e); see J. QUARLES & W. LEWIS, *THE NEW CLEAN AIR ACT: A GUIDE TO THE CLEAN AIR PROGRAM AS AMENDED IN 1990*, at 31 (1990).

13. The North Carolina regulations follow a predominately health-based approach, while the federal regulations, in contrast, follow a technology-based approach. For an explanation of each regulatory method, see *infra* text accompanying notes 103-09.

14. See 1 *HEARINGS, supra* note 7, at I-1 (noting the "relatively limited ability" of § 112 of the old CAA to control toxic air pollutants).

15. 42 U.S.C. § 7412(b) (1988), amended by Clean Air Act of 1990, Pub. L. No. 101-549, § 301, 104 Stat. 2399, 2531 (1990) (codified at 42 U.S.C.A. § 7412 (West Supp. 1991)); see Note, *Toward Sensible Regulation of Hazardous Air Pollutants Under Section 112 of the Clean Air Act*, 63 N.Y.U. L. REV. 612, 617 (1988).

16. Marchant & Danzeisen, "Acceptable" Risk for Hazardous Air Pollutants, 13 HARV. ENVTL. L. REV. 535, 536 (1989). The eight pollutants currently listed are asbestos, benzene, beryllium, coke

There are several reasons why the EPA has allowed many potentially hazardous pollutants to go unregulated. Part of the problem stems from the statute itself: the statute's time constraints made expansive listing impracticable. Section 112 required the EPA to list substances that the EPA Administrator decided "may reasonably be anticipated to result in an increase in mortality or an increase in serious irreversible, or incapacitating reversible illness."¹⁷ Within 180 days after listing a pollutant, the Administrator had to publish proposed regulations and give notice of public hearings for comment to be held within thirty days.¹⁸ Within 180 days of the proposal publication, the EPA had to promulgate final emission standards¹⁹ setting "the level which in [the Administrator's] judgment provide[d] an ample margin of safety to protect the public health from such hazardous air pollutants."²⁰ These time constraints forced the EPA to conduct the extensive research necessary to determine appropriate emissions standards for a pollutant without time for a thorough assessment.²¹ Moreover, after each listing the EPA had to identify all emission sources, investigate various kinds of control strategies, and obtain information on compliance costs before it could publish final emissions standards. Such preparation for final emission standards usually takes a minimum of two years.²² The EPA, knowing it could not present final emissions standards for numerous TAPs within six months, began slowing the process by listing only one pollutant at a time.²³

Although unrealistic time constraints may have delayed the listing process somewhat, a more serious problem stemmed from the "ample margin of safety"²⁴ concept. Most toxic substances regulated under section 112 are carcinogens that have no known "threshold" below which adverse human health effects do not occur.²⁵ This means that no level of exposure has been identified as inherently "safe." If no safe exposure level exists, then section 112 literally required complete and immediate prohibition of all emissions of a pollutant once the EPA listed it.²⁶ Such an all-or-nothing approach caused the EPA to choose "nothing" more often than not. Complete elimination of many pollutants may be technologically impossible. Moreover, outright bans could induce the shutdown of major industries, resulting in massive social dislocation.²⁷ To avoid such dire consequences and in response to pressure from industry, the EPA in-

oven emissions, inorganic arsenic, mercury, radionuclides, and vinyl chloride. National Emission Standards for Hazardous Air Pollutants, 40 C.F.R. § 61.01 (1990).

17. 42 U.S.C. § 7412(a)(1).

18. *Id.* § 7412(b)(1)(B).

19. *Id.*

20. *Id.*

21. Graham, *The Failure of Agency-Forcing: The Regulation of Airborne Carcinogens Under Section 112 of the Clean Air Act*, 1985 DUKE L.J. 100, 123-24.

22. *Id.* at 124.

23. *Id.* at 124-25. Failure to list a pollutant within a reasonable time is legally defensible because § 112 places no time limit on the listing of a pollutant. *Id.* at 125.

24. 42 U.S.C. § 7412(b)(1)(B).

25. Marchant & Danzeisen, *supra* note 16, at 536-37; see *Industrial Union Dept., AFL-CIO v. American Petroleum Inst.*, 448 U.S. 607, 613, 638 n.43 (1980) (recognizing that no "safe exposure level" exists for certain pollutants).

26. See Marchant & Danzeisen, *supra* note 16, at 537.

27. *Id.*

stead chose to regulate only a few pollutants.²⁸ To avert widespread industry shutdown, the EPA based the few standards it set on "best available" control technology.²⁹ In short, the EPA's steadfast refusal to entertain major industry closings as a solution to the TAP problem rendered Congress's strict health-based approach to toxic emissions limitations a dead-end path to progressive pollution control.

Once the federal system proved inadequate, states began filling in the gaps with their own air toxics legislation. By 1987, several states, including California and New Jersey, already had remedial legislation in place.³⁰ North Carolina, however, had not reacted yet to the growing toxic air emissions problem. In 1985, North Carolina decided to take action for several reasons. First, heightened public concern about toxic chemicals in the atmosphere necessitated action.³¹ Second, the State was aware that the EPA had failed to regulate these pollutants effectively under section 112.³² Third, scientific reports indicated an increase in the release of numerous TAPs—including carcinogens, mutagens, and teratogens—into North Carolina skies.³³

With funding from the EPA, the North Carolina Division of Environmental Management (DEM) began a survey of sources of North Carolina toxic air pollutants.³⁴ After compiling a list of pollutants, DEM initially sought to emulate other state programs by proposing a "strict factored approach" to obtain ambient air levels.³⁵ This approach uses the threshold limit value (TLV)³⁶ of a pollutant multiplied by a particular safety factor.³⁷ Industry strongly objected to this approach, arguing that the DEM applied the TLVs out of context and that a constant safety factor was inappropriate under the circumstances.³⁸

28. *Id.*

29. Note, *supra* note 15, at 615. Best available technology (BAT) standards consider both technological feasibility and economic cost of implementation. *Id.* The statutory language of § 112, however, precludes tradeoffs between health factors and nonhealth feasibility factors. See 42 U.S.C. 7412(b)(1)(B); Note, *supra* note 15, at 613. In fact, Congress envisioned the possible shutdown of industries in the absence of control technology available to assure zero emissions. See 116 CONG. REC. 42,385 (1970). At least one court has upheld the EPA's decision to incorporate feasibility factors in setting emissions standards. See *Natural Resources Defense Council v. EPA*, 824 F.2d 1146, 1157 (D.C. Cir. 1987).

30. NORTH CAROLINA ENVIRONMENTAL DEFENSE FUND, TOXIC AIR EMISSIONS IN NORTH CAROLINA: AN UPDATE FOR 1988, at 21 (1989).

31. See 1 HEARINGS, *supra* note 7, at I-3 (listing reasons for instituting a North Carolina air toxics program).

32. *Id.*

33. *Id.* A mutagen is a substance that tends to increase the frequency of alterations in genetic or hereditary material. 2 WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 1492 (1976). A teratogen is a substance that causes developmental malformations in fetuses. *Id.* at 2358.

34. 1 HEARINGS, *supra* note 7, at I-3.

35. *Id.* at I-2. For an explanation of a modified (as opposed to a strict) factored threshold limit value approach, see *infra* note 54.

36. A threshold limit value (TLV) is the "time-weighted average concentration, for a normal 8-hour workday and 40-hour workweek, to which nearly all workers may be repeatedly exposed . . . without adverse effect." AMERICAN CONFERENCE OF GOVERNMENT AND INDUSTRIAL HYGIENISTS, THRESHOLD LIMIT VALUES AND BIOLOGICAL EXPOSURE INDICES FOR 1988-1989, at 4 (1988) [hereinafter ACGIH]; NCEDF, *supra* note 1, app. C.

37. The strict factored approach applies only one safety factor (1/200) to the threshold limit value of a pollutant to obtain acceptable ambient levels (AALs).

38. 1 HEARINGS, *supra* note 7, at I-2.

Therefore, industry suggested that North Carolina establish an independent scientific panel to develop acceptable ambient levels (AAL) for the pollutants.³⁹ The North Carolina Academy of Sciences established the Air Toxics Panel, charging it with reviewing the list of proposed toxic air pollutants recommending a suitable approach for determining AALs.⁴⁰

In reviewing the list of pollutants the Panel limited its choices to chemicals that the American Conference of Government and Industrial Hygienists already had assigned a TLV, that the EPA had listed as carcinogens in the category of Group A (human carcinogens) or Group B (probable human carcinogens),⁴¹ or that North Carolina Division of Health Services considered to be of public health concern. The Panel further limited its work to chemicals for which there was potential exposure in North Carolina.⁴² Additionally, the scientific panel proposed a "modified factored approach"⁴³ in determining the AALs.

After these initial studies, the EMC entered into a contract with Radian Corporation to do an economic impact study of the proposed regulations.⁴⁴ At public hearings on the proposed regulations, various industrial representatives argued that the economic impact statement was incomplete and that it underestimated the regulations' impact by "ignoring the loss of competitiveness and the resulting loss of profitability that would accompany a slight increase in prices for

39. *Id.*

40. *Id.*

41. Group A human carcinogens are those chemicals for which there is sufficient evidence from epidemiological studies to support a causal association between exposure to the agent and cancer. Group B probable human carcinogens are chemicals for which there is sufficient evidence of carcinogenicity from animal studies and limited or inadequate evidence from epidemiological studies. EPA Guidelines for Carcinogen Risk Assessment, 51 Fed. Reg. 33,992, 34,000 (1986). North Carolina's AALs for carcinogens are based on unit risk factors developed by the EPA's Carcinogen Assessment Group (CAG). Unit risk factor is an estimate of the incremental lifetime cancer risk over the background occurring in a population in which all individuals are exposed continuously to a concentration of 1 ug/m³ of the agent in the air that they breathe. EPA Proposed Guidelines for Carcinogen Risk Assessment, 49 Fed. Reg. 46,294, 46,299 (1984); 2 HEARINGS, *supra* note 7, at V-650.

42. NORTH CAROLINA ACADEMY OF SCIENCES, REPORT AND RECOMMENDATIONS OF THE AIR TOXICS PANEL OF THE NORTH CAROLINA ACADEMY OF SCIENCES 6 (Sept. 1986), *reprinted in* 4 HEARINGS, *supra* note 7, at VI-7 [hereinafter AIR TOXICS PANEL]. Limiting the lists of air toxics to those that pose an actual threat to human health in North Carolina saves both time and money. The EMC's resources should be allocated properly to control only those pollutants that North Carolina's industries actually emit.

43. 1 HEARINGS, *supra* note 7, at I-2; *see infra* note 54.

44. 1 HEARINGS, *supra* note 7, at I-2; *see* RADIAN CORPORATION, ASSESSMENT OF THE ECONOMIC IMPACTS OF NORTH CAROLINA'S PROPOSED AIR TOXICS REGULATION (Apr. 1988) [hereinafter RADIAN]. Section 143-215.107(f) of the North Carolina General Statutes provides that in the event the federal government has not adopted regulations on a matter, the EMC may not adopt regulations until it considers "an assessment of the economic impact of the proposed standards." N.C. GEN. STAT. § 143-215.107(f) (1990).

The North Carolina legislature adopted this section—called the Hardison Amendment after Senator Harold Hardison—in 1975. The statute constrains the EMC's development of air quality standards in two ways. First, the EMC may not adopt standards more stringent than those promulgated by the federal government. Second, where the federal government has not regulated, the EMC must conduct an economic impact assessment before the standards are adopted. *Id.* The economic impact study must be part of the rule-making record and include an "estimate of the economic and social costs to commerce and industry, units of local government, and agriculture necessary to comply with the proposed standards and an examination of the economic and social benefits of such compliance." *Id.*

products produced by North Carolina industries."⁴⁵ Further, they argued that the assessment failed to evaluate certain social impacts of the regulations, including job losses and jobs that would not be created.⁴⁶ Environmentalists, in contrast, asserted that the statute requires only an "estimate" of regulatory costs, and does not specify with any detail what the assessment should contain, how extensive it needs to be, or what type of comparisons need to be made.⁴⁷ Moreover, environmentalists argued that the DEM compiled a legitimate report, and that the study could be reasonably relied upon given the limited resources available.⁴⁸ Following the public hearings, the EMC approved the economic impact statement⁴⁹ and the new guidelines for TAPs became effective as of May 1, 1990.⁵⁰

The North Carolina regulations contain a list of TAPs that the Air Toxics Panel divided into noncarcinogens (acute irritants, acute toxicants, and chronic toxicants) and carcinogens.⁵¹ The Panel chose acceptable ambient levels for both types of pollutants. For noncarcinogens, the Panel applied various safety factors to TLVs to obtain a concentration that protects the human population exposed outside of the abutting property line of any emissions source.⁵² These safety factors consider variability in human susceptibility, continuous exposure over a 168-hour week as compared to a 40-hour week, uncertainties inherent in studies of chronic effects, and the severity of effects.⁵³ The Panel then applied these factors to the TLVs to determine the concentrations acceptable as ambient air levels.⁵⁴

45. 2 HEARINGS, *supra* note 7, at V-461 (comments by Charles Case on behalf of the Chemical Industry Council of North Carolina). Radian Corporation's impact study reported that under worst case conditions less than three percent of 325 North Carolina facilities surveyed would experience significant economic hardship, and 19% would be able to comply with minimal added controls or by raising stack height. See RADIAN, *supra* note 44, at xiii.

46. 2 HEARINGS, *supra* note 7, at V-461 (comments by Charles Case on behalf of Chemical Industry Council of North Carolina).

47. See N.C. GEN. STAT. § 143-215.107(f) (1990).

48. 2 HEARINGS, *supra* note 7, at V-550 (comments at public hearing by Steven Levitas on behalf of the North Carolina Environmental Defense Fund). Environmentalists assert that limited resources hamper North Carolina's ability to devote the time and energy necessary to achieve a dynamic air toxics program. NCEDF, *supra* note 1, at 23.

49. The North Carolina Attorney General has ruled that the EMC has the ultimate responsibility and authority to determine whether an economic impact assessment is sufficiently complete. HEARINGS, *supra* note 7, at I-88.

50. N.C. ADMIN. CODE tit. 15A, r. 2D.1104 (Aug. 1990).

51. *Id.* Acute irritants are chemicals that cause irritation at the site of contact immediately following exposure of eight hours or less. Acute toxicants are chemicals that cause adverse effects at sites distant from the point of exposure within eight hours. Chronic toxicants are chemicals that cause adverse effects after multiple or prolonged exposures. 4 HEARINGS, *supra* note 7, at VI-32 to VI-33.

52. 1 HEARINGS, *supra* note 7, at I-2.

53. AIR TOXICS PANEL, *supra* note 42, at 11-16.

54. 1 HEARINGS, *supra* note 7, at I-2. The application of various safety factors to TLV values is called a "modified factored approach." The use of several factors is designed to account for the multiple differences between community and occupational exposures. For example, the Air Toxics Panel decided on a factor of 10 to account for the variability in human susceptibility. AIR TOXICS PANEL, *supra* note 42, at 12. The TLV does not always reflect this variation because workers are usually healthy adults. In contrast, the population at large includes children, elderly persons, and other sensitive subgroups. *Id.* The Panel also adjusted for continuous community exposure by adopting a factor of four to account for the difference between a 40-hour workweek and a 168-hour

The Air Toxics Panel determined the AALs for known carcinogens and probable carcinogens in a different manner. Carcinogens, as opposed to non-carcinogens, are nonthreshold agents.⁵⁵ Because no data confirm a safe level of exposure for these chemicals, the Panel had to decide on an "acceptable" level of risk.⁵⁶ The Panel first categorized each agent as either a known human carcinogen (Group A) or probable human carcinogen (Group B), based on studies by the Carcinogen Assessment Group of the EPA.⁵⁷ For known carcinogens, the Panel determined AALs based on a lifetime risk of one incidence of cancer in one million exposed persons.⁵⁸ For probable human carcinogens, the Panel determined a less stringent AAL based on a lifetime risk of one incidence of cancer in 100,000 exposed persons.⁵⁹ These "acceptable" risk factors are based on current federal and state practice and are not considered extreme.⁶⁰ For known human carcinogens, the Panel used quantitative risk assessment⁶¹ techniques, based on available human data, to determine the incremental air concentration (concentration attributable to an emission source) associated with an additional lifetime cancer risk of one in a million persons exposed.⁶² For probable human carcinogens, the Panel determined AALs similarly, except it used extrapolations from animal studies to determine the incremental air concentration associated

week. *Id.* at 12-13. The Panel used a factor of two to adjust for the uncertainty inherent in studies of chronic effects. *Id.* at 14. It also decided to apply a safety factor of two to those agents eliciting irreversible or life-threatening effects at concentrations that might be reasonably expected to occur in the ambient air. *Id.* at 14-15.

55. "Nonthreshold" means that there is no "safe" level of exposure. *Industrial Union Dept., AFL-CIO v. American Petroleum Inst.*, 448 U.S. 607, 638 n.43 (1980); see Comment, *The Clean Air Act: Economic and Technological Feasibility in Setting Standards Under Section 112*, 20 LAND & WATER L. REV. 397, 403 n.73 (1987). A minority of scientists believe that carcinogens do have a threshold. Doniger, *Federal Regulation of Vinyl Chloride: A Short Course in the Law and Policy of Toxic Substances Control*, 7 ECOLOGY L.Q. 500, 511 n.42 (1978).

56. The "acceptable" level of risk is not the product of any actual data, but is primarily a policy decision. AIR TOXICS PANEL, *supra* note 42, at 26.

57. See *supra* note 41 and accompanying text.

58. AIR TOXICS PANEL, *supra* note 42, at 16.

59. *Id.* Industry officials critical of the AALs established by the Air Toxics Panel noted that lifetime risk estimates assume that the average life span is 70 years and that exposure to a particular cancer-causing agent is continuous during life. 2 HEARINGS, *supra* note 7, at V-522 (comments Richard V. Hargitt on behalf of E.I. Du Pont de Nemours).

60. EPA, the Occupational Safety and Health Administration, and the Food and Drug Administration use the same criteria for control of carcinogenic chemicals. 2 HEARINGS, *supra* note 7, at V-472 (comments at public hearing by Dr. Carl M. Shy, M.D., Chairman, Air Toxics Panel, North Carolina Academy of Sciences).

61. Quantitative risk assessment, as defined by the National Academy of Sciences, is the calculation of the probability of potentially adverse health effects from human exposure to environmental hazards. ENVIROLOGIC DATA, REVIEW OF THE NORTH CAROLINA PROPOSED AIR TOXICS PROGRAM § 1.3 (Jan. 1989), reprinted in 2 HEARINGS, *supra* note 7, at V-579. The risk assessment process usually involves four steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization. EPA Proposed Guidelines for Carcinogen Risk Assessment, 49 Fed. Reg. 46,294, 46,295 (1984). Hazard identification is a qualitative assessment that looks at the weight of the evidence to determine whether a chemical poses a hazard to human health. *Id.* Dose-response assessment characterizes the relationship between the dose of a chemical and the incidence of an adverse health effect in humans. This process usually involves extrapolations made from high to low doses and from animal to human exposures. *Id.* Exposure assessment estimates the intensity, frequency, and duration of human exposures to a chemical in the environment. *Id.* Risk characterization is the final step of combining exposure and dose-response assessments to reach a quantitative estimate of the risk. *Id.*

62. AIR TOXICS PANEL, *supra* note 42, at 16.

with an additional lifetime cancer risk of one in 100,000 persons exposed.⁶³

The AALs are not strict regulations but merely guidelines that aid the EMC in deciding whether human health is adequately protected.⁶⁴ The regulations provide that a facility may not emit any of the toxic air pollutants listed "in such quantities that may cause or contribute beyond the premises . . . to any *significant* ambient air concentration that may adversely affect human health."⁶⁵ The EMC is to rely on the AAL guidelines in determining what concentrations are "significant."⁶⁶

To date, North Carolina has established AAL guidelines for 105 TAPs—eighty-four AALs effective May 1, 1990, and twenty-one AALs effective May 1, 1991.⁶⁷ With few exceptions, all sources of air toxics must have a permit to emit any of these air toxics.⁶⁸ The regulations require existing sources to apply for a permit or permit modification 180 days after they receive notice from the DEM requesting that they apply for a permit to emit TAPs.⁶⁹ The DEM makes such notification or "permit calls" on the basis of Standard Industrial Classification (SIC) codes, which group industries of similar type.⁷⁰ This grouping benefits both the DEM and industries. It helps reduce the workload of the permitting staff by calling only a manageable number of industries at one time. Phasing in the program also spreads out the demand for air pollution control equipment, making it easier and cheaper for companies to obtain and install required technology.⁷¹

To acquire a permit, a new source must demonstrate through modeling⁷² that the AALs or guidelines will not be exceeded because of the facility's emissions.⁷³ Alternatively, a new source may avoid compliance with AALs by proving that a greater concentration than that set forth would not adversely affect human health.⁷⁴ Although this poses a difficult burden, a new source may support a request for a higher concentration in one of two ways. First, the new source may establish that the areas where the ambient concentrations are expected to exceed the AALs "are not inhabitable or occupied for the duration of the averaging time of the pollutant of concern."⁷⁵ For instance, a new source

63. *Id.* By definition, probable human carcinogens have no conclusive data based on human epidemiological studies, but the carcinogenicity may be based on "sufficient" animal studies. See EPA Guidelines for Carcinogen Risk Assessment, 51 Fed. Reg. 33,992, 34,000 (1986) (EPA's definition of Group B probable human carcinogens).

64. See 1 HEARINGS, *supra* note 7, at I-12 (agency response to comments).

65. N.C. ADMIN. CODE tit. 15A, r. 2D.1104 (Aug. 1990) (emphasis added).

66. *Id.*

67. NCEDF, *supra* note 1, at 22; see N.C. ADMIN. CODE tit. 15A, r. 2D.1104.

68. 1 HEARINGS, *supra* note 7, at I-3; see N.C. ADMIN. CODE tit. 15A, r. 2H.0610 (Aug. 1990). With the exception of new sources combusting only unadulterated fossil fuels or wood, all new sources of air toxics must have a permit. *Id.*

69. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(a)(3).

70. *Id.* r. 2H.0610(a)(4).

71. 1 HEARINGS, *supra* note 7, at I-46.

72. For a discussion of modeling, see *infra* text accompanying notes 86-98.

73. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(b)(1).

74. *Id.* r. 2H.0610(b)(2).

75. *Id.* r. 2H.0610(b)(2)(A).

may assert that the area in question is over a body of water and cannot be inhabited. Of course, for pollutants with a short averaging time,⁷⁶ an owner may still have difficulty showing that persons will not occupy the area during that time. For example, even if the area in question is an uninhabited body of water, fishing or other recreational activities may bring people into the area for short periods of time. Second, a new source may produce new toxicological data showing that the AAL for the pollutant in question is too low and the facility's ambient impact is below the level indicated by the new data.⁷⁷ Such data, however, is not easy to come by and may prove quite costly to obtain.⁷⁸ Furthermore, because of the uncertainties inherently involved in air toxics research, it is doubtful that a source will be able to prove with any degree of accuracy that a higher concentration will not adversely affect human health in any way.

Likewise, existing sources may obtain a permit by demonstrating either that they will comply with the guidelines, that noncompliance would not adversely affect human health, or that they qualify for one of three exceptions.⁷⁹ An existing source may receive a permit, even if it is not in compliance, if it submits a schedule satisfying the agency that it will comply with the air toxics guidelines within three years after receiving a permit call.⁸⁰ An existing source also may exceed the guidelines for a pollutant if the source can demonstrate that compliance is technologically infeasible⁸¹ or would result in significant economic hardship.⁸² These exceptions are premised on the belief that it is easier to adapt new sources to technological innovations in pollution control than to retrofit existing sources for the same pollution control. Generally, standards are more lax for existing sources than for new sources. Existing sources falling into a technological or economic infeasibility exception have three years after receiving written

76. Each category of toxic pollutants has a different averaging time. For carcinogens, facilities must average data over a period of a year. See *id.* r. 2H.1104(a). For chronic toxicants, acute systemic toxicants, and acute irritants, facilities must average data for 24 hours, one hour, and 15 minutes respectively. *Id.* To show that no adverse effects on human health will result from emissions exceeding the AALs, a source must demonstrate that the area in question is uninhabited and that most persons probably will not occupy the area for the duration of the averaging time. *Id.* r. 2H.0610(b)(2).

77. *Id.* r. 2H.0610(b)(2)(B).

78. The Air Toxics Panel collected the best scientific data and used state-of-the-art techniques for determining AALs. See 1 HEARINGS, *supra* note 7, at I-14, 19. Further, for carcinogens, the Panel relied on judgments of the EPA's Carcinogen Assessment Group because it felt that "these risk assessments, which are made at a national level and are subject to extensive peer review, provide the State with the best current scientific judgment about carcinogenic risks." 2 *id.* at V-472 (comments at public hearing by Dr. Carl M. Shy, M.D., Chairman, Air Toxics Panel, North Carolina Academy of Sciences). To obtain new toxicological data that would meet a facility's burden of proof, the owner would have to hire a team of experts, which would prove quite costly and offer no guarantee of success.

79. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(c).

80. *Id.* r. 2H.0610(c)(1).

81. *Id.* r. 2H.0610(c)(2). "Technologically infeasible" means that the technology necessary to reduce emissions to a level that does not exceed the AALs does not exist. 1 HEARINGS, *supra* note 7, at I-51 to I-52 (agency response to comments).

82. N.C. ADMIN. CODE tit. 15A, r. 2H.0160(c)(3). "'Significant economic hardship' means the cost of installing the technology necessary to prevent the acceptable ambient levels from being exceeded would result in a negative net profit when the installation of the technology is amortized over five years." 1 HEARINGS, *supra* note 7, at I-52 (agency response to comments).

notification from the DEM to achieve maximum feasible control⁸³—not necessarily AAL compliance. It may be quite some time, however, before the DEM decides to notify certain industries.⁸⁴

How do owners or operators determine the ambient concentrations around their facility? Owners are likely to use what is known as dispersion modeling.⁸⁵ Dispersion modeling estimates the concentration of a pollutant using mathematical simulations based on information about atmospheric and stack emission conditions.⁸⁶ The purpose of the models is to “‘predict pollutant concentrations at any point in the neighborhood of the source.’”⁸⁷ These models operate under the assumption that a pollutant disperses vertically out of the smokestack and then “disperses laterally as predicted by the laws of fluid dynamics.”⁸⁸ Dispersion modeling analysis typically consists of two stages: the screening stage and the refining stage.⁸⁹ The screening stage employs relatively simple techniques to estimate a maximum pollutant concentration by using an array of “worst case” meteorological data. Using “worst case” meteorology accounts for all possible atmospheric conditions. This screening model is intentionally designed to over-predict concentration levels.⁹⁰

If the screening model exhibits concentrations in excess of any applicable state or federal standard for any pollutant, the owner performs refined dispersion modeling. The refined modeling is more exacting, time consuming, and expensive and involves using actual meteorological data. The results of the refined modeling are more precise and invariably predict lower concentrations than the screening model.⁹¹

The modeling process incorporates various facility-specific and meteorological conditions by plugging these factors into a mathematical equation as variables or input. Some of these variables include wind speed, wind direction, atmospheric stability, temperature, mixing height, emission rate, stack gas temperature, stack gas exit velocity, stack diameter, stack height, and terrain features.⁹² The EPA explains how a slight variation in a few of these parameters can make a significant difference in ambient concentrations of a pollutant:

“[A] gaseous pollutant emitted over a grassy field will disperse much

83. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(d). “Maximum feasible control” means “the maximum degree of reduction for each pollutant . . . using the best technology that is available taking into account, on a case-by-case basis, energy, environmental, and economic impacts and other costs.” *Id.* r. 2H.0602(6).

84. *Cf.* 1 HEARINGS, *supra* note 7, at I-45 to I-46 (facilities will receive permit calls only as quickly and to the extent that the permitting staff can manage the workload).

85. *See id.* at I-35 (agency response to comments).

86. *Case, Problems in Judicial Review Arising from the Use of Computer Models and Other Quantitative Methodologies in Environmental Decisionmaking*, 10 B.C. ENVTL. AFF. L. REV. 251, 324 n.362 (1982).

87. *Id.* at 317 n.327 (quoting *Alabama Power Co. v. Costle*, 636 F.2d 323, 348 (D.C. Cir. 1980)).

88. *Id.* at 324 n.362.

89. 1 HEARINGS, *supra* note 7, at I-35 (agency response to comments).

90. *Id.* at I-35 to I-36.

91. *Id.* at I-36.

92. *Id.* at I-37.

differently than if the pollutant is emitted over a large urban area. There the dispersion will be affected not only by the local weather conditions but also by the greater turbulence caused by the different types of surface areas and heat sources throughout a city."⁹³

The problem with using so many highly sensitive variables is that modeling is arguably an inaccurate method of determining ambient concentrations. Some critics, for example, consider modeling "a very poor approximation of reality."⁹⁴ On the other hand, one must keep in mind that facility owners use modeling for permit purposes only; modeling determines only operational processes, air pollution control equipment parameters, and emissions rates to be included in the permit.⁹⁵ Once a permit is issued, its conditions are enforced through monitoring, not modeling.

The EMC chose modeling for this limited purpose because of its concern with cost-effective methods to achieve its clean air goals. The permit program itself can prove quite costly to industry, but the use of modeling to determine ambient concentrations of a pollutant is much cheaper than establishing, maintaining, and operating an adequate ambient monitoring network.⁹⁶ Although modeling is a plus for larger industries, which would be faced with installing extensive monitoring equipment and hiring qualified personnel to staff their many monitoring stations, smaller industries may find dispersion modeling to be the more cumbersome choice.⁹⁷ Dispersion modeling is highly technical and requires computer time and expertise that may not be readily available in smaller industries. The advantage of modeling, however, is that it allows all calculations to occur on the premises in front of a computer, and industries are more likely to get results quickly, even if it means sacrificing some accuracy.⁹⁸ Facilities need permits before operating, and the convenience of modeling ensures fast and efficient permit issuance.

Once permits are in place, facilities must use monitoring and reporting devices to determine compliance with the AALs.⁹⁹ The regulations require facilities with air quality permits to report emissions of listed TAPs.¹⁰⁰ Methods of quantifying emissions include stack testing, mass balance calculations, and emissions factors.¹⁰¹ The owner or operator of the facility is responsible for both the determination of emission rates and the accuracy of the data.¹⁰²

93. *Cleveland Elec. Illuminating Co. v. EPA*, 572 F.2d 1150, 1161 (6th Cir. 1978) (quoting EPA brief).

94. 3 HEARINGS, *supra* note 7, at V-2486 (comment at public hearings by William M. Deal, Vice President of Manufacturing, Bernhardt Furniture Co.).

95. 1 *id.* at I-36 (agency response to comments).

96. *Id.* at I-35.

97. 2 *id.* at V-716 (comments by Larry Runyan on behalf of the American Furniture Manufacturers Association) (refined modeling requires "additional time and expertise which most industries will not have available in-house").

98. *Id.* at I-36 to I-37.

99. N.C. ADMIN. CODE tit. 15A, r. 2D.1105.

100. *Id.* r. 2D.1105(c)(1).

101. 1 HEARINGS, *supra* note 7, at I-40.

102. See N.C. ADMIN. CODE tit 15A, r. 2D.1105; 1 HEARINGS, *supra* note 7, at I-40.

II. NORTH CAROLINA'S HEALTH-BASED APPROACH TO AIR TOXICS CONTROL

North Carolina adopted acceptable ambient levels for each toxic air pollutant without considering technological or economic infeasibility factors in its assessment. This health-based approach¹⁰³ presupposes that there is a certain level of pollution to which the public may be exposed without sacrificing public health and safety. This approach then establishes an acceptable or ideal limit on the level of pollutants in the ambient air. The ambient air level is derived by assessing various health-related factors and deciding on a level of risk that is acceptable for the maintenance of public health and welfare.¹⁰⁴ Subsequently, control measures are developed to achieve this ambient level. The health-based approach in its purest form disregards technological or economic feasibility factors in its analysis.¹⁰⁵ North Carolina, however, makes exception for existing sources when compliance with AAL guidelines would constitute economic hardship or when compliance is technologically infeasible.¹⁰⁶ North Carolina, by providing this exception, combines its health-based approach with another approach to pollution control, the technology-based approach.

The technology-based approach focuses on technology or pollution control measures that are available and feasible in light of the type of industry involved and often includes an inquiry into whether the industry is an existing or new source.¹⁰⁷ Then, the best available technology (BAT)¹⁰⁸ or a similar standard is applied regardless of whether any ambient goals are reached.¹⁰⁹ Title III of the 1990 CAA amendments exemplifies the technology-based approach to pollution control.

III. TITLE III OF THE 1990 CLEAN AIR ACT AMENDMENTS

In response to the failure of the health-based approach to air toxics regulation in section 112,¹¹⁰ Congress enacted new legislation designed to remedy some of the problems plaguing the previous CAA amendments.¹¹¹ First, Con-

103. For a discussion of a technology-based approach to regulation, see *infra* notes 107-09 and accompanying text.

104. The goal in setting ambient standards is to ensure that the air we breathe is sufficiently safe. See 42 U.S.C. § 7409 (1988).

105. The health-based approach provides no room for consideration of technological and economic feasibility. It often imposes standards that force industry to develop the control technologies needed to protect public health fully. Cf. Note, *supra* note 15, at 619 (provisions of CAA § 112 are technology-forcing); Marchant & Danzeisen, *supra* note 16, at 539 (same).

106. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(c) (Aug. 1990); see *supra* notes 81-82 and accompanying text.

107. Cf. Ackerman & Stewart, *Reforming Environmental Law*, 37 STAN. L. REV. 1333, 1334-37 (1985) (criticizing the best available technology approach for distinguishing between existing and new sources).

108. The BAT approach contemplates applying the best control technology currently available to all similar types of industries uniformly. *Id.* at 1334-37. For the purposes of this Note, the BAT approach shall be used when referring to a technology-based approach.

109. Graham, *supra* note 21, at 133.

110. See *supra* notes 3-6 and accompanying text.

111. See Clean Air Act of 1990, Pub. L. No. 101-549, § 301, 104 Stat. 2399, 2531-74 (1990) (codified at 42 U.S.C.A. § 7412 (West Supp. 1991)).

gress took the initiative and established its own list of hazardous air pollutants instead of delegating "listing" decisions to the EPA.¹¹² In the past, the Agency refused to identify substances producing adverse health effects because it knew that after listing a pollutant it would not be able to write standards to regulate it.¹¹³

Instead of creating control standards for individual pollutants, the Administrator now establishes categories of industrial sources that emit substantial amounts of each TAP.¹¹⁴ The EPA must publish the list of categories by November 15, 1991, and thereafter must provide an opportunity for public comment on the list.¹¹⁵ Once the EPA publishes the final list, the Act requires regulation of all major sources in each category.¹¹⁶ This source-by-source approach (as opposed to a pollutant-by-pollutant approach), combined with the mandatory list of pollutants, should speed the regulation process and help prevent foot-dragging by the EPA.

The Administrator must promulgate regulations establishing emission standards for all categories of major sources and area sources of TAPs.¹¹⁷ The emission standards

shall require the maximum degree of reduction in emissions of the hazardous air pollutants . . . that the Administrator, taking into consideration the cost of achieving such emission reduction, and any non-air quality health and environmental impacts and energy requirements, determines is achievable for new or existing sources in the category or subcategory to which such emission standard applies.¹¹⁸

Industries may implement this so-called maximum achievable control technology (MACT) using a wide variety of control measures including installation of control mechanisms, substitution of materials, change in work-practice methodology, and increased operational standards.¹¹⁹ The MACT control standard represents a marked departure from the strict risk-based or health-based approach of the old air toxic regulations¹²⁰ and allows the EPA to consider economic as well as other non-health-based criteria in determining MACT.

112. 42 U.S.C.A. § 7412(b)(1) (West Supp. 1991). Under the 1990 CAA amendments, the EPA Administrator has discretion to add to the list any pollutants that pose adverse health effects, but the Administrator may delete pollutants from the list only upon a showing that the substance "may not reasonably be anticipated to cause any adverse effects to the human health or adverse environmental effects." *Id.* § 7412(b)(3)(B). Although it may be in the best interest of some industries to try to petition for a change, scientific uncertainty will likely obviate any chance of success.

113. See *supra* text accompanying notes 24-29.

114. J. QUARLES & W. LEWIS, *supra* note 12, at 34.

115. 42 U.S.C.A. § 7412(c)(1).

116. *Id.* § 7412(d)(1). Further, the EPA must assure that "90 per cent of the area source emissions of the 30 hazardous air pollutants that present the greatest threat to public health in the largest number of urban areas" achieve compliance within ten years. *Id.* § 7412(c)(3).

117. *Id.* § 7412(d)(1).

118. *Id.* § 7412(d)(2).

119. *Id.* § 7412(d)(2)(A)-(E). This provision offers greater flexibility in control measures than was permitted under § 112 of the 1977 amendments. See 42 U.S.C. § 7412(d) (1988), amended by 42 U.S.C.A. § 7412(d) (West Supp. 1991).

120. See J. QUARLES & W. LEWIS, *supra* note 12, at 33; *supra* notes 24-29 and accompanying text. In the past, the EPA sought to graft such considerations onto the old regulations despite the fact that the literal language of § 112 did not allow inclusion of technological or economic feasibility

As is common to technology-based statutes,¹²¹ the 1990 amendments impose more stringent standards on new sources than on existing sources. For new sources, the MACT may be no less stringent than the emission control currently achieved in practice by what the Administrator determines is the best controlled similar source.¹²² Existing sources may have less stringent standards than new sources in the same category; the standard, however, may not be less stringent than the average emission limitation achieved by the best performing twelve percent of existing sources in the category of thirty or more sources, or the average emission limitation achieved by the best performing five sources in the category with fewer than thirty sources.¹²³ Although new source standards are tougher than those for existing sources, technological and economic feasibility continue to be appropriate factors to consider in determining new source MACT.¹²⁴

The new amendments also provide a schedule for the promulgation of MACT standards. By November 15, 1992, EPA must establish priorities for its list of source categories by considering the adverse effects of TAPs on the public health and the environment, the quantity and location of TAP emissions that each category will emit, and the efficiency of grouping categories according to the pollutants emitted.¹²⁵ Once the EPA sets priorities, it must promulgate MACT standards for the first forty categories no later than two years after enactment of the 1990 CAA.¹²⁶ Within four years, the EPA must establish standards for twenty-five percent of all of the listed categories. Within seven years standards must be promulgated for an additional twenty-five percent of the source categories, and at the end of ten years, all listed categories must have standards in place.¹²⁷

All sources must comply with MACT standards within three years after promulgation.¹²⁸ Congress, however, has built in an incentive for sources to reduce emissions on their own. The amendments provide that an existing source may obtain a six-year extension for compliance if it achieves a ninety percent reduction in emissions prior to the proposal of an applicable MACT standard.¹²⁹ Existing sources may want to analyze technological mechanisms, economic factors, and regulatory schemes to decide whether they can make a few minor changes now in order to qualify for the six-year extension later.¹³⁰ Because these decisions must be made before MACT standards are issued, however, industry must predict now whether achieving an immediate ninety-percent reduction will result in a savings over the cost of future compliance. Most likely,

factors in the equation. See *supra* note 29. Now Congress has chosen to incorporate those factors in the 1990 amendments. See 42 U.S.C.A. § 7412(d)(2) (West Supp. 1991).

121. Ackerman & Stewart, *supra* note 107, at 1335-36.

122. 42 U.S.C.A. § 7412(d)(3).

123. *Id.*

124. *Id.* § 7412(d)(2), (3); see J. QUARLES & W. LEWIS, *supra* note 12, at 34.

125. 42 U.S.C.A. § 7412(e).

126. *Id.*

127. *Id.*

128. *Id.* § 7412(i)(3)(A).

129. *Id.* § 7412(i)(5).

130. J. QUARLES & W. LEWIS, *supra* note 12, at 35.

industries will have to conduct a cost-benefit analysis individually to decide whether the cost of voluntary reductions now is less than any benefits associated with a six-year delay of the MACT standards.

Because of the EPA's prior record of laxity in promulgating air toxics control regulations,¹³¹ Congress was not content to leave the implementation of MACT standards to the Agency completely, so it designed the so-called "MACT hammer."¹³² If the EPA fails to promulgate MACT standards as required, sources will become subject to case-by-case MACT standards after states have their permit programs approved.¹³³ Permits will contain emission limitations for TAPs subject to regulation so that there will be equivalent standards under this alternative regulatory method.

Once the EPA establishes MACT standards, it must decide whether more stringent standards are required after application of MACT controls to protect the public health with an "ample margin of safety . . . [and] taking into consideration costs, energy, safety, and other relevant factors, [to prevent] an adverse environmental effect."¹³⁴ This plan, designed to address the problem of residual risks, is one health-based aspect that Congress has not abandoned. The EPA must promulgate residual risk standards with an ample margin of safety nine years after establishing MACT standards in the case of the first source categories regulated and eight years for all other categories.¹³⁵ Legislation mandates that the EPA establish residual risk standards for carcinogens that present a cancer risk greater than one in one million after MACT controls have been installed.¹³⁶ Because of the uncertainty of risk assessment, the 1990 amendments call for the National Academy of Sciences to review the EPA's risk assessment methodology for determining carcinogenic risk associated with air toxics exposure and to recommend improvements in methodology.¹³⁷

The above regulations primarily deal with normal, everyday releases from facilities. In addition, the new amendments propose methods of preventing and responding to accidental releases of toxic substances.¹³⁸ Here, the amendments require a completely new regulatory program. The EPA must publish, not later than November 15, 1992, an "initial list of 100 substances which, in the case of

131. See *supra* 26-29 and accompanying text.

132. J. QUARLES & W. LEWIS, *supra* note 12, at 34.

133. *Id.*; see 42 U.S.C.A. § 7412(j). The state-issued permits "shall contain emission limitations for the hazardous air pollutants . . . emitted by the source that the Administrator (or the State) determines, on a case-by-case basis, to be equivalent to the limitation that would apply to such source if an emission standard had been promulgated in a timely manner." *Id.* § 7412(j)(5). This provision adds a layer of protection in case the EPA fails to promulgate an emissions standard. It also requires the states to apply the MACT standards required under federal law.

134. 42 U.S.C.A. § 7412(f)(2). The EPA must promulgate these more rigorous standards in the event Congress fails to act on its risk assessment report. *Id.* § 7412(f)(1)-(2).

135. *Id.* § 7412(f)(2)(C).

136. *Id.* The federal act requires that categories of sources emitting a pollutant classified as a known, probable, or possible human carcinogen reduce emissions to a level associated with a lifetime cancer risk of less than one in one million. The North Carolina regulations, however, distinguish between known and probable carcinogens by allowing an acceptable risk of one in 100,000 for probable human carcinogens. See *supra* note 59 and accompanying text.

137. 42 U.S.C.A. § 7412(f)(2)(A).

138. J. QUARLES & W. LEWIS, *supra* note 12, at 36; see 42 U.S.C.A. § 7412(r).

accidental release, are known to cause or may reasonably be anticipated to cause death, injury, or serious adverse effects to human health or the environment."¹³⁹ The amendments identify sixteen pollutants that must be included.¹⁴⁰ By 1993 the EPA must promulgate regulations for the prevention and detection of accidental releases and measures for emergency response, including the preparation of risk management plans.¹⁴¹ Owners or operators of a facility subject to accidental-release regulations subsequently will be required to develop risk management plans that comply with the EPA regulations and include a hazard assessment, a prevention program, and a response plan.¹⁴²

In sum, the new federal regulation of air toxics has shifted from a health- or risk-based approach to primarily a technology-based approach, applying MACT standards to similarly situated sources and allowing for more flexible methods of control technology.

IV. RETENTION OF NORTH CAROLINA'S AIR TOXICS REGULATIONS

The North Carolina and federal regulations take two different approaches to the problem of cleaning the ambient air. North Carolina relied on a predominantly health-based approach in formulating AALs,¹⁴³ whereas the 1990 CAA amendments use a technology-based approach in prescribing emissions standards.¹⁴⁴ Although both the North Carolina and the federal programs combine elements of each approach, both are essentially predicated on one theory or the other. Neither approach to environmental regulation has obtained overwhelming support.¹⁴⁵ Environmentalists advocate preservation of public health and a clean environment at any cost and therefore favor a strict health-based approach. Industrialists, in contrast, view cost minimization as a primary goal and favor a technology-based approach. As a result, two opposing groups actively dispute the value of these two methods of regulation.¹⁴⁶ The choices are not as clear cut as they may seem, however, for each method contains inherent weaknesses.

Neither the health-based nor the technology-based approach can be effective on its own because each addresses different issues. The technology-based approach responds to immediate problems of scarcity and technological infeasibility but fails to project a long-term plan for pollution reduction.¹⁴⁷ The health-based approach, in contrast, takes an idealistic approach to long-term

139. 42 U.S.C.A. § 7412(r)(3).

140. *Id.*

141. *Id.* § 7412(r)(7)(B).

142. *Id.*; see J. QUARLES & W. LEWIS, *supra* note 12, at 36.

143. See 1 HEARINGS, *supra* note 7, at I-11 (agency response to comments).

144. See J. QUARLES & W. LEWIS, *supra* note 12, at 33.

145. See 1 HEARINGS, *supra* note 7, at I-66 to I-69 (agency summary of testimony in favor of technology-based regulatory scheme); cf. Doniger, *supra* note 55, at 555-56 (comparing health-based and technology-based statutes regulating toxics at state and federal levels).

146. See 1 HEARINGS, *supra* note 7, at I-4 to I-11 (agency summary of the competing arguments of industry and environmentalists); Latin, *Ideal Versus Real Regulatory Efficiency: Implementation of Uniform Standards and "Fine-Tuning" Regulatory Reforms*, 37 STAN. L. REV. 1267, 1293 (1985).

147. Ackerman & Stewart, *supra* note 107, at 1335-37.

goals, but ignores the short-term problem of limited technical innovation and resources.¹⁴⁸ In addition to these theoretical difficulties, practical problems plague both approaches. The efficacy of the health-based approach is limited by scientific uncertainties associated with risk assessment,¹⁴⁹ while the technology-based approach avoids the risk assessment difficulties, but faces problems of substantial cost and limited flexibility needed to address the idiosyncracies of individual facilities.¹⁵⁰

Despite the substantial difficulties encountered in each regulatory approach, North Carolina should retain its health-based regulations in conjunction with the federal scheme for several reasons. First, although risk assessment techniques are fraught with uncertainties, the Air Toxics Panel used the best possible research techniques and based its assumptions on the most current data available.¹⁵¹ The major problem facing the Panel and agencies applying the health-based approach to TAP regulation is the treatment of carcinogens and the corresponding risks. Risk assessment, the most common technique used in epidemiological studies, requires scientific determinations under conditions of substantial uncertainty concerning the risks involved, limited availability of scientific data, and economic barriers to conducting adequate research.¹⁵²

Ultimately, the health-based or risk-based approach is subject to the confines of present scientific knowledge. Many statutes using the health-based approach result in two-part risk assessment analysis. The first part involves a political decision as to what level of cancer risk is acceptable. The second part is a scientific determination of the level of exposure to a pollutant which causes that particular risk.¹⁵³ Unfortunately, the causal relationship between a pollutant and cancer is often difficult to assess.¹⁵⁴ Part of the problem is that cancer is a latent disease that fails to manifest itself until fifteen to forty years after exposure begins.¹⁵⁵ Also, the operation of chemical carcinogens and their effect on human metabolic processes confound scientists.¹⁵⁶ There is substantial disagreement in the scientific community as to how much exposure to a potential carcinogen actually begins cell mutation. Scientists question whether one brief exposure is enough or whether several extended exposures within a short period

148. 1 HEARINGS, *supra* note 7, at I-68 to I-69 (agency summary of comments).

149. See Doniger, *supra* note 55, at 508-14.

150. 1 HEARINGS, *supra* note 7, at I-68 (agency response to comments); Ackerman & Stewart, *supra* note 107, at 1335-36.

151. See *supra* note 78.

152. Doniger, *supra* note 55, at 508-14. These factors prevent agencies from making precise and informed decisions about risks at varying degrees of pollutant exposure. The limitations of science and the controversial nature of risk management techniques result in substantial litigation challenging these acceptable risk levels as unfounded and capricious, leaving the agencies constantly fighting for ground. See, e.g., *Industrial Union Dept., AFL-CIO v. American Petroleum Inst.*, 448 U.S. 607, 630-38 (1980).

153. See Latin, *Good Science, Bad Regulation, and Toxic Risk Assessment*, 5 YALE J. ON REG. 89, 89 (1988).

154. Doniger, *supra* note 55, at 510. To account for gaps in scientific knowledge, regulators incorporate conservative estimates into risk assessment models. See Latin, *supra* note 153, at 94 (noting the inconsistent treatment of uncertainty in risk assessment).

155. Doniger, *supra* note 55, at 511-12.

156. *Id.* at 510.

of time will cause cancer to appear.¹⁵⁷ The most perplexing yet important question is what risks can be expected from a varying range of doses.¹⁵⁸ If this correlation was known, risk levels for carcinogens probably could withstand judicial scrutiny. Because of the lack of human studies, scientists have had to resort to animal studies to predict human exposure risks.¹⁵⁹ Two problems arise when using animal studies. First, there is uncertainty in extrapolating dose-response levels across species lines.¹⁶⁰ Second, because scientists must use relatively small test groups,¹⁶¹ chemicals must be administered at high dosages to establish a strong correlation between the dose and an adverse effect.¹⁶² Therefore, scientists must extrapolate to low doses to mimic the actual exposure level in the environment.¹⁶³

In short, agencies cannot predict how a difference in risk correlates to small changes in dose strength. The North Carolina Air Toxics Panel had to recommend guidelines with these scientific uncertainties in mind. Thus, the Panel combined policy decisions with scientific determinations to propose air toxics regulations. The Panel emphasized that, given scientific uncertainties, its approach "should not be considered a precise method distinguishing safe from unsafe levels of contaminants, but rather a means to establish flexible guidelines which can be used to raise flags of concern and set priorities for action."¹⁶⁴ AALs may be riddled with educated guesses that tend to err on the side of conservatism in order to protect the public health. Nonetheless, the Panel based the AALs on value judgments reflecting the best available scientific information at the time¹⁶⁵ and cannot be accused of being arbitrary or capricious.

The second reason North Carolina should apply its regulations in addition to those established under federal law is that the State has refined its regulatory system by dividing the hazardous air pollutants into noncarcinogens and carcinogens and using assessment methods that cater to known properties of each.¹⁶⁶ This category-specific approach affords the greatest degree of protection because guidelines may be more or less stringent based on varying degrees of toxicity. For the categories of noncarcinogens, the Air Toxics Panel used TLVs for chem-

157. *Id.* at 510-11. Scientists also differ over whether the human body has defense mechanisms against a single exposure and the possible effects of combined carcinogenic substances. *Id.*

158. *Id.* at 511.

159. *Id.* at 512.

160. *Id.* at 513.

161. In order to conduct an animal study with a 95% accuracy rate, the test must involve at least six million animals. These experiments are not practical because of tremendous expense and vulnerability to statistical errors. *Id.*

162. *Id.* at 512-13. Scientists cannot trace sufficiently low doses of chemicals in small test groups. Thus, to induce cancer at detectable rates, scientists must administer chemicals to test animals at much higher doses than humans normally would experience. *Id.* at 512.

163. *Id.* at 513.

164. AIR TOXICS PANEL, *supra* note 42, at 5.

165. 1 HEARINGS, *supra* note 7, at I-14 (agency response to comments questioning the scientific data used to determine AALs).

166. See *supra* notes 51-63. Varying treatment for carcinogens and noncarcinogens not only adds an element of flexibility to North Carolina's approach, but also adds credibility to health-based standards.

icals and multiplied them by various safety factors to come up with AALs.¹⁶⁷ The TLVs are limits assigned to industrial chemicals—chemicals found in the work place by agencies such as the American Conference of Government and Industrial Hygienists (ACGIH) and the Occupational Safety and Health Administration (OSHA).¹⁶⁸ The ACGIH has assessed threshold levels of chemicals in the work place for over forty years, and these threshold limits for noncarcinogenic pollutants are fairly well established.¹⁶⁹ The ACGIH conducted extensive research on limited groups of workers exposed to a limited number of pollutants at specific times and for an exact duration. The TLV determinations represent the best judgment of safe occupational levels given the present state of knowledge.¹⁷⁰ The Air Toxics Panel took these values and compensated for the differing environment, exposure times, and varying susceptibilities to reflect outdoor conditions.¹⁷¹ These adjustment factors, however, are not well established and involve discretionary judgments in converting TLVs to legitimate AALs. Industry argued that this modified TLV factored approach was not appropriate for outdoor exposure, and the addition of strict numerical factors could not simulate the variance between the work place and continuous outdoor exposure.¹⁷² Although this may be true, the final ambient levels were a product of established scientific data and the best educated assumptions based on this reliable data. Furthermore, if clean air goals are to be attained, regulators must bite the bullet and undauntedly push for health-based regulatory schemes in the face of scientific uncertainty.

This same argument applies to the regulation of carcinogens, which inherently involves more uncertainty, requiring even more guesswork. Applying the most advanced scientific methods, the Air Toxics Panel accepted the risk of one cancer per million persons exposed, and it performed quantitative risk analysis to determine the incremental concentration of a pollutant associated with that risk.¹⁷³ Because it is not feasible to conduct experiments on animals or make human observations to define this low risk level, the Panel had to choose a statistical model to estimate the dose associated with this acceptable low risk. The modeling process is scientific in that it incorporates principles of extrapolation and estimation. There is no proof that one particular model is clearly right or valid, however, and these models usually lead to widely divergent results. The Panel decided to use the fairly conservative linearized multistage model to extrapolate from high dose animal studies.¹⁷⁴ The Panel chose this method because it

167. See *supra* note 54.

168. See 1 HEARINGS, *supra* note 7, at I-14 (agency response to comments).

169. *Id.*

170. See 4 HEARINGS, *supra* note 7, at VI-35; ACGIH, *supra* note 36, at 3.

171. See 1 HEARINGS, *supra* note 7, at I-15; AIR TOXICS PANEL, *supra* note 42, at 9-16.

172. 1 HEARINGS, *supra* note 7, at I-13 to I-14 (agency summary of comments by industry representatives).

173. In comments responding to a proposal by the EPA for regulation of benzene in 1988, the National Resources Defense Council accepted this same risk. The Council relied on various sources that cite a risk of one in one million as the upper bound for de minimis risk. Marchant & Danzeisen, *supra* note 16, at 543-44.

174. Letter from Dr. Carl M. Shy, M.D., Chairman, Air Toxics Panel, North Carolina Academy of Sciences to Gladys Van Pelt, Ph.D., Chair, Air Quality Committee, Environmental Management

incorporated an adequate set of conservative assumptions regarding human cancer risks at low doses.¹⁷⁵ To ensure that the most accurate, up-to-date scientific information will be used for future regulation, the North Carolina Department of Environment, Health and Natural Resources is establishing a Secretary's Scientific Advisory Board to advise the EMC and DEM on current developments in toxicology and health risks of air toxics.¹⁷⁶ This step adds a layer of protection so that guesswork in establishing acceptable ambient levels is kept to a minimum.

North Carolina should retain the air toxics regulations because of the flexibility written into its provisions. While maintaining its health-based goals, the state program makes allowances to exceed AALs and in limited circumstances permits variances for existing sources. A variance may be granted if a facility demonstrates that compliance is impracticable due to technological infeasibility.¹⁷⁷ This provision allowing for variances recognizes the difficulty in retrofitting old sources with control devices. If an existing source is successful in demonstrating that meeting the ambient standards is technologically infeasible, then it need only install "maximum feasible control technology."¹⁷⁸ Allowing for variances represents a melding of the technology-based and health-based approaches and is one of the most apparent concessions North Carolina made to industry.

The state regulations also offer flexibility by regulating on a facility-by-facility basis rather than in the uniform manner applied under the 1990 CAA amendments. This distinction can be illustrated by comparing aspects of the North Carolina air toxics program with parallel provisions of the federal scheme.

First, the North Carolina regulations mandate that each owner or operator of a facility apply for a permit in order to emit a toxic pollutant.¹⁷⁹ The North Carolina regulations also require each owner to conduct computer dispersion

Commission (July 29, 1988) (discussing Air Toxic Panel's recommendations), *reprinted in* 4 HEARINGS, *supra* note 7, at VI-175.

175. *Id.* This model assumes that risk is linearly proportional to concentration.

176. 1 HEARINGS, *supra* note 7, at I-91; *see* Division of Environmental Management, Department of Environment, Health, and Natural Resources, Draft: Secretary's Advisory Board on Toxic Air Pollutants (undated) (describing Advisory Board's composition and function), *reprinted in* 4 HEARINGS, *supra* note 7, at VI-193.

177. A variance gives a facility an opportunity to avoid the stringent standards and possibly opt for lesser standards by meeting certain conditions. North Carolina allows variances for existing sources if they can demonstrate technological infeasibility, *see supra* note 81, or economic hardship, *see supra* note 82, in complying with the ambient guidelines. N.C. ADMIN. CODE tit. 15A, r. 2D.0610(c)(2)-(3) (Aug. 1990). An existing source may delay compliance by submitting an acceptable schedule that provides for compliance within three years. *Id.* r. 2D.0610(c)(1). The Code does not specify how quickly reductions must proceed during the three-year interval nor whether reductions must be constant and gradual. In any case, a facility may be able to emit pollutants that exceed ambient levels for a period of three years if the DEM accepts its schedule. *Id.* The variance for a new source is much more difficult to obtain because it requires that the source prove either that its emissions do not exceed acceptable ambient levels beyond its property boundary, or that such emissions will not adversely affect human health. *Id.*

178. *See supra* note 83.

179. N.C. ADMIN. CODE tit. 15A, r. 2H.0610(a).

modeling to determine its ambient contribution for a pollutant.¹⁸⁰ Based on these models, the State will implement control technology on an individualized basis to assure that the facility will not contribute significant amounts of a pollutant into the ambient air.

In contrast, the 1990 CAA amendments require that once the EPA lists a category of sources, all sources within the category must achieve emission reductions by applying the MACT standard.¹⁸¹ This approach is a highly centralized, industry-uniform system often used to implement BAT: the standards apply to all industries of similar types regardless of their geographic location, climate, or population size of those affected. Uniform standards neglect certain areas and allow for creation of high concentrations of pollutants in small areas, or so-called "hot spots."¹⁸²

State systems like North Carolina's can alleviate particularly problematic areas by requiring each source to have a pollution reduction permit and demanding that each facility reduce its own emissions so that collective emissions do not exceed AALs. Whereas uniform controls may prove dysfunctional by regulating some areas more or less than necessary, highly decentralized controls can be tailored to redress specific pollution problems. For example, each of several sources located in close proximity may comply with uniform federal standards on an individual basis, yet the combination of their emissions may cause a significant ambient concentration within the vicinity. A decentralized system like North Carolina's specifically addresses each facility's emissions based on ambient concentrations in the neighboring area and thereby can assure that the ambient concentration in a particular industrial area is not exceeded. Furthermore, North Carolina's regulations require facilities to apply additional control technology if emissions of two or more sources located in a small area exceed AALs.¹⁸³

A highly centralized, uniform approach does have some advantages over a more individualized approach. They include:

decreased information collection and evaluation costs, greater consistency and predictability of results, greater accessibility of decisions to public scrutiny and participation, increased likelihood that regulations will withstand judicial review, reduced opportunities for manipulative behavior by agencies in response to political or bureaucratic pressures, reduced opportunities for obstructive behavior by regulated parties, and decreased likelihood of social dislocation and "forum shopping" resulting from competitive disadvantages between geographical regions

180. *Id.* r. 2H.0610(b)(1).

181. See 42 U.S.C.A. § 7412(c) (West Supp. 1991); J. QUARLES & W. LEWIS, *supra* note 12, at 34.

182. Ackerman & Stewart, *supra* note 107, at 1350.

183. N.C. ADMIN. CODE tit. 15A, r. 2D.1107(a) (Aug. 1990). "The owner of a facility shall not be required to conduct [a] multi-facility ambient impact analysis . . . This type of analysis shall be done by the Division of Environmental Management." *Id.* r. 2D.1107(c). The Division may require a facility to install additional control technology if ambient impact analysis reveals that it is necessary to protect the public from the combined effect of multifacility pollutants. *Id.* r. 2D.1107(a).

or between firms in regulated industries.¹⁸⁴

At the federal level, this uniform approach has fostered significant improvements in environmental quality with a cost to society that has not proved excessive.¹⁸⁵ Nevertheless, opponents still argue that uniform standards are not cost-effective.¹⁸⁶ Uniform requirements may waste billions of dollars annually by ignoring variations among plants and industries in the cost of reducing pollution and by ignoring geographic variations in pollution effects.¹⁸⁷ Some argue that North Carolina's health-based approach allows for the most cost-effective methods by regulating facility by facility.¹⁸⁸ Only facilities that need emissions reduction will put on control technology, and sources that need emissions reductions will use only those methods that assure that ambient guidelines will not be exceeded.

This argument also supports North Carolina's use of a health-based regulatory system rather than a technology-based system, such as BAT. The technology-based approach used by the federal government would prove much more costly for North Carolina than implementing a health-based approach, because under the control technology approach, every industry emitting toxic air pollutants would have to install the best control equipment whether or not such controls were necessary to meet the AALs.¹⁸⁹ Because the implicit goal of BAT is the air quality level attainable if every facility installed BAT,¹⁹⁰ the technology-based approach does not assure or even attempt to assure that public health is protected from any adverse health effects.¹⁹¹ What happens if the best available technology is in place and an unacceptable risk to human health is still present? Additional controls are not technologically or economically feasible in the BAT scenario.¹⁹² The health-based approach allows for many options to achieve ambient goals, including reduced emissions or substitution of materials, because it is not constrained by best available technology.

Furthermore, BAT approaches, although assuring that sources install established control technologies, do not provide a strong incentive for the development of new technology.¹⁹³ Once BAT is in place, industry has complied with the standard and no more need be done. Such an approach may even discourage the development of new technology, and the long-term effect may be devastating. Similarly, others contend that the "BAT strategy is inconsistent with intelligent priority setting."¹⁹⁴ Merely applying maximum pollution controls to the pollutant that makes the regulatory agenda may prevent an agency from ad-

184. Latin, *supra* note 146, at 1271.

185. *Id.* at 1273. This is not to say that even excessive costs could not be absorbed by society if necessary.

186. *Id.* at 1273 n.25.

187. See Ackerman & Stewart, *supra* note 107, at 1335.

188. See 1 HEARINGS, *supra* note 7, at I-68 (agency response to comments).

189. *Id.* at I-68 to I-69.

190. Ackerman & Stewart, *supra* note 107, at 1341.

191. See 1 HEARINGS, *supra* note 7, at I-68 (agency response to comments).

192. *Id.*

193. Ackerman & Stewart, *supra* note 107, at 1336.

194. *Id.* at 1337.

addressing the most serious pollutants first. Once a new pollutant is identified, BAT requires costly inquiries into the state of control technology in the industries emitting such pollutants. Then BAT requires implementation of control to the full extent of available technology for each particular pollutant.¹⁹⁵ Not only is this an inefficient method of prioritizing, it is inefficient in terms of cost.

North Carolina should retain the regulations not only because they better meet North Carolina's needs, but also because they may meet its needs sooner. The 1990 CAA amendments contain a complex regulatory timetable. The amendments do not require facilities to put on control technology until after the EPA establishes categories of sources and decides how to rank the categories.¹⁹⁶ The EPA then must promulgate emissions standards for each source category. The schedule requires the EPA to decide on emissions standards for only forty source categories by November 1992.¹⁹⁷ Once the EPA establishes emissions standards, new CAA provisions require facilities to apply MACT within three years.¹⁹⁸ This means that it may be November 1995 before the first industries (listed in the forty source categories) will have to comply with the federal regulations, and it will be November 2000 before fifty percent of the source categories will be regulated.¹⁹⁹

The North Carolina scheme has the potential to ensure regulation sooner than Title III. Regulations for eighty-four pollutants became effective in May 1990.²⁰⁰ Already, new sources must apply for permits before beginning construction so that AALs will not be exceeded.²⁰¹ Existing sources must apply for permits within 180 days after receiving a permit call.²⁰² Once an existing source applies for a permit, it must apply MACT no later than three years after receiving notification. Ideally, if the DEM calls existing sources now, they must comply with the air toxics guidelines by May 1993 because, unlike the federal statutory provisions, the DEM need only call a facility's SIC, and compliance must follow within three years.²⁰³ The DEM need not decide on source categories and emissions limitations before requiring any control technology.

North Carolina should not only retain its regulations but should strive for implementation as soon as possible. Accordingly, the DEM should commence permit calling as soon as feasible. The CAA should not provide an excuse to hold back implementation of the state regulations and wait for the EPA to set national emissions standards. Because North Carolina's regulations place no

195. *Id.* at 1359.

196. See 42 U.S.C.A. § 7412(e) (West Supp. 1991); J. QUARLES & W. LEWIS, *supra* note 12, at 37-38.

197. 42 U.S.C.A. § 7412(e). These first categories may or may not include pollutants that are prevalent in North Carolina.

198. *Id.* § 7412(i).

199. *Id.* § 7412(e) (standards for half of the categories must be promulgated by November 1997); *id.* § 7412(i) (industries have three years from promulgation to comply). This timetable assumes no bureaucratic delay or foot dragging by the EPA and that the EPA has sufficient resources to accomplish the CAA regulatory program.

200. N.C. ADMIN. CODE tit. 15A, r. 2D.1104 (Aug. 1990).

201. *Id.* r. 2H.0610(a)(1).

202. *Id.* r. 2H.0610(a)(3).

203. The DEM, however, may call a SIC code at any time. See *supra* note 84.

time constraints on the DEM regarding permit calls, the DEM may wait an indefinite period of time before notifying existing sources. Although such delay may be permitted by the Code, the DEM would sacrifice public health by its inaction. North Carolina toxic air pollutants continue to be a major problem that the federal program may not address for five years. Furthermore, even when the CAA requires compliance with the first standards within five years after enactment of the statute, the government-regulated pollutants may not pose a threat to North Carolina. In other words, state facilities may emit certain pollutants that the EPA decides warrant a lower priority. In that case, facilities need not apply control technology until 1997 or later. Finally, the 1990 CAA amendments do not address North Carolina's "hot spots" sufficiently, whereas the state regulations contemplate multiple-facility emissions.

V. CONCLUSION

North Carolina has abandoned the industry-uniform, technology-based approach in favor of a more flexible health-based approach, even though this approach had proved unworkable at the federal level. The state has not fully implemented the plan yet and will not know its efficacy for some time. The North Carolina regulations emphasize goals slightly different from those of the CAA. First, risk management and public safety are of the utmost importance in the state system and, as such, the regulators are willing to face the uncertainties of risk assessment and opt for intelligent guesses based on the best scientific information available, rather than sacrificing air quality and suppressing uncertainties in a BAT system. The second state goal is improving long-term air quality by encouraging scientific innovation and variation in control technology at the industry level. In implementing an individualized facility-by-facility approach, only those facilities in need of controls install them; this method of regulation is not only cost-effective but also resource-conscious. Finally, the state program advocates the use of a scientific panel to review listed TAPs and research possible air toxics in trying to assure that the EMC will revise present standards upon receipt of new scientific data and that facilities will install new control technology as needed.

North Carolina should retain its regulations in conjunction with the federal statutes because the state regulations offer a bold, individualized approach to health-based environmental regulation with a view to cost-effectiveness and opportunity for limited variances for economic or technological infeasibility. North Carolina's decentralized regulations can address specific problem areas that the federal system of uniform standards may overlook. While the North Carolina regulations set ambient air concentration guidelines, the CAA amendments require emission reduction based on MACT standards. Theoretically, neither regulatory system excludes the other and industries can comply with both.²⁰⁴ Keeping the state's health-based interests in mind, the DEM should

204. 1 HEARINGS, *supra* note 7, at I-73 to I-74 (agency response to comments). Furthermore, the state probably will adopt the EPA's standards. There should be no conflict in having two sets of standards or regulations—one for emission rates and one for ambient concentrations. *See id.*

implement the state guidelines for all sources as quickly as possible. The option to delay permit calls and subsequent regulatory control may seem attractive to the DEM now;²⁰⁵ a "wait-to-see" attitude, however, subverts the original intent behind these health-based regulations. The EMC considered the possibility of concurrent federal regulations and decided that "[t]o wait for Congress to act would produce needless and unnecessary delay."²⁰⁶ If the state sits on its hands and fails to implement its own program, it gives facility owners a license to pollute. Where North Carolina has the opportunity and the mechanism to control its own environmental destiny it should do so, and not rely on a federal program that contains no assurance of success.

DEANNA SCHMITT

205. Although industry predicts the permitting process will produce a bottleneck because of limited resources of the permitting staff, the EMC remains convinced that the permit calling process will spread the workload sufficiently to allow for adequate review of permits. *Id.* at I-45 to I-46.

206. *Id.* at I-73.

Indelible Ink in the Milk*: Adoption of the Inclusionary Approach to Uncharged Misconduct Evidence in *State v. Coffey*

*It was tragic. The defendant had, on several occasions, taken up with divorced women so he could have access to their children. He had sexually abused four other children before this little girl. The government either couldn't or didn't introduce any of this prior conduct at the defendant's trial. The jury acquitted. When some of the jurors subsequently found out about the prior incidents, they were furious. "If only we'd known about them, we'd have convicted the guy!"*¹

The issue of admission of uncharged misconduct evidence is the most litigated evidentiary issue in most federal and state appellate courts.² The multitude of litigation in this area no doubt results from the highly prejudicial effect this type of evidence has on judges, juries, and lay-persons³ in both civil and criminal cases.⁴ The North Carolina Supreme Court, like most state and federal courts, traditionally recognized the impact of this type of evidence on a defendant's right to a fair trial by adopting a general exclusionary rule, with a limited number of well-recognized exceptions.⁵ The court further protected defendants from unrestrained admission of prior uncharged acts by requiring that evidence of these acts be substantially similar to and not too remote in time from the charged crime.⁶ Since Congress adopted Federal Rule of Evidence 404(b) and numerous state legislatures adopted corresponding state versions of that rule,⁷ many courts in these jurisdictions have relaxed significantly—if not eliminated—the evidentiary constraints on the admissibility of this highly prejudicial class of evidence.⁸

* The title originates from a colorful remark in a decision by the United States Court of Appeals for the Third Circuit involving the admission of uncharged misconduct evidence: "A drop of ink cannot be removed from a glass of milk." *Virgin Islands v. Toto*, 529 F.2d 278, 283 (3rd Cir. 1976).

1. The opening quotation is borrowed from the beginning of a law review article on admission of evidence of a defendant's prior sexual misconduct with children. Hutton, *Prior Bad Acts Evidence in Cases of Sexual Contact with a Child*, 34 S.D.L. REV. 604, 604 (1989).

2. See 22 C. WRIGHT & K. GRAHAM, *FEDERAL PRACTICE AND PROCEDURE: EVIDENCE* § 5239, at 427 (1978).

3. E. IMWINKELRIED, *UNCHARGED MISCONDUCT EVIDENCE* § 1:03 (1984 & Supp. 1991).

4. See *Huddleston v. United States*, 485 U.S. 681, 685 (1988) (noting that rule 404(b) applies to both civil and criminal cases).

5. See *State v. McClain*, 240 N.C. 171, 173-76, 81 S.E.2d 364, 366-68 (1954).

6. See, e.g., *State v. Boyd*, 321 N.C. 574, 577, 364 S.E.2d 118, 119 (1988); *State v. Scott*, 318 N.C. 237, 248, 347 S.E.2d 414, 420 (1986); *State v. Moore*, 309 N.C. 102, 106-07, 305 S.E.2d 542, 544-45 (1983); *State v. Shane*, 304 N.C. 643, 654-56, 285 S.E.2d 813, 820-21 (1982), cert. denied, 465 U.S. 1104 (1984). For a discussion of these cases, see *infra* notes 57-77 and accompanying text.

7. See, e.g., N.C.R. EVID. 404(b) (1986). For North Carolina's version of rule 404(b), see *infra* note 43.

8. See Imwinkelried, *The Need to Amend Federal Rule of Evidence 404(b): The Threat to the Future of the Federal Rules of Evidence*, 30 VILL. L. REV. 1465, 1467-68 (1985). For a discussion of Congress' adoption of Federal Rule of Evidence 404(b) and the rule's subsequent effects on the federal courts' approach to uncharged misconduct evidence, see *infra* notes 83-90 and accompanying text. For a discussion of state courts' reactions to their legislatures' adoption of a version of rule 404(b), see *infra* notes 104-10 and accompanying text.

In *State v. Coffey*⁹ the North Carolina Supreme Court followed this trend and abrogated the traditional procedural safeguards against admission of this highly prejudicial class of evidence. The court ruled that evidence of the defendant's prior sexual misconduct was admissible to show motive and intent in a trial for the murder of a ten-year-old girl.¹⁰ In upholding the trial court's admission of the evidence, the *Coffey* court announced that a "clear general rule of inclusion" governs the admissibility of uncharged misconduct evidence under rule 404(b).¹¹ The court stated that the adoption of rule 404(b) superseded any language to the contrary in its previous opinions.¹² Moreover, the conspicuous absence of an analysis under the previously applied two-pronged test of similarity and proximity in time¹³ intimated that the court interpreted its adoption of the inclusionary approach as obviating the need to prove either of these previously required independent safeguards.¹⁴

This Note begins with a brief discussion of the evidence presented at the trial in *Coffey* and the court's rationale for upholding its introduction on appeal. The Note then outlines two lines of state precedent relating to the court's opinion in *Coffey*: cases delineating the general standard for admissibility of uncharged misconduct evidence and cases applying this general standard in situations involving evidence of uncharged sexual misconduct, similar to that in *Coffey*. The Note next analyzes two significant aspects of the court's decision. First, the Note discusses the propriety of the court's express adoption of a rule of inclusion for uncharged misconduct evidence and contends that the *Coffey* court's adoption of this approach is justified by the concomitant intent of the North Carolina General Assembly. Second, the Note analyzes the court's application of this approach to the present case and its apparent abandonment of the previously applied two-pronged test of similarity and proximity in time. The Note asserts that the *Coffey* court erroneously discarded these important safeguards. Specifically, it allowed the admission of uncharged misconduct evidence for the sole purpose of proving that the defendant was guilty of the charged offense because he had acted in accordance with such behavior in the past, precisely the purpose rule 404(b) prohibits. The Note concludes that only the North Carolina General Assembly has the authority to restore the common-law presumption against admission of uncharged misconduct evidence. The Note, however, recommends that the court restore the previously required two-pronged test for admission of similar uncharged misconduct evidence.

On July 19, 1979, the beaten and asphyxiated body of ten-year-old Amanda Ray was found near a lake in Mecklenberg County, North Carolina.¹⁵ Eight years later the State indicted the defendant for the first degree murder of the

9. 326 N.C. 268, 389 S.E.2d 48 (1990).

10. *Id.* at 280-81, 389 S.E.2d at 55-56.

11. *Id.* at 278, 389 S.E.2d at 54.

12. *Id.*

13. See *infra* notes 57-77 and accompanying text for a discussion of the application of this two-pronged standard in previous North Carolina cases.

14. See *Coffey*, 326 N.C. at 278-81, 389 S.E.2d at 54-56.

15. *Id.* at 274, 389 S.E.2d at 51.

child.¹⁶ The prosecution based the first degree murder charge on the theory that the defendant kidnapped the victim with the intent to commit indecent liberties and subsequently murdered her to avoid exposure.¹⁷

The prosecution's proof at trial relied exclusively on various types of circumstantial evidence. The State produced numerous eyewitnesses who testified that they had seen Amanda talking with a man they identified as the defendant in the two days prior to the discovery of her body.¹⁸ In addition to eyewitness identification, the State offered evidence that fibers found on the victim's body matched fibers found on several items proven to be in the defendant's possession around the time of the murder.¹⁹

Finally, the State offered the testimony of both Janet Ashe and her pastor, Reverend James Hall.²⁰ Mrs. Ashe testified about an incident involving the defendant and her three-year-old daughter, Angel. Mrs. Ashe stated that in May 1979 she left Angel with the defendant and, after returning home, learned from the child that the defendant had masturbated in front of her.²¹ Both Janet Ashe and her pastor testified that, when confronted, the defendant admitted to masturbating in front of the child.²²

Based on this circumstantial evidence, the jury convicted the defendant of first degree murder and, following the jury's recommendation, the trial court entered a death sentence.²³ On direct appeal to the North Carolina Supreme Court, the defendant raised numerous assignments of error from both the trial and sentencing proceedings.²⁴ The court rejected all of the defendant's objections arising from the trial and conviction phase, but remanded for a new sentencing proceeding because of prejudicial errors that occurred during that phase.²⁵

16. *Id.* at 274, 389 S.E.2d at 52.

17. *Id.* at 280, 389 S.E.2d at 55.

18. *Id.* at 274-75, 389 S.E.2d at 52. More specifically, several eyewitnesses testified that they saw the victim fishing on a lake with the defendant on July 18, 1979, the day before Amanda's body was discovered near the same lake. *Id.* at 275, 389 S.E.2d at 52. Several of the witnesses also saw the defendant in a white van at both the lake and Amanda's apartment complex. *Id.* at 274-75, 389 S.E.2d at 52.

19. The State introduced fibers taken from a sofa and van which belonged to the defendant at the time of the murder. *Id.* at 276, 389 S.E.2d at 53. Both the defendant and his ex-wife confirmed that the defendant owned a white and blue van in 1979. *Id.* at 276-77, 389 S.E.2d at 53. Apparently, the sofa had not been cleaned in approximately ten years; thus the police were able to obtain hair fibers from a dog, since deceased but admittedly owned by the defendant. *Id.* An expert in trace evidence testified that dog hairs found in the sofa matched those found on the victim's body and in the defendant's van. *Id.* The expert also stated that fibers from the carpet in the defendant's van matched fibers found on Amanda's body. *Id.*

20. *Id.* at 277-78, 389 S.E.2d at 54.

21. *Id.*

22. *Id.* In a *voir dire* hearing prior to the admission of this evidence, the court concluded that the incident involving Angel Ashe was admissible but excluded evidence of two other incidents where the defendant had taken indecent liberties with other young girls. *Id.* at 279, 389 S.E.2d at 55. The court ruled that the cumulative effect of this evidence would be more prejudicial than probative under Rule 403 of the North Carolina Rules of Evidence. *Id.*

23. *Id.* at 274, 389 S.E.2d at 51.

24. *Id.*

25. *Id.* The court reversed the death sentence based on a failure to follow procedures outlined in § 15A-2000(c)(3) of the North Carolina General Statutes. *Id.* at 296, 389 S.E.2d at 64-65. This

The court's rationale for upholding the admission of the testimony involving the incident with Angel Ashe is especially significant. The court rejected the defendant's contention that evidence of other crimes, wrongs, or acts falls under a general rule of exclusion, subject only to certain exceptions.²⁶ Instead, the court embraced the position that the adoption of North Carolina Rule of Evidence 404(b) manifested a clear legislative intent on the part of the North Carolina General Assembly to codify a general rule of inclusion.²⁷ Under this standard of inclusion, the court stated, such uncharged misconduct evidence generally is admitted, "subject to but *one exception* requiring its exclusion if its *only* probative value is to show that the defendant has the propensity or disposition to commit an offense of the nature of the crime charged."²⁸ The court added that its prior decisions have been "markedly liberal" in admitting evidence of prior sex offenses of an "unnatural" character, such as in the present case.²⁹

Under this general standard of inclusion, the court ruled that the trial court properly admitted evidence of the Angel Ashe episode because its admission satisfied two distinct purposes specifically enumerated in rule 404(b).³⁰ First, the court ruled that the incident with Angel Ashe could provide a motive for Amanda's murder.³¹ In addition, the court held that the lower court properly admitted this evidence to prove specific intent, a necessary element of the underlying felony of kidnapping.³² Finally, the court ruled that in admitting the evidence the trial court did not abuse its discretion under North Carolina Rule of Evidence 403.³³

statute requires that the foreman return a signed writing on behalf of the jury finding that mitigating circumstances were insufficient to outweigh the aggravating factors before imposing the death penalty. N.C. GEN. STAT. § 15A-2000(c)(3) (1988). Because the trial court did not follow this procedure, the supreme court remanded the case for a new sentencing proceeding. *Coffey*, 326 N.C. at 297, 389 S.E.2d at 65.

26. *Coffey*, 326 N.C. at 278, 389 S.E.2d at 54.

27. *Id.*

28. *Id.* at 278-79, 389 S.E.2d at 54 (emphasis in original).

29. *Id.* at 279, 389 S.E.2d at 54-55 (quoting 1 H. BRANDIS, BRANDIS ON NORTH CAROLINA EVIDENCE § 92 (3d ed. 1988)).

30. The rule lists as proper purposes proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake, entrapment, or accident. N.C.R. EVID. 404(b) (1986). This list is identical to that contained in Federal Rule of Evidence 404(b) except for the addition of the word "entrapment." *Id.* 404(b) commentary.

31. *Coffey*, 326 N.C. at 280, 389 S.E.2d at 55. This reasoning reflects the prosecution's theory that the defendant killed the victim to avoid any exposure similar to that which occurred with Angel Ashe. *Id.* Although the prosecution was not required to prove a motive, the court held that motive is always admissible when the act is in question. *Id.* Consequently, the court ruled that the trial court properly admitted the evidence for this purpose. *Id.*

32. *Id.* at 280-81, 389 S.E.2d at 55-56. The prosecution theorized that the defendant took the victim with the intent to commit indecent liberties, thereby committing a kidnapping. *Id.* at 281, 389 S.E.2d at 55-56. The court agreed that the episode involving Angel Ashe tended to prove the necessary specific intent and ruled that the evidence was admitted properly on this theory as well. *Id.*

33. *Id.* at 281, 389 S.E.2d at 56. The court stated that under rule 403 the trial judge has wide discretion to determine if the probative value of the evidence is outweighed substantially by the danger of undue prejudice to the defendant. See N.C.R. EVID. 403. The court held that the trial court had remained within the bounds of its discretion on this issue. *Coffey*, 326 N.C. at 281, 389 S.E.2d at 56.

In addressing the admissibility of uncharged misconduct evidence, the North Carolina Supreme Court confronted a number of its previous decisions, decided both before and after the enactment of the North Carolina Rules of Evidence, which took effect on July 1, 1984.³⁴ The cases decided after the adoption of rule 404(b) retain some of the pre-rule safeguards, but also show a marked trend toward the admission of "other crimes evidence."³⁵ This trend toward admissibility is particularly prominent in the area of prior sexual acts, especially those of an "unusual" or "unnatural" character.³⁶

Before the adoption of rule 404(b) and to some extent even after its adoption, the seminal case concerning other crimes evidence in North Carolina was *State v. McClain*.³⁷ In *McClain*, a jury convicted the defendant of prostitution. The defendant appealed and excepted to the admission of evidence that she surreptitiously stole \$135 from her client subsequent to the alleged prostitution.³⁸ The North Carolina Supreme Court held that the trial court erred in admitting the evidence. The court announced that a general rule of exclusion "subject to certain well recognized exceptions" governed admission of evidence of uncharged offenses.³⁹ Applying this standard, the *McClain* court excluded evidence relating to the defendant's uncharged larceny because it did not fall within one of the listed exceptions to the general rule of exclusion.⁴⁰ Thus, under *McClain*, to admit evidence of other uncharged crimes the state had to prove that the evidence satisfied one of the finite number of well-recognized exceptions.

The *McClain* court expounded several strong justifications for the general rule of exclusion:

"the dangerous tendency and misleading probative force of this class of evidence require that its admission should be subjected by the courts to rigid scrutiny. Whether the requisite degree of relevancy exists is a judicial question to be resolved in the light of the consideration that the inevitable tendency of such evidence is to raise a legally spurious presumption of guilt in the minds of the jurors. Hence, if the court does not clearly perceive the connection between the extraneous criminal transaction and the crime charged, that is, its logical relevancy, the accused should be given the benefit of the doubt, and the evidence should be rejected."⁴¹

34. An Act to Simplify and Codify the Rules of Evidence, ch. 701, § 1, 1983 N.C. Sess. Laws 666, 668-69 (codified at N.C.R. EVID. 404(b)).

35. Other crimes evidence is a term of art used to refer to proffered evidence relating to an act that is independent of the charged crime. This is the type of evidence at issue in rule 404(b) cases such as *Coffey*.

36. See *infra* note 77 and accompanying text.

37. 240 N.C. 171, 81 S.E.2d 364 (1954).

38. *Id.* at 172-73, 81 S.E.2d at 365.

39. *Id.* at 174, 81 S.E.2d at 366. The Court then delineated eight exceptions to the exclusionary rule. Compare *id.* with N.C.R. EVID. 404(b) (1986) (the exceptions provided in *McClain* are extremely similar to those specifically enumerated in rule 404(b)).

40. *McClain*, 240 N.C. at 177, 81 S.E.2d at 368. The court subsequently held that admission of such evidence would be prejudicial to the defendant's right to a fair trial. *Id.*

41. *Id.* at 177, 81 S.E.2d at 368 (quoting *State v. Gregory*, 191 S.C. 212, 221, 4 S.E.2d 1, 4 (1939)); see also *id.* at 173-74, 81 S.E.2d at 356-66 (the court also discusses the reasons for an exclusionary rule); *State v. Shane*, 304 N.C. 643, 654, 285 S.E.2d 813, 820 (1982) ("[F]undamental fairness

On July 1, 1984, the North Carolina Rules of Evidence took effect,⁴² with rule 404(b) specifically addressing the admissibility of uncharged misconduct evidence.⁴³ Originally, the North Carolina Supreme Court interpreted the new rule to be consistent with prior North Carolina practice.⁴⁴ Despite this avowed consistency, North Carolina courts' interpretation of the general standard under rule 404(b) has been subtly, yet significantly, modified from the pre-rule standard to one allowing greater admissibility.

For example, in *State v. Morgan*,⁴⁵ an early case interpreting the new rule, the court stated that uncharged misconduct evidence is admissible so long as it is offered for some purpose other than to show action in conformity.⁴⁶ In *Morgan* the defendant allegedly came out of his place of business and shot the victim.⁴⁷ At trial, the State proffered evidence that the defendant had pointed a gun at an unrelated party three months earlier to disprove his claim of self-defense.⁴⁸ The court held that offering evidence of previous violent behavior to prove that the defendant was the aggressor in the affray was exactly the type of propensity evidence excluded by rule 404(b).⁴⁹ Thus, under *Morgan*, the court's announced standard appears to be more lenient than under *McClain* because it focuses not on the exclusion of evidence but instead on the admission of evidence assuming the prosecution can prove the existence of any proper purpose.⁵⁰ The *Morgan* court's application of this standard to the facts of the case, however, retains the same protections against propensity evidence present in pre-rule cases.⁵¹

In a case decided in the same year as *Morgan*, the court took an even broader reading of the admissibility of uncharged misconduct evidence under rule 404(b). In *State v. Weaver*,⁵² the court asserted that under both rule 404(b) and previously under *McClain*, "the purposes for which evidence of other crimes, wrongs or acts is admissible is not limited to those enumerated either in

requires giving defendant the benefit of the doubt and excluding the evidence. [Or, as it is more descriptively said in the game of baseball, the tie must go to the runner]." (brackets in original).

42. An Act to Simplify and Codify the Rules of Evidence, ch. 701, § 1, 1983 N.C. Sess. Laws 666, 668-69 (codified at N.C.R. EVID. 404(b)).

43. Rule 404(b) reads:

Other crimes, wrongs, or acts.—Evidence of other crimes, wrongs or acts is not admissible to prove the character of a person in order to show that he acted in conformity therewith. It may, however, be admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake, entrapment or accident.

N.C.R. EVID. 404(b).

44. *State v. Young*, 317 N.C. 396, 412, 346 S.E.2d 626, 635 (1986). In fact, the commentary to rule 404(b) indicates this same point. See N.C.R. EVID. 404(b) commentary.

45. 315 N.C. 626, 340 S.E.2d 84 (1986).

46. *Id.* at 636, 340 S.E.2d at 91.

47. *Id.* at 629, 340 S.E.2d at 86. Apparently this incident evolved from a dispute between the defendant and the victim over closing down his business. *Id.*

48. *Id.* at 637-38, 340 S.E.2d at 91-92.

49. *Id.* at 638, 340 S.E.2d at 92.

50. *Id.* at 635-40, 340 S.E.2d at 90-93.

51. See *id.* at 637-38, 340 S.E.2d at 91-92.

52. 318 N.C. 400, 348 S.E.2d 791 (1986). The State charged the defendant with felonious breaking and entering and larceny of a chain saw and socket set. *Id.* at 400, 348 S.E.2d at 792.

the rule or in *McClain*.”⁵³ The court stated, in even broader language, “[i]n fact, as a careful reading of rule 404(b) clearly shows, evidence of other offenses is admissible so long as it is relevant to any fact or issue other than the character of the accused.”⁵⁴ Under this expansive standard, the court found evidence of the defendant’s prior dealings with a government witness admissible to show a common plan or scheme.⁵⁵ Thus, in both its language and its application, the *Weaver* court significantly relaxed the general admissibility standard for uncharged misconduct evidence.

In addition to cases delineating the general admissibility standard, numerous North Carolina cases have addressed the admissibility of “other crimes” evidence in the specific context of sexual misconduct. These cases, also decided before and after the adoption of rule 404(b), clearly show that the court’s approach to admitting evidence of prior sex acts has been markedly different from its approach when evaluating cases involving all other crimes. The court stated in a pre-rule case: “Our Court has been very liberal in admitting evidence of similar sex crimes in construing the exceptions to the general rule [of exclusion].”⁵⁶ Even in cases of prior sex acts, however, the court has required that the evidence meet certain safeguards in order to protect the defendant’s right to a fundamentally fair trial.

In a pre-rule case in this specific context, *State v. Shane*,⁵⁷ the State charged a police officer with first degree sexual offense for allegedly using his position to gain sexual favors from employees of a massage parlor.⁵⁸ The trial court admitted testimony from the defendant’s former employer about a similar but independent incident.⁵⁹ Despite the supreme court’s finding that a “striking similarity” existed between the two episodes, it held that the period of time elapsing between the two events, approximately one year, “substantially negated the plausibility of the existence of an ongoing and continuous plan to engage persistently in such deviant activities.”⁶⁰ Consequently, the court held that the erroneous and prejudicial admission of such evidence required a new trial.⁶¹

53. *Id.* at 402-03, 348 S.E.2d at 793. This quotation is at variance with the clear holding of *McClain*. See *supra* note 39 and accompanying text.

54. *Weaver*, 318 N.C. at 403, 348 S.E.2d at 794 (citing 1 H. BRANDIS, BRANDIS ON NORTH CAROLINA EVIDENCE § 91 (1982)).

55. *Id.* at 404, 348 S.E.2d at 793-94. A defense witness testified that he, and not the defendant, had stolen and sold the tools. *Id.* at 402, 348 S.E.2d at 793. The court allowed admission of the defendant’s prior dealings with the State’s witness to disprove the claim of the defense witness. *Id.* at 403-04, 348 S.E.2d at 793-94.

56. *State v. Greene*, 294 N.C. 418, 423, 241 S.E.2d 662, 665 (1978). In *Greene*, the court ruled that two cases of assault and rape had been properly consolidated and that testimony of the complaining witness in one case was admitted properly as to the other case under both the identity and common plan exceptions. *Id.*

57. 304 N.C. 643, 285 S.E.2d 813, *cert. denied*, 465 U.S. 1104 (1982).

58. *Id.* at 644-46, 285 S.E.2d at 815.

59. *Id.* at 652, 285 S.E.2d at 819. This other incident, like the one charged, involved the defendant’s use of his position of authority to coerce the victim to engage in oral sex. *Id.*

60. *Id.* at 655-56, 285 S.E.2d at 820-21. The State had argued that this evidence should be admitted under the common plan or scheme purpose. *Id.* at 653-54, 285 S.E.2d at 820.

61. *Id.* at 656-57, 285 S.E.2d at 821-22.

Thus, in *Shane*, the court recognized the first essential requirement in admitting sexual misconduct evidence—proximity in time to the charged crime.

Another pre-rule case involving admission of uncharged sexual crimes illustrates a second requirement for admitting such evidence. In *State v. Moore*,⁶² the trial court, during the defendant's trial for first degree sexual offense, admitted the testimony of a victim of a separate uncharged rape, who identified the defendant as her assailant.⁶³ The trial court admitted the evidence of the previous sexual offense under the theory that it helped identify the perpetrator of the charged crime—a recognized exception to the general rule of exclusion.⁶⁴ The *Moore* court, however, ruled that the testimony relating to the uncharged crime should have been excluded because of the limited similarities between the two crimes.⁶⁵ The court noted that "[t]o allow the admission of evidence of other crimes without such a showing of similarities would defeat the purpose of the general rule of exclusion."⁶⁶

Thus, prior to the adoption of rule 404(b), the North Carolina Supreme Court recognized, in separate cases, the existence of two important constraints on the admissibility of sexual misconduct: proximity in time and similarity. The court initially reaffirmed the continued existence of these requirements after the adoption of rule 404(b) in *State v. Scott*.⁶⁷ In *Scott*, a jury convicted the defendant of a first degree sexual offense perpetrated on his three and four-year-old nieces.⁶⁸ The trial court admitted testimony that eight years previously, when the defendant was only thirteen, he had forced his sister to have sexual intercourse with him at knife-point.⁶⁹

In analyzing the propriety of admitting such evidence, the court recognized that "no rule exists generally permitting evidence of a defendant's 'unnatural disposition.'"⁷⁰ Nevertheless, the court observed that it had made "exceptions under *McClain* or Rule 404(b) if the incidents . . . [were] sufficiently similar and not too remote in time so as to be more probative than prejudicial under the Rule 403 balancing test."⁷¹ Applying this standard, the court ruled that the

62. 309 N.C. 102, 305 S.E.2d 542 (1983).

63. *Id.* at 103, 305 S.E.2d at 542-43.

64. *Id.* at 106, 305 S.E.2d at 544.

65. *Id.* at 108, 305 S.E.2d at 545-46. The court found that certain similarities did exist; specifically, the assailant in both instances had used a knife, both involved oral sex, and both occurred in Greensboro, North Carolina, within a two month period. *Id.* at 107, 305 S.E.2d at 545. The court, however, found that differences between the two—primarily the contrast between the relatively non-threatening demeanor of the assailant in this case and the violent disposition of the assailant in the first assault—outweighed any similarities that might have existed. *Id.* Justice Meyer wrote a scathing dissent asserting that the appellate court should have focused upon the similarities and not the differences in evaluating the propriety of admitting the testimonial evidence. *Id.* at 109-10, 305 S.E.2d at 546-47 (Meyer, J., dissenting).

66. *Id.* at 106-107, 305 S.E.2d at 545. The court found the admission of evidence prejudicial and granted the defendant a new trial. *Id.* at 109, 305 S.E.2d at 546.

67. 318 N.C. 237, 347 S.E.2d 414 (1986).

68. *Id.* at 239, 347 S.E.2d at 415.

69. *Id.* at 244-45, 347 S.E.2d at 418-19. The trial court had allowed testimony regarding the incident to be elicited during the cross-examination of the defendant and his sister. *Id.*

70. *Id.* at 248, 347 S.E.2d at 420.

71. *Id.*

incident with defendant's sister eight years ago was too remote in time and dissimilar from the charged crime to be admitted under any theory recognized by the court.⁷²

In other cases decided after the adoption of rule 404(b), however, the court has repeatedly held admissible under this two-pronged standard of proximity in time and similarity evidence of "unnatural" prior sex acts. The court frequently admits evidence of uncharged sexual wrongs where a defendant has sexually abused members of his family on various independent occasions.⁷³ For example, in *State v. DeLeonardo*,⁷⁴ the defendant faced charges of sexually molesting his sons.⁷⁵ The trial court, however, admitted evidence that the defendant also had sexually abused his three-year-old daughter.⁷⁶ Without reservation, the supreme court held that the evidence relating to defendant's daughter was admissible to establish a common plan or scheme on the part of the defendant to sexually abuse his children.⁷⁷

Thus the court has been more liberal in allowing evidence of uncharged sexual misconduct than it has been concerning evidence of other uncharged crimes. The court, however, has placed some restraint on the admissibility of this type of evidence by employing consistently the two-pronged standard requiring sufficient similarity and proximity in time to the charged act. The decisions of the North Carolina Supreme Court reflect a general adherence to these procedural safeguards in evaluating evidence of uncharged sexual misconduct.

The court's decision in *Coffey* is significant for two distinct but related reasons. First, the *Coffey* court expressly adopted the inclusionary approach to

72. *Id.* The court granted the defendant a new trial as a result of the erroneous and prejudicial admission of this evidence. *Id.*

73. See, e.g., *State v. Miller*, 321 N.C. 445, 454, 364 S.E.2d 387, 392 (1988); *State v. Boyd*, 321 N.C. 574, 577-78, 364 S.E.2d 118, 120 (1988); *State v. DeLeonardo*, 315 N.C. 762, 771-72, 340 S.E.2d 350, 357 (1986).

74. 315 N.C. 762, 340 S.E.2d 350 (1986).

75. *Id.* at 763, 340 S.E.2d at 352.

76. *Id.* at 767, 340 S.E.2d at 354-55.

77. *Id.* at 770-71, 340 S.E.2d at 356-57. In a similar case, the court affirmed the admission of evidence, pursuant to the same common-plan exception, that the defendant had sexually abused the victim, his son, the day after the charged incident. *State v. Miller*, 321 N.C. 445, 454, 364 S.E.2d 387, 392 (1988).

On the same day as the *Miller* decision, the court handed down another case approving the admission of evidence of uncharged sex crimes with another member of the defendant's family. *State v. Boyd*, 321 N.C. 574, 364 S.E.2d 118 (1988). In *Boyd*, the defendant was convicted of the rape of his thirteen-year-old step-daughter. *Id.* at 575-76, 364 S.E.2d at 118-19. The trial court admitted testimony that within a year of the charged rape, the defendant's wife had discovered the defendant in the step-daughter's bed with an eight-year-old female cousin. *Id.* at 576, 364 S.E.2d at 119. Applying the two-pronged standard announced in *Scott*, the *Boyd* court found the incident "sufficiently similar to the act charged and not too remote in time" so as to be properly admitted. *Id.* at 578, 364 S.E.2d at 120.

In a case decided a few months after *Boyd*, the court again applied this standard to uphold the admission of similar evidence in *State v. Rosier*, 322 N.C. 826, 370 S.E.2d 359 (1988). In *Rosier*, the defendant faced charges of first degree sexual offense for forcing an unrelated seven-year-old to have anal intercourse. *Id.* at 827, 370 S.E.2d at 360. The court held that testimony regarding the defendant's prior conviction for fondling other young children only three months prior to the charged offense was similar and not too remote in time. *Id.* at 828-29, 370 S.E.2d at 360-61. Accordingly, the court ruled that the evidence was properly admitted because it established a common scheme or plan. *Id.*

uncharged misconduct evidence under rule 404(b). Although the court attempted to de-emphasize the importance of this decision by stating that it was merely announcing a principle recognized by its former cases, in fact the court's adoption of a general inclusionary approach is an important transformation of court policy.⁷⁸ Second, the *Coffey* decision is significant because the court inexplicably abandoned the two-pronged test it previously employed to determine the admissibility of uncharged sexual misconduct evidence.

The exclusionary approach to uncharged misconduct generally prohibits the admission of evidence of other crimes subject only to a finite number of well-recognized exceptions.⁷⁹ Under this traditional doctrine, evidence that does not fit within one of the previously acknowledged "pigeonhole[s]" is automatically excluded.⁸⁰ In contrast, the inclusionary approach allows the admission of other crimes evidence unless its only probative value is to show the accused's propensity to commit the charged crime. In other words, admissibility of such evidence is not limited to one of the specifically enumerated exceptions, but can be introduced if it satisfies any valid purpose other than to illustrate the defendant's propensity to commit crimes.⁸¹

To evaluate the *Coffey* court's express adoption of the inclusionary approach and equally explicit rejection of the exclusionary approach, two independent factors must be examined: the original intent of the North Carolina

78. For a discussion of the propriety of the *Coffey* court's depiction of its precedent as unquestionably establishing a rule of inclusion under rule 404(b), see *infra* note 123 and accompanying text.

79. Reed, *Trial by Propensity: Admission of Other Criminal Acts Evidenced in Federal Criminal Trials*, 50 U. CIN. L. REV. 713, 713 (1981) [hereinafter Reed, *Trial by Propensity*].

80. Imwinkelried, *supra* note 8, at 1468. The exclusionary approach is founded on nothing less than the nature of the Anglo-American accusatorial system, as opposed to the European inquisitorial system. Reed, *Trial by Propensity*, *supra* note 79, at 713. In an accusatorial system, the state has the burden of proving that a criminal defendant committed some act prohibited by law. In juxtaposition, the inquisitorial system assumes the accused committed a crime and requires him to prove his innocence. *Id.* More specifically, the emergence of the propensity rule can be linked to a seventeenth-century reaction against the use of the Star Chamber during the Tudor and Stuart Regimes. The Star Chamber was a royal court designed to eliminate political and religious rivals of the monarchy through treason trials. *Id.* at 716-17.

The exclusionary approach was the traditional view in American jurisdictions, with a majority of the states subscribing to it in the early twentieth century. Stone, *The Rule of Exclusion of Similar Fact Evidence: America*, 51 HARV. L. REV. 988, 1036 n.221 (1938). The North Carolina Supreme Court traditionally followed this approach, as indicated by the court's decision in *State v. McClain*, 240 N.C. 171, 173, 81 S.E.2d 364, 365 (1954). For a discussion of the *McClain* decision, see *supra* notes 37-41 and accompanying text. The exclusionary approach remained in force in a majority of states at the time of the adoption of the Federal Rules of Evidence. Reed, *The Development of the Propensity Rule in Federal Criminal Causes 1840-1975*, 51 U. CIN. L. REV. 299, 303-04 (1982) [hereinafter Reed, *Federal Causes*].

81. Imwinkelried, *supra* note 8, at 1468. This approach originated in Dean Wigmore's character rule. J. WIGMORE, EVIDENCE § 216 (1904); see Reed, *Trial by Propensity*, *supra* note 79, at 736-37. Dean Wigmore asserted that evidence of an accused's character was highly probative and that courts were too cautious in excluding this evidence to protect a criminal defendant. As a result of this view, Wigmore proposed instead that character evidence should be excluded only when it was offered solely to establish an accused's propensity to criminal activity—essentially the modern inclusionary rule. J. WIGMORE, § 216; see also Reed, *Trial by Propensity*, *supra* note 79, at 736-37 (explaining Dean Wigmore's views on the subject).

This inclusionary approach prevailed in a minority of federal jurisdictions before the adoption of the Federal Rules of Evidence. See Reed, *Federal Causes*, *supra* note 80, at 303-04; Reed, *Admission of Other Criminal Act Evidence After Adoption of The Federal Rules of Evidence*, 53 U. CIN. L. REV. 113, 113-14 (1984) [hereinafter Reed, *After Adoption*].

General Assembly in adopting rule 404(b) and, conversely, the justifications typically given by many state courts for retaining the exclusionary rule. Because North Carolina rule 404(b) is modeled after its counterpart in the federal rules,⁸² determining congressional intent is a useful starting point for any attempt to determine the intent of the North Carolina General Assembly in enacting its corresponding state version. The congressional history of the House⁸³ and Senate,⁸⁴ as well as the explanations set forth in the advisory committee's note⁸⁵ to Federal Rule of Evidence 404(b), indicate that Congress intended to place more emphasis on the admissibility of uncharged misconduct evidence. As a result, many commentators concluded that Congress intended to adopt an inclusionary approach in enacting federal rule 404(b).⁸⁶

82. See N.C.R. EVID. 404(b) commentary (1986).

83. See H.R. REP. No. 650, 93d Cong., 1st Sess. 7, reprinted in 1974 U.S. CODE CONG. & ADMIN. NEWS 7051, 7081;

The second sentence of Rule 404(b) as submitted to the Congress began with the words "This subdivision does not exclude the evidence when offered." The Committee amended this language to read "It may, however, be admissible", the words used in the 1971 Advisory Committee draft, on the ground that *this formulation properly placed greater emphasis on admissibility* than did the final Court version.

Id. (emphasis added).

84. See S. REP. No. 1277, 93d Cong., 2d Sess. 24-25, reprinted in 1974 U.S. CODE CONG. & ADMIN. NEWS 7051, 7071;

[Rule 404(b)] provides that evidence of other crimes, wrongs, or acts is not admissible to prove character but may be admissible for other specified purposes such as proof of motive.

Although your committee sees no necessity in amending the rule itself, it anticipates that the use of the discretionary word "may" with respect to the admissibility of evidence of crimes, wrongs, or acts is not intended to confer any arbitrary discretion on the trial judge. Rather, it is anticipated that with respect to permissible uses for such evidence, the trial judge *may exclude it only on the basis of those considerations set forth in Rule 403*, i.e. prejudice, confusion or waste of time.

Id. (emphasis added).

85. See FED. R. EVID. 404(b) advisory committee note:

Subdivision (b) deals with a specialized but important application of the general rule excluding circumstantial use of character evidence. Consistently [sic] with that rule, evidence of other crimes, wrongs, or acts is not admissible to prove character as a basis for suggesting the inference that conduct on a particular occasion was in conformity with it. However, the evidence may be offered for another purpose, such as proof of motive, opportunity, and so on, which does not fall within the prohibition. In this situation the rule does not require that the evidence be excluded. No mechanical solution is offered. The determination must be made whether the danger of undue prejudice outweighs the probative value of the evidence in view of the availability of other means of proof and other factors appropriate for making decisions of this kind under Rule 403.

Id.

86. See Imwinkelried, *supra* note 8, at 1484; Reed, *After Adoption*, *supra* note 81, at 156.

Professor Imwinkelried conducted an extensive survey of the available legislative history surrounding the adoption of Federal Rule of Evidence 404(b). Specifically, he questioned whether rule 404(b) retained the common-law allocation of burden to the proponent of evidence or, alternatively, whether the rule switched the burden to the defendant by incorporating the rule 403 balancing test. Imwinkelried, *supra* note 8, at 1479-80. The rule 403 balancing test places the burden of excluding evidence on the defendant by requiring that relevant evidence should be admitted unless the danger of unfair prejudice substantially outweighs its probative value. FED. R. EVID. 403. Professor Imwinkelried concluded that Congress intended to incorporate the rule 403 balancing test, which in effect adopts an inclusionary approach by allocating the burden of excluding evidence to the defendant. See Imwinkelried, *supra* note 8, at 1479-80. His conclusion was based on three main arguments.

First, he asserted that rule 403 was incorporated into all rules contained within the federal rules

Although the United States Supreme Court has not directly addressed the issue,⁸⁷ the overwhelming majority of federal circuits have interpreted rule 404(b) as adopting an inclusionary approach.⁸⁸ These courts stressed the significance of the placement of the words "such as" preceding the list of exceptions,

that did not expressly preclude its application. Because rule 404(b) does not exclude the rule 403 balancing test, he deduced that the test is included. *Id.* at 1480.

Second, Imwinkelried rejected the partial incorporation theory, which maintains that rule 404(b) allows a judge to consider the rule 403 factors but preserves the common-law burden. *Id.* at 1481. Although conceding that this theory is consistent with the literal language of the advisory committee note, he rejected this notion as contrary to the congressional intent to broaden admissibility for uncharged misconduct evidence. *Id.* at 1481-82.

Finally, he rejected the argument that Congress would have made such a sweeping change explicit in the rule if such a change was truly intended. *Id.* at 1483. Imwinkelried asserted that Congress interpreted the language of rule 402, permitting admission of relevant evidence unless otherwise excluded, as precluding the need for an express repudiation of the common law standard. *Id.* at 1483-84.

87. The Supreme Court, however, has addressed the related issue of whether a trial court must make a preliminary finding that the uncharged misconduct occurred before admitting rule 404(b) evidence. See *Huddleston v. United States*, 485 U.S. 681, 685 (1988). The Court held that the proponent of 404(b) evidence is not required to prove by a preponderance of the evidence that the act occurred. *Id.* at 687-89.

Although not directly addressing the issue of which approach Congress intended to adopt in rule 404(b), the Supreme Court, in two distinct portions of its opinion in *Huddleston*, gave conflicting indications of its interpretation of this matter. The Court stated: "Federal Rule of Evidence 404(b) . . . generally prohibits the introduction of evidence of extrinsic acts that might adversely reflect on the actor's character, unless that evidence bears upon a relevant issue in the case such as motive, opportunity, or knowledge." *Id.* at 685 (emphasis added). Although not conclusive, this language intimates a general rule of exclusion.

In contrast, at the conclusion of its opinion, the Court implied that rule 404(b) incorporates an inclusionary approach. To allay the fears of the petitioner, the Court observed that even without a preliminary finding requirement the admission of uncharged misconduct evidence would not be unrestrained because other protective measures bar introduction of unduly prejudicial evidence. *Id.* at 691-92. In dicta the Court stated that one such measure is the incorporation of the rule 403 balancing test into rule 404(b). *Id.* at 691. This interpretation of rule 404(b), if accepted as binding, would effectively reverse the common-law presumption of exclusion. See Imwinkelried, *supra* note 8, at 1484.

88. See, e.g., *United States v. Ayers*, 924 F.2d 1468, 1473 (9th Cir. 1991) ("We have construed Rule 404(b) as being 'a rule of inclusion.'") (quoting *Heath v. Cast*, 813 F.2d 254, 259 (9th Cir.), *cert. denied*, 484 U.S. 849 (1987)); *Berkovich v. Hicks*, 922 F.2d 1018, 1022 (2d Cir. 1991) ("But such evidence may be admitted for any other relevant purpose under our 'inclusionary' approach.") (quoting *United States v. Brennan*, 798 F.2d 581, 589 (2d Cir. 1986), *cert. denied*, 490 U.S. 1022 (1989)); *United States v. Yerks*, 918 F.2d 1371, 1373 (8th Cir. 1990) ("We view Rule 404(b) as a rule of inclusion . . ."); *Virgin Islands v. Edwards*, 903 F.2d 267, 270 (3d Cir. 1990) ("Rule 404(b) is a rule of inclusion, not exclusion . . ."); *United States v. Cohen*, 888 F.2d 770, 776 (11th Cir. 1989) ("The rule [404(b)] is one of inclusion . . ."); *United States v. Acosta-Cazares*, 878 F.2d 945, 948 (6th Cir.) ("Thus, we have explained that Rule 404(b) 'is actually a rule of inclusion rather than exclusion . . .'") (quoting *United States v. Blankenship*, 775 F.2d 735, 739 (6th Cir. 1985)), *cert. denied*, 110 S.Ct. 255 (1989); *Morgan v. Foretich*, 846 F.2d 941, 944 (4th Cir. 1988) ("This Court has held Rule 404(b) to be an 'inclusionary rule . . .'") (quoting *United States v. Masters*, 622 F.2d 83, 85 (4th Cir. 1980)); *United States v. Cuch*, 842 F.2d 1173, 1176 (10th Cir. 1988) ("It is well settled that the rule [404(b)] is one of inclusion . . ."); *United States v. Moore*, 732 F.2d 983, 987 & n.30 (D.C. Cir. 1984) (finding congressional intent in enacting rule 404(b) to adopt an inclusionary approach); *United States v. Jordan*, 722 F.2d 353, 356 (7th Cir. 1983) ("The draftsmen of Rule 404(b) intended it to be construed as one of inclusion, and not exclusion.") (quoting *United States v. Long*, 574 F.2d 761, 766 (3d Cir.), *cert. denied*, 439 U.S. 985 (1978)); *United States v. Ackal*, 706 F.2d 523, 531 (5th Cir. 1983) (en banc) ("This test [adopted by the circuit] takes an inclusionary and not an exclusionary approach.") (quoting *United States v. King*, 703 F.2d 119, 125 (5th Cir.), *reh'g denied*, 711 F.2d 1054 (1983)). But see *United States v. Rodriguez-Cardona*, 924 F.2d 1148, 1153 (1st Cir. 1991) ("Rule 404(b) is a rule of exclusion.").

and concluded that the list of exceptions is only illustrative, not exhaustive.⁸⁹ Consequently, these courts have held that the federal version of rule 404(b) adopts an inclusionary approach, rather than one characterized by a general rule of exclusion with a limited number of enumerated exceptions.⁹⁰

The accepted interpretation of the federal rule is important because the North Carolina General Assembly apparently intended to follow the federal model. The most authoritative support for this interpretation, besides the almost identical language of the two rules, is the commentary following the North Carolina rule. This commentary repeats verbatim the language contained in the federal advisory committee's note.⁹¹ If this was not itself a sufficient indication of an intent to follow the federal rule, the North Carolina commentary appends an illuminating sentence: "The list in the last sentence of subdivision (b) is nonexclusive and the fact that evidence cannot be brought within a category does not mean that the evidence is inadmissible."⁹² Thus it appears that, like Congress, the North Carolina General Assembly's intent in enacting its version of rule 404(b) was to adopt an inclusionary approach.

Although the legislative intent is relatively clear, there are other important policy concerns that warrant discussion in analyzing the *Coffey* court's holding. State courts have repeatedly expressed these policies in retaining the exclusionary approach despite the adoption of rule 404(b) in their respective jurisdictions. For example, in the seminal pre-rule case on "other crimes" evidence in North Carolina, *State v. McClain*,⁹³ the supreme court delineated several of these important policy concerns.

First, the *McClain* court observed that logically the commission of an independent offense did not, in itself, prove the commission of the charged offense.⁹⁴ Second, the admission of evidence that the defendant has been guilty of another heinous crime falsely leads the jury to a belief that he is guilty of the charged crime and, therefore, effectively strips him of the presumption of inno-

89. Imwinkelried, *supra* note 8, at 1468; see, e.g., *Cohen*, 888 F.2d at 776 ("The list provided by the rule is not exhaustive . . ."); *Acosta-Cazares*, 878 F.2d at 948 ("[T]he list of permissible uses of evidence of other crimes or acts set forth in Rule 404(b) is neither exhaustive nor conclusive.") (quoting *United States v. Mendez-Ortiz*, 810 F.2d 76, 79 (6th Cir. 1986), cert. denied, 480 U.S. 922 (1987)); *Moore*, 732 F.2d at 987 n.31 ("The uses specified in the rule [404(b)] are not meant to be exhaustive, but merely illustrative."); *Jordan*, 722 F.2d at 356 ("The rule [404(b)] does not exhaust the purposes for which evidence of other wrongs or acts may be admitted."); *United States v. Johnson*, 634 F.2d 735, 737 (4th Cir. 1980) ("The Rule's [404(b)] list is merely illustrative, not exclusive."), cert. denied, 451 U.S. 907 (1981).

90. Imwinkelried, *supra* note 8, at 1468.

91. See N.C.R. EVID. 404(b) commentary (1986). For the text of the federal advisory committee's note, see *supra* note 85.

92. N.C.R. EVID. 404(b) commentary. The sentence following this one, however, seems contradictory: "Subdivision (b) is consistent with North Carolina practice." Standing alone and interpreted in light of North Carolina practice up until the adoption of rule 404(b), this sentence might be interpreted as intending the application of an exclusionary approach, or so the literal language suggests. In light of various other statements to the contrary in the commentary, however, this sentence is probably best viewed as an anomaly or legislative misunderstanding of the current state of the law in North Carolina prior to adoption.

93. 240 N.C. 171, 81 S.E.2d 364 (1954). For a discussion of the historical significance of *McClain*, see *supra* notes 37-41 and accompanying text.

94. *McClain*, 240 N.C. at 173-74, 81 S.E.2d at 365.

cence.⁹⁵ Finally, introduction of this type of evidence unnecessarily diverts the attention of the jury away from the point in issue and compels the defendant to meet accusations for which he was not charged.⁹⁶

Numerous empirical studies reinforce the underlying purposes of the exclusionary rule by confirming that admission of uncharged misconduct evidence has a significant impact on juries. For example, the Chicago Jury Project⁹⁷ determined that after the proponent disclosed uncharged misconduct evidence, juries employed a "different . . . calculus of probabilities" when determining the guilt or innocence of an accused.⁹⁸ Specifically, the study documented that conviction rates were significantly greater after such disclosure.⁹⁹ Furthermore, in a study conducted on behalf of the National Law and Social Science Foundation,¹⁰⁰ researchers concluded that potential jurors were in substantial agreement in ranking any immoral acts by the defendant as one of the most prejudicial types of evidence.¹⁰¹ The significance of this finding is augmented because the potential jurors exhibited virtually no common evaluations regarding the degree of prejudice for other types of evidence.¹⁰²

The inclusionary approach adopted by the *Coffey* court, however, discards these important arguments against relaxed admissibility of other crimes evidence. A survey of state court decisions on this issue decided under statutes similar to Federal Rule of Evidence 404(b) discloses a striking disparity of interpretation when compared to the rulings of the federal courts.¹⁰³ To date, thirty-four states have enacted an uncharged misconduct rule modeled after Federal Rule of Evidence 404(b).¹⁰⁴ Thirty of these states have adopted statutes that are either verbatim copies of the federal rule or reflect only minor technical changes.¹⁰⁵ Of these states, sixteen have retained the exclusionary approach,¹⁰⁶

95. *Id.* at 174, 81 S.E.2d at 366.

96. *Id.*

97. This title refers to an extensive, exhaustive study of the American jury system conducted at the University of Chicago Law School. This project attempted to combine the research of attorneys and social scientists in studying various aspects of the jury system. See H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* at v (1966).

98. *Id.* at 179.

99. *Id.* at 178-79.

100. This study focused on the effects that different aspects of the legal system have on modeling jury behavior. Professors from the Psychology Department and the School of Law at the University of New Mexico analyzed the aspect of the study dealing with the prejudicial effect of certain evidence. Teitelbaum, Sutton-Barbere & Johnson, *Evaluating the Prejudicial Effect of Evidence: Can Judges Identify the Impact of Improper Evidence on Juries?*, 1983 WIS. L. REV. 1147, 1147.

101. *Id.* at 1162.

102. *Id.* at 1163.

103. For a discussion of the federal circuit courts' almost unanimous interpretation of the federal version of rule 404(b) as a rule of inclusion, see *supra* note 88 and accompanying text.

104. See 1 G. JOSEPH & S. SALTZBURG, *EVIDENCE IN AMERICA: THE FEDERAL RULES IN THE STATES* ch. 14: Rule 404, at 11-24 (1987 & Supp. 1990) (listing the state statutes).

The states adopting a rule modeled after the federal rule include: Alaska, Arizona, Arkansas, Colorado, Delaware, Florida, Hawaii, Idaho, Iowa, Louisiana, Maine, Michigan, Minnesota, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Rhode Island, South Dakota, Texas, Tennessee, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. *Id.* at 1-24.

105. See *id.* Of the states listed in note 104 *supra*, all except Florida, Louisiana, Maine, and

six have expressly adopted the inclusionary approach,¹⁰⁷ and eight states either have not addressed the issue or have reached inconclusive results.¹⁰⁸

The discrepancy between federal and state court interpretations of rule 404(b) stems primarily from most state courts' refusal to disregard the policies underlying the exclusionary rule. These state courts looked primarily to the purposes underlying the exclusionary rule as a basis for their retention of it.¹⁰⁹ Consequently, they implicitly have chosen to disregard the apparent intent of the legislators to adopt an inclusionary rule¹¹⁰ in favor of the policies behind the exclusionary rule.

The *Coffey* court could have followed these state courts and chosen to ignore the intent of the North Carolina General Assembly in favor of the policies governing its case law prior to adoption of rule 404(b). A judicial decision following this route, however, arguably presents a serious danger to the whole structure of the rules of evidence because it ignores the intent of the legislature in favor of judge-made rules.¹¹¹ Rule 402 allows admission of evidence unless it is expressly prohibited by another rule, the United States or North Carolina Constitutions, or other legislative acts.¹¹² Rule 402's fundamental premise—all

Tennessee have adopted a rule which essentially codifies Federal Rule of Evidence 404(b) in their respective states. *See id.*

106. *See, e.g.,* *Burke v. State*, 624 P.2d 1240, 1247-48 (Alaska 1980); *Soper v. State*, 731 P.2d 587, 589-90 (Alaska Ct. App. 1987); *State v. Roscoe*, 145 Ariz. 212, 216-18, 700 P.2d 1312, 1316-18 (1984), *cert. denied*, 471 U.S. 1094 (1985); *State v. Bainbridge*, 108 Idaho 273, 278, 698 P.2d 335, 340 (1985); *People v. Devine*, 168 Mich. App. 56, 58, 423 N.W.2d 594, 596 (1988); *Elmore v. State*, 510 So. 2d 127, 130 (Miss. 1987); *State v. Fitzgerald*, 238 Mont. 261, 265, 776 P.2d 1222, 1225 (1989); *Daly v. State*, 99 Nev. 564, 567, 665 P.2d 798, 801 (1983); *State v. Ferguson*, 391 N.W.2d 172, 174-75 (N.D. 1986); *State v. Broom*, 40 Ohio St. 3d 277, 281-82, 533 N.E.2d 682, 689-90 (1988), *cert. denied*, 490 U.S. 1075 (1989); *Little v. State*, 725 P.2d 606, 607 (Okla. Crim. App. 1986); *State v. Houghton*, 272 N.W.2d 788, 790-91 (S.D. 1978) (en banc), *overruled on other grounds*, *State v. Willis*, 370 N.W.2d 193 (S.D. 1985); *Turner v. State*, 754 S.W.2d 668, 671-72 (Tex. Crim. App. 1988); *State v. Shickles*, 760 P.2d 291, 295 (Utah 1988); *State v. Catsam*, 148 Vt. 366, 380-82, 534 A.2d 184, 193 (1987); *State v. Dolin*, 347 S.E.2d 208, 212 (W. Va. 1986), *overruled on other grounds*, *State v. Edward*, 398 S.E.2d 123 (W. Va. 1990); *Gezzi v. State*, 780 P.2d 972, 974 (Wyo. 1989).

107. *See, e.g.,* *Sullivan v. State*, 289 Ark. 323, 328, 711 S.W.2d 469, 471-72 (1986); *Getz v. State*, 538 A.2d 726, 730-31 (Del. 1988); *State v. Plaster*, 424 N.W.2d 226, 228-29 (Iowa 1988); *State v. Parker*, 127 N.H. 525, 532, 503 A.2d 809, 813 (1985); *Coffey*, 326 N.C. at 278-79, 389 S.E.2d at 54; *State v. Johns*, 301 Or. 535, 544, 725 P.2d 312, 317-18 (1986).

In addition, two other states with substantially modified versions of rule 404(b) have also adopted an inclusionary approach. *See* *Garron v. State*, 528 So. 2d 353, 358 (Fla. 1988); *State v. DeLong*, 505 A.2d 803, 805-06 (Me. 1986).

108. These states include Colorado, Hawaii, Minnesota, Nebraska, New Mexico, Rhode Island, Washington, and Wisconsin.

109. *See, e.g.,* *Houghton*, 272 N.W.2d at 790 (South Dakota); *Turner*, 754 S.W.2d at 671-72 (Texas).

110. Because these decisions are from states with statutes almost identical to the federal rule, it is logical that the intent of the adopters from these states was to follow Congress' intent to adopt an inclusionary rule. For a discussion of congressional intent in enacting rule 404(b), *see supra* notes 83-86 and accompanying text.

111. Imwinkelried, *supra* note 8, at 1491-95.

112. The North Carolina version of Rule 402 reads: "All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by the Constitution of North Carolina, by an Act of Congress, by an Act of the General Assembly, or by these rules. Evidence which is not relevant is not admissible." N.C.R. EVID. 402 (1986).

The commentary following the rule adds: "This rule is identical to Fed. R. Evid. 402 except that the phrases 'by the Constitution of North Carolina' and 'by Act of the General Assembly' were

relevant evidence should be admitted unless expressly excluded by legislative act—is the basis of the federal rules and is designed to eliminate the existence of judge-made rules of evidence.¹¹³ Interpreting rule 404(b) as exclusionary, contrary to its clear language, is in essence a judicial decision which subverts the comprehensive codification effort of the rules.¹¹⁴ Therefore, ignoring legislative intent and preferring the common-law purposes behind the exclusionary approach could precipitate continuous undermining of the foundational principle behind the rules of evidence.¹¹⁵

Thus, although strong support exists in both empirical studies and state case law for the continued application of the exclusionary approach to uncharged misconduct evidence, following these policy concerns could have detrimental effects on the integrity of the North Carolina Rules of Evidence. The North Carolina General Assembly expressed a relatively clear intent to adopt an inclusionary approach when it enacted North Carolina Rule of Evidence 404(b).¹¹⁶ Given the intent of the General Assembly and the limited role of the courts, the *Coffey* court's adoption of an inclusionary approach is wholly justified. Furthermore, because subverting the intent of the General Assembly could have potentially pernicious effects on the integrity and success of the entire structure of the North Carolina Rules of Evidence, the *Coffey* court cannot be faulted for adopting an inclusionary approach, despite the strong policies supporting the exclusionary rule.¹¹⁷

A second significant aspect of the *Coffey* decision is more dubious, however. The North Carolina Supreme Court consistently has held, both before and after the adoption of rule 404(b), that evidence of prior sexual misconduct must satisfy a two-pronged test of substantial similarity and proximity in time to the charged crime.¹¹⁸ For example, in *State v. Boyd*¹¹⁹ the court cited the broad admissibility standard enunciated in post-rule cases that relate to sexual misconduct evidence.¹²⁰ In the next sentence, however, the court stated: "Nevertheless, the ultimate test for determining whether such [uncharged misconduct] evidence is admissible is whether the incidents are sufficiently similar and not so remote in time as to be more probative than prejudicial under the balancing test of . . . Rule 403."¹²¹ Thus, as *Boyd* illustrates, the court previously did not

added and the phrase 'by other rules prescribed by the Supreme Court pursuant to statutory authority' was deleted." *Id.* 402 commentary.

113. 22 C. WRIGHT & K. GRAHAM, *supra* note 2, § 5199, at 220-23.

114. See Imwinkelried, *supra* note 8, at 1493-94.

115. See *id.* at 1494-95. Professor Imwinkelried discusses other instances where courts have ignored the framers' intent in adopting a rule and its deleterious effect on the corresponding evidentiary code. *Id.* at 1494-96.

116. N.C.R. EVID. 404(b) commentary.

117. For a discussion of the appropriate legislative response in light of this empirical support for the exclusionary rule, see *infra* text accompanying notes 141-43.

118. See, e.g., *State v. Rosier*, 322 N.C. 826, 828-29, 370 S.E.2d 359, 360-61 (1988); *State v. Boyd*, 321 N.C. 574, 577, 364 S.E.2d 118, 119 (1988); *State v. Scott*, 318 N.C. 237, 248, 347 S.E.2d 414, 420 (1986). For a discussion of North Carolina cases applying this rule, see *supra* notes 57-77 and accompanying text.

119. 321 N.C. 574, 364 S.E.2d 118 (1988). For a discussion of *Boyd*, see *supra* note 77.

120. *Boyd*, 321 N.C. at 577, 364 S.E.2d at 119.

121. *Id.*

interpret the adoption of a broad admissibility standard as superseding the application of this independent two-pronged test.¹²²

Furthermore, the *Boyd* court expressly acknowledged the grave risks inherent in the adoption of a broad admissibility standard for uncharged misconduct evidence. The court stated: "We are not unmindful of the danger of allowing Rule 404(b) exceptions to become so pervasive that they swallow the rule, a danger vigorously argued in defendant's brief. We, however, do not find that its application to the facts of this case encourages that danger."¹²³

This express concern for the untrammelled admission of other crime evidence is entirely lacking from the court's decision in *Coffey*. In fact, an analysis under the two-pronged test is conspicuously absent from the court's opinion.¹²⁴ The court mentioned the defendant's contention that the incident with Angel Ashe was not similar to the charged murder.¹²⁵ It did not, however, employ the two-pronged test of similarity and proximity in time to reject this argument; instead, the court implied that the broad admissibility standard under rule 404(b) disposed of the defendant's contention.¹²⁶ The court's analysis intimates that it believed the adoption of a broad inclusionary approach under rule 404(b) superseded any previously recognized requirements that the proponent of the other crimes evidence prove its similarity and timeliness.¹²⁷ This reasoning is a dramatic departure from the analysis employed by the court just two years previously in *Boyd*.¹²⁸

It is possible that the *Coffey* court implicitly considered the two-pronged standard in making its analysis. The court concluded that evidence of the Angel Ashe episode satisfied both the motive and specific intent purposes under rule 404(b).¹²⁹ The court found this evidence relevant to prove the intent and motive for the underlying felony of kidnapping for the purpose of committing indecent liberties.¹³⁰ Presumably, the court implicitly concluded that the incident involving Angel Ashe was similar to the State's contention that the defendant had taken indecent liberties with the victim.

The fallacy of this argument, however, is patent: the State had no extrinsic

122. In fact, the *Boyd* court explicitly found that evidence of the defendant's prior acts with a related victim fell squarely within this two-pronged test. *Id.* at 578, 364 S.E.2d at 120.

123. This quotation also calls into question the *Coffey* court's depiction of its post-rule cases as effectively adopting an inclusionary approach. This language suggests that the court is still following a general standard of exclusion. The reference to "exceptions swallowing up the rule" is more consistent with an exclusionary than an inclusionary approach. The standard announced in *State v. Weaver*, two years prior to *Boyd*, however, supports the court's view. *See State v. Weaver*, 318 N.C. 400, 402-03, 348 S.E.2d 791, 793 (1986). Thus, it is probably most accurate to describe the court's approach to rule 404(b) prior to *Coffey* as inconclusive or perhaps inconsistent. This inconsistency only reinforces the view that the *Coffey* court's express adoption of the inclusionary rule is an extremely significant ruling, not the mere reaffirmance of prior case law, as the court asserts. *See Coffey*, 326 N.C. at 278-79, 389 S.E.2d at 54.

124. *See Coffey*, 326 N.C. at 278-81, 389 S.E.2d at 54-56.

125. *Id.* at 278, 389 S.E.2d at 54.

126. *See id.* at 278-79, 389 S.E.2d at 54-55.

127. *See id.*

128. *See supra* notes 119-23 and accompanying text.

129. *Coffey*, 326 N.C. at 280-81, 389 S.E.2d at 55-56.

130. *Id.*

evidence that Coffey took any indecent liberties with Amanda Ray. The only connection between the completely unrelated Angel Ashe incident and the present case was the prosecutor's naked theory.¹³¹ In other words, the court upheld the defendant's conviction of kidnapping with the intent to commit indecent liberties¹³² based solely on the testimony that he committed similar acts in the past in an entirely unrelated incident. Application of this reasoning is tantamount to trying the defendant based on his propensity to commit such acts. As the court stated in *State v. Morgan*,¹³³ the inference that the defendant acted a certain way in the past and thus acted the same way in the present case is "precisely what is prohibited by Rule 404(b)."¹³⁴ In previous cases, the North Carolina Supreme Court consistently required independent extrinsic evidence to establish that the charged sexual crime occurred before it would admit evidence of defendant's prior sexual wrongs.¹³⁵

Thus, the *Coffey* court not only abandoned the two-pronged test requiring evidence of uncharged misconduct to be substantially similar to and not too remote in time from the charged act,¹³⁶ but also failed to require that there be any independent extrinsic evidence that such a similar act actually occurred in the present case. The *Coffey* court disregarded its own post-rule precedents recognizing these requirements as important protections against the admission of marginally relevant but extremely prejudicial evidence.¹³⁷ By abandoning these important procedural prerequisites, the court has unleashed a "prosecutor's delight,"¹³⁸ where the admission of uncharged misconduct evidence is unanchored

131. The State conceded that the Angel Ashe incident was its only evidence of the underlying felony of kidnapping with the intent to commit indecent liberties. Brief for Respondent at 35, *Coffey*, (No. 613A87). As counsel for petitioner noted, without such extrinsic evidence, "the jury is afloat in a sea of speculation." Brief for Petitioner at 49, *Coffey*, (No. 613A87).

132. This conviction supported the felony-murder charge, which in turn resulted in a conviction of first degree murder and a sentence of death. *Coffey*, 326 N.C. at 277, 389 S.E.2d at 53.

133. 315 N.C. 626, 340 S.E.2d 84 (1986). For a discussion of *Morgan*, see *supra* notes 45-51 and accompanying text.

134. *Id.* at 638, 340 S.E.2d at 92 (emphasis added).

135. See, e.g., *State v. Rosier*, 322 N.C. 826, 827, 370 S.E.2d 359, 360 (1988) (victim's testimony that defendant forced her to engage in anal intercourse); *State v. Boyd*, 321 N.C. 574, 577, 364 S.E.2d 118, 120 (1988) (victim testified that stepfather sexually assaulted her on four occasions); *State v. DeLeonardo*, 315 N.C. 762, 768, 340 S.E.2d 350, 355 (1986) (testimony of son that father sexually abused him, his brother, and his sister).

Although the victim's testimony is obviously unavailable in *Coffey*, there are many other forms of evidence which could have been used to prove that the victim had been sexually abused, such as presence of semen or abnormalities in the victim's sexual organs.

136. Once the pretext of the unsupported underlying felony is sheared away, it is apparent that the court has no firm basis on which to admit such evidence, besides propensity—masturbation in front of a young girl has no similarity to murder. Furthermore, numerous North Carolina cases excluded other crimes evidence with less tenuous similarities than those offered in *Coffey*. See, e.g., *State v. Scott*, 318 N.C. 237, 248, 347 S.E.2d 414, 420 (1986) (acts of cunnilingus on three and four-year-old nieces ruled not similar to forced intercourse with sister, the mother of the girls); *State v. Moore*, 309 N.C. 102, 107-08, 305 S.E.2d 542, 545-46 (1983) (two attacks both involving the use of a knife to force the victim to engage in oral sex occurring within Greensboro within a two month period of each other ruled not sufficiently similar).

137. See, e.g., *State v. Boyd*, 321 N.C. 574, 577, 364 S.E.2d 118, 119 (1988); *State v. Scott*, 318 N.C. 237, 244, 248, 349 S.E.2d 414, 418, 420 (1986).

138. This term comes from the characterization of uncharged misconduct evidence in a student-written law review article. Comment, *Evidence of Prior Acquittals: An Attack on the "Prosecutors Delight,"* 21 UCLA L. REV. 892, 896 (1974).

by any protective safeguards to a defendant's right to a fair trial.

Moreover, the language of the opinion is alarmingly general and does not appear to be limited to the particular facts in *Coffey* or even just to evidence of prior sexual acts.¹³⁹ Thus, the ruling potentially could be applied to all types of cases involving the introduction of uncharged misconduct evidence. Without the two-pronged requirements, the State may freely admit evidence of prior misconduct that is dissimilar and distant in time from the charged act, so long as the prosecutor can conceive of any possible theory under which it may be relevant. Furthermore, without a requirement that there be some extrinsic evidence of the charged crime, a prosecutor desiring to introduce evidence of a defendant's uncharged behavior could merely add a plausible, related charge to the indictment and thereby gain admittance of evidence that is damaging and highly prejudicial to the defendant.

The North Carolina Supreme Court's ruling in *Coffey* is significant in several important respects. First, the court expressly adopted an inclusionary approach to rule 404(b) and rejected the traditional exclusionary approach followed at common law. Second, the court's ruling implies that the adoption of this broad approach to admissibility of other crimes evidence precludes the application of the previously recognized two-pronged test which served to protect the defendant against introduction of unrelated and unduly prejudicial evidence. Finally, the general language of the court's opinion intimates that this holding is not limited to the particular facts of *Coffey*, but could apply to all cases adjudicating the admissibility of uncharged misconduct evidence.

Strong empirical data and other states' case law support the continued application of the exclusionary approach to uncharged misconduct evidence.¹⁴⁰ The *Coffey* court is bound jurisprudentially, however, to follow the North Carolina General Assembly's intent to adopt an inclusionary approach in enacting rule 404(b). Given the limited role of the courts, a better alternative to court renunciation of legislative intent would be for the general assembly to amend rule 404(b).¹⁴¹ The amendment could take into account the strong empirical support indicating the highly prejudicial effect of uncharged misconduct evidence as well as the decisions of other state courts retaining the exclusionary approach without jeopardizing the integrity of either the North Carolina Rules of Evidence or the North Carolina court system.

This Note recommends adoption of the following amendment, proposed by a distinguished commentator on rule 404(b) evidence:

Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show that he acted in conformity therewith. It may however, be admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident. Before the judge admits evidence for such a purpose, the proponent of the evidence must

139. See *Coffey*, 326 N.C. at 278-79, 389 S.E.2d at 54-55.

140. See *supra* notes 97-102 & 103-10 and accompanying texts.

141. Imwinkelried, *supra* note 8, at 1497.

persuade the judge that the probative value of the evidence outweighs the danger of unfair prejudice.¹⁴²

This amendment would restore the common-law burden of proving that the evidence should be admitted to the proponent of uncharged misconduct evidence.¹⁴³ Consequently, this reformulation of rule 404(b) would recognize the important policies behind the original adoption of the exclusionary rule by re-establishing the presumption against admittance of this often unduly prejudicial type of evidence.

In addition, the court should reaffirm its commitment to the requirements of substantial similarity and proximity in time as prerequisites to admission of uncharged misconduct evidence. These requirements serve a useful and necessary purpose by guarding against untrammelled admission of propensity evidence. Without these important safeguards, as the ruling in *Coffey* indicates, there exists a substantial danger that the prosecutor will misuse uncharged misconduct evidence essentially to put a defendant on trial for prior bad acts and not for those crimes properly contained in the indictment.

Coffey was not a sympathetic defendant and it is difficult to be outraged by the procedural error which occurred in this particular case. Nevertheless, this rationale does not justify the *Coffey* court's ruling establishing a potentially pernicious precedent in this important area of evidence law. The *Coffey* court allows the State to introduce evidence of a defendant's unrelated prior sexual acts without any supporting extrinsic evidence that a similar act occurred in connection with the actions for which the defendant is specifically charged. This approach is contrary not only to all North Carolina precedent on the subject, but also, and more importantly, to the fundamental postulate of the American accusatorial system of criminal justice—an accused can be tried only on the charges against her.¹⁴⁴

DOUGLAS J. BROCKER

142. *Id.*

143. *Id.*

144. *Cf. Reed, Trial by Propensity, supra* note 79, at 713 (asserting that the exclusionary rule originated in the accusatorial system, which requires the State to prove the defendant is guilty of a particular crime and not merely past uncharged misconduct).

State v. Strickland: Evening the Odds in Rape Trials! North Carolina Allows Expert Testimony on Post Traumatic Stress Disorder to Disprove Victim Consent

Rape trials traditionally have stood alone among criminal proceedings as examinations not of the defendant's actions, but of the victim's conduct, life-style, and personal history.¹ As a result, rape consistently has been the most difficult violent crime to prove.² Within the last twenty years, however, a concerted, albeit scattered, effort has begun to sweep away some of the archaic barriers facing prosecutors at trial.³ Recently one of the largest of these barriers, the difficulty of disproving consent,⁴ has come under fire as prosecutors seek to use expert testimony concerning post traumatic stress disorder (PTSD)⁵ to ex-

1. N. GAGER & C. SCHURR, *SEXUAL ASSAULT: CONFRONTING RAPE IN AMERICA* 129-30 (1976) (pointing to the patriarchal nature of society and its ramifications on rape trials, which focus on any possible "seduction, lying, mistaken identification, or 'wanton' behavior" by the victim). For further examinations on the tendency of rape trials to investigate the victim, see S. BESSMER, *THE LAWS OF RAPE* (1984); A. BURGESS & L. HOLMSTROM, *RAPE: CRISIS AND RECOVERY* (1979); T. MCCAHERN, L. MEYER & A. FISCHMAN, *THE AFTERMATH OF RAPE* (1979); C. WARNER, *RAPE AND SEXUAL ASSAULT* (1980); J. WILLIAMS & K. HOLMES, *THE SECOND ASSAULT: RAPE AND PUBLIC ATTITUDES* (1981).

2. Sir Mathew Hale's often-quoted statement on rape prosecution symbolizes the enormous barriers standing before rape victims:

It is true rape is a most detestable crime, and therefore ought severely and impartially to be punished with death; but it must be remembered, that it is an accusation easily to be made and hard to be proved, and harder to be defended by the party accused, tho never so innocent.

1 M. HALE, *HISTORY OF THE PLEAS OF THE CROWN* 634 (1847).

Wigmore carried this fear of a false complaint and the resulting caution toward victims' claims into modern times. "One form taken by these complexes is that of contriving false charges of sexual offenses by men. The unchaste (let us call it) mentality finds incidental but direct expression in the narration of imaginary sex incidents of which the narrator is the heroine or the victim." 3A J. WIGMORE, *WIGMORE ON EVIDENCE* 924a, at 736 (Chadbourn rev. 1970).

3. Note, *Checking the Allure of Increased Conviction Rates: The Admissibility of Expert Testimony on Rape Trauma Syndrome in Criminal Proceedings*, 70 VA. L. REV. 1657, 1662-63 (1984).

4. Gager and Schurr described the uniqueness of the legal system's focus on consent in rape trials, a focus not found in other criminal trials, and the difficulty this focus creates for victims:

The evolution of our legal process has thrust the consent issue into the center of the storm about rape. The legal obverse of consent is resistance, resistance sufficient to prove lack of consent. However, resistance, or nonconsent, is largely a subjective matter, and hundreds, if not thousands, of rapists have been allowed to go free by police, juries, and judges who have decided arbitrarily that the victim did not resist "enough."

N. GAGER & C. SCHURR, *supra* note 1, at 139-40.

5. Post traumatic stress disorder (PTSD) involves the development of a series of characteristic symptoms (physical and psychological) after an individual undergoes a particularly stressful situation. In the context of rape, psychologists describe a subcategory of PTSD as "the acute phase and long-term reorganization process that occurs as a result of forcible rape or attempted rape." Burgess & Hunter, *Rape Trauma Syndrome*, in *THE RAPE VICTIM* 121 (D. Nass ed. 1977). When rape is the cause of this process commentators sometimes refer to the resulting PTSD syndrome as "rape trauma syndrome" (RTS). Burgess & Holmstrom, *Rape Trauma Syndrome*, 131 AM. J. PSYCHIATRY 981, 982 (1974) [hereinafter Burgess & Holmstrom, *Rape Trauma Syndrome*]; Burgess & Holmstrom, *Rape Trauma Syndrome and Post Traumatic Stress Response*, in *RAPE AND SEXUAL ASSAULT: A RESEARCH HANDBOOK* 49 (A. Burgess ed. 1985) [hereinafter Burgess & Holmstrom, *RTS and PTSD*]; Wilson, Smith & Johnson, *A Comparative Analysis of PTSD Among Various Survivor Groups*, in *TRAUMA AND ITS WAKE: THE STUDY AND TREATMENT OF POST-TRAUMATIC*

plain the victim's suffering and defeat the defense's contention that the victim consented.

Four years after the issue first came before North Carolina courts,⁶ the North Carolina Court of Appeals ruled in *State v. Strickland*⁷ that expert testimony on PTSD is relevant and admissible to disprove consent in rape trials.⁸ In *Strickland* the trial court admitted, over the defendant's objections, expert testimony that the victim suffered from PTSD.⁹ The appellate court held that such testimony is relevant in rape cases and affirmed the ruling, but failed to address directly whether PTSD evidence is universally admissible or remains limited to cases similar to *Strickland*.¹⁰

This Note examines the *Strickland* decision in light of the controversial history of PTSD expert testimony and the resulting split among the states regarding its admissibility. After exploring the arguments proffered against allowing PTSD testimony, this Note concludes that both the relevance and reliability of the technique justify the court's decision to allow such testimony in limited situations. Finally, this Note discusses the failure of the *Strickland* court to discuss the boundaries of PTSD testimony's admissibility, as well as the implications of the appellate court's vague holding and the North Carolina Supreme Court's refusal to review the decision.¹¹

Strickland centered on the abduction of a woman as she left a mall in Raleigh, North Carolina. The defendant, Wendell Wade Strickland, forced the vic-

STRESS DISORDER 142, 167 (C. Figley ed. 1985). For a discussion of the specific nature of PTSD and RTS, see *infra* notes 44-50 and accompanying text.

Whether one refers to this syndrome as PTSD or RTS, the basic symptoms remain the same, but within this general PTSD framework rape victims display certain peculiar behavior patterns, such as a tendency to delay reporting the incident, memory failure immediately following the rape, a desire to change daily living patterns, and a fear of men. *People v. Taylor*, 75 N.Y.2d 277, 281, 552 N.E.2d 131, 135, 552 N.Y.S.2d 883, 887 (1990) (noting the likelihood that RTS sufferers will exhibit a heightened fear of men); Burgess & Holmstrom, *Rape Trauma Syndrome*, *supra*, at 983-84 (describing the desire of many rape victims to move in order to overcome their fear that the rapist will find them again); Veronen, Kilpatrick & Resick, *Treating Fear and Anxiety in Rape Victims: Implications for the Criminal Justice System*, in PERSPECTIVES ON VICTIMOLOGY 148, 151 (W. Parsonage ed. 1979) (immediately after a rape occurs many rape victims suffer from short-term memory loss); S. KATZ & M. MAZUR, UNDERSTANDING THE RAPE VICTIM: A SYNTHESIS OF RESEARCH FINDINGS, 188-90 (1979) (citing several studies indicating that a large percentage of women delay in reporting the rape and noting that women who know their assailants wait even longer before filing a police report).

6. *State v. Stafford*, 317 N.C. 568, 575, 346 S.E.2d 463, 468 (1986) (The court stated "we do not deem it necessary to reach on this record the question whether in a proper case testimony about rape trauma syndrome will be admissible in the courts of this state."), *abrogation recognized by State v. Strickland*, 96 N.C. App. 642, 387 S.E.2d 62 (1990).

7. 96 N.C. App. 642, 387 S.E.2d 62, *disc. rev. denied*, 326 N.C. 486, 392 S.E.2d 100 (1990).

8. *Id.* at 648, 387 S.E.2d at 66.

9. *Id.* at 646, 387 S.E.2d at 65.

10. *Id.* at 648, 387 S.E.2d at 66.

11. A recent appellate court case has clarified, somewhat, the boundaries of PTSD testimony in sexual abuse cases. *State v. Hall*, 98 N.C. App. 1, 390 S.E.2d 169, *disc. rev. allowed*, 327 N.C. 486, 397 S.E.2d 228 (1990). In *Hall* the court upheld the trial court's admission of expert testimony to help the jury determine if a rape had in fact occurred. *Id.* at 8, 390 S.E.2d at 172-73. This holding answered one question that remained after *Strickland*, but failed to address a number of other concerns stemming from the *Strickland* decision. See *infra* text accompanying notes 163-73.

The North Carolina Supreme Court has granted certiorari in *Hall* and one hopes it will establish a clear framework for the proper usage of PTSD and RTS testimony.

tim into her car, demanded that she drive to a deserted road, and then raped her.¹² Following this incident Strickland took her to his home, where the two spent the night in his bedroom.¹³ The next morning he took the victim to another location in Raleigh and released her.¹⁴

At trial Strickland's roommate, who was present during the night in question, testified that he heard lovemaking sounds coming from Strickland's bedroom.¹⁵ He also stated that he followed Strickland and the victim on the following morning and witnessed the victim kissing Strickland good-bye as he got out of her car.¹⁶ The State called four witnesses to contradict this testimony,¹⁷ including Dr. Susan Roth, a clinical psychologist, who testified that the victim's symptoms were consistent with PTSD.¹⁸ The defense objected, but the court allowed her testimony,¹⁹ and the jury later convicted Strickland of second degree rape.²⁰

On appeal the defense argued that Dr. Roth was not qualified to offer testimony on PTSD and, further, that testimony regarding PTSD is inadmissible in rape trials.²¹ In a perfunctory opinion the appellate court disagreed. The court first found Dr. Roth qualified²² and then followed a line of cases from various states that held PTSD testimony admissible in rape trials by affirming the allowance of Dr. Roth's testimony.²³ The court also found support for its decision in a dissenting opinion from Judge Martin in an earlier case, in which he argued that PTSD is relevant in rape cases,²⁴ and in the recent case of *State v. Clemmons*,²⁵ which implied that PTSD is now admissible in rape cases.²⁶ The North

12. *Strickland*, 96 N.C. App. at 643, 387 S.E.2d at 63.

13. *Id.*

14. *Id.*

15. *Id.*

16. *Id.* at 644, 387 S.E.2d at 63.

17. Three of these witnesses testified that the defendant had urged his roommate to exaggerate and lie to the police about the incident, and that the roommate initially gave more "exaggerated" and deceptive responses to police questions. *Id.* at 644, 387 S.E.2d at 64.

18. *Id.* at 645, 387 S.E.2d at 64. She also testified as to the nature of PTSD and its appearance in rape victims. *Id.*

19. *Id.*

20. *Id.* at 644, 387 S.E.2d at 63.

21. *Id.* at 645-46, 387 S.E.2d at 64-65.

22. *Id.* at 646, 387 S.E.2d at 65. The court noted Dr. Roth's position as an associate professor at Duke University and her extensive research in the areas of sexual trauma, sexual aggression, stress, coping, and helplessness. *Id.*

23. *Id.* at 646-48, 387 S.E.2d at 65-66. The court recognized, but rejected, another line of cases holding PTSD or RTS inadmissible and also mentioned that the American Psychiatric Association recognizes PTSD. *Id.* (citing AMERICAN PSYCHIATRIC ASSOCIATION, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 236 (3d ed. 1980) [hereinafter DSM III]).

24. Specifically, Judge Martin asserted that PTSD is relevant "to assist jurors in understanding the evidence and in drawing appropriate conclusions therefrom." *Id.* at 648, 387 S.E.2d at 66 (quoting *State v. Stafford*, 77 N.C. App. 13, 24, 334 S.E.2d 799, 803 (1985) (Martin, J., dissenting), *aff'd*, 317 N.C. 568, 346 S.E.2d 463 (1986)).

25. 319 N.C. 192, 353 S.E.2d 209 (1987). In concluding that admission of testimony regarding the defendant's prior sexual misconduct was improper but harmless error, the court pointed to PTSD testimony as a deciding factor supporting the defendant's conviction.

Considering the general consistency between the victim's testimony and her pre-trial statements and conduct, the evidence that the victim's house was in disarray following what the defendant contended was a consensual sexual union, and particularly the medical evidence

Carolina Supreme Court subsequently denied discretionary review.²⁷

Traditionally in rape cases the victims have undergone two ordeals—the initial rape and the equally traumatic trial procedure.²⁸ Ironically, society's fear and loathing of this crime helped shift attention to the victim rather than her attacker.²⁹ Virtually the entire inquiry has centered on the victim's behavior, appearance, and past conduct.³⁰ This skewed focus resulted from a number of popular misconceptions regarding rape and its victims.³¹ First and foremost, society and the courts have viewed rape as a sexual crime rather than as one of violence.³² Thus, the investigation often hinged on the impact the victim's behavior and appearance had upon her assailant. If she expressed, in any manner, a desire for his affections, the attacker could seize upon a ready-made defense to his crime.³³ Similarly, the fear of false accusations of rape and the perception that "it is more probable that an unchaste woman would assent . . . than a

of the victim's severe post-traumatic stress disorder for a lengthy period immediately following the incident, we conclude that there is no reasonable possibility that the jury would not have convicted defendant even if the evidence in question had not been admitted for any purpose.

Id. at 199-200, 353 S.E.2d at 213.

26. *Strickland*, 96 N.C. App. at 648, 387 S.E.2d at 66 (citing *Clemmons*, 319 N.C. at 199, 353 S.E.2d at 213).

27. *State v. Strickland*, 326 N.C. 486, 392 S.E.2d 100 (1990).

28. See N. GAGER & C. SCHURR, *supra* note 1, at 129-30; Berger, *Man's Trial, Woman's Tribulation: Rape Cases in the Courtroom*, 77 COLUM. L. REV. 1, 13-14 (1977); Bohmer, *Judicial Attitudes Toward Rape Victims*, 57 JUDICATURE 303, 303 (1974). Bohmer stated that "[v]ictims frequently report that their encounters with the police, district attorneys and courtroom personnel were more traumatic than the rape incident itself." Bohmer, *supra*, at 303.

29. One author explains the tendency to concentrate on the victim rather than the actual crime observed:

"[T]he mention of rape makes us all uneasy—for different reasons depending on who we are. . . . [T]he thoughtful normal man, after hearing the details of a forcible rape, finds it difficult to believe. . . . He knows that all thoughts of sex—which he equates with fun, romance, and mutual admiration—would leave him if the woman were really struggling to get free. . . . He does not realize that, to the rapist, the act is not 'love,' nor ardor, and usually not even passion; it is a way of debasing and degrading a woman. . . . This gives rise to the commonly held view that: 'There is no such thing as rape.' . . . To most women, [rape] is almost as unreal as it is to most men because they themselves have not experienced it, and few people who have done so are in the habit of talking about it. . . .

At the same time, an occasional newspaper story about a particularly brutal rape-murder makes all women shudder. They wonder if it could possibly happen to them, and if it did, how would they react. . . .

. . . .

One way of coping . . . is to imagine, and then believe, that women to whom rape happens are in some way vastly different from oneself. Deciding that they must have been taller, shorter, fatter, thinner, older, or younger will not work since rape victims come in all variations of these attributes. It is far easier to settle on some impalpable quality which is not so easily measured with ruler or scale. This accounts for the overwhelming number of women who believe that most claims of being raped are either outright lies, or that the rapes were brought on by the victim herself. . . ."

Massaro, *Experts, Psychology, Credibility, and Rape: The Rape Trauma Syndrome Issue and Its Implications for Expert Psychological Testimony*, 69 MINN. L. REV. 395, 399 n.27 (1985) (quoting C. HURSCH, *THE TROUBLE WITH RAPE* 5-7 (1977)).

30. Berger, *supra* note 28, at 12-15; Note, *supra* note 3, at 1661-62.

31. See *infra* note 129 and accompanying text.

32. See Note, *supra* note 3, at 1657.

33. See *supra* note 1.

virtuous woman' ”³⁴ led to a system that put the victim on trial along with her attacker.

Consequently, the issue of consent is central in most rape trials. The North Carolina rape statute provides: “A person is guilty of rape in the first degree if the person engages in vaginal intercourse . . . [w]ith another person by force and against the will of the other person”³⁵ Hence, any evidence implying that the victim assented to the accused’s sexual advances becomes relevant to whether a rape occurred. Traditionally courts allowed *any* testimony tending to show victim consent or a failure to resist physically,³⁶ including evidence of the victim’s prior sexual history.³⁷ This practice caused rape trials to revolve around the victim’s behavior and personal history.³⁸

In response to growing pressures for a solution to the imbalances in rape trials³⁹ virtually all states have enacted rape shield statutes, which limit exploration into the victim’s past.⁴⁰ North Carolina passed a rape shield statute in 1977 strictly confining explorations into the victim’s past sexual behavior.⁴¹ On several occasions the North Carolina courts have upheld the constitutionality of this statute and “‘[rejected] the notion that all sexual behavior, however proved, has some intrinsic relevance in a sexual assault proceeding, and [required] a

34. Berger, *supra* note 28, at 15 (quoting *People v. Collins*, 25 Ill. 2d 605, 611, 186 N.E.2d 30, 33 (1962)). In fact, while the F.B.I. reported that on a national average 15% of all rape reports were ‘unfounded’ and research based on police statistics turned up even higher percentages, studies performed by medical and social workers yield much lower figures (from 1 to 7%). These differences were explained by the fact that the police statistics included cases not investigated because the officers believed that the victim was lying. S. KATZ & M. MAZUR, *supra* note 5, at 208-14.

35. N.C. GEN. STAT. § 14-27.2(a) (1986). For a similar description of second-degree rape, see *id.* § 14-27.3 (a)(1) (requiring force against the will of another person).

36. *State v. Dill*, 184 N.C. 645, 652, 113 S.E. 609, 613 (1922) (allowing testimony that the victim failed to cry out and subsequently failed to report the rape for three days as relevant to her credibility and the issue of consent).

37. See, e.g., *State v. Banks*, 295 N.C. 399, 409, 245 S.E.2d 743, 750 (1978); *State v. Goss*, 293 N.C. 147, 153, 235 S.E.2d 845, 849 (1977); *State v. Grundler*, 251 N.C. 177, 191-92, 111 S.E.2d 1, 11-12 (1959), *cert. denied*, 362 U.S. 917 (1960).

38. This focus upon the victim is a major reason for the massive underreporting of rapes. In 1978 only eight percent of the total number of rapes and attempted rape incidents were reported to the police. Russell, *The Prevalence and Incidence of Forcible Rape and Attempted Rape of Females*, 7 VICTIMOLOGY: AN INT’L J. 81, 81 (1982).

39. The pressure to reform has come from all angles, including the feminist movement, social commentators, and changing public attitudes toward rape. Note, *supra* note 3, at 1657 n.1.

40. See Galvin, *Shielding Rape Victims in the State and Federal Courts: A Proposal for the Second Decade*, 70 MINN. L. REV. 763, 765 n.3 (1986) (noting that 48 jurisdictions have rejected the notion that sexual acts of victims are per se admissible).

41. N.C.R. EVID. 412. This statute provides:

[T]he sexual behavior of the complainant is irrelevant to any issue in the prosecution unless such behavior:

- (1) Was between the complaint [sic] and the defendant; or
- (2) Is evidence of specific instances of sexual behavior offered for the purpose of showing that the act or acts charged were not committed by the defendant; or
- (3) Is evidence of a pattern of sexual behavior so distinctive and so closely resembling the defendant’s version of the alleged encounter . . . as to tend to prove that complainant consented to the act . . . or behaved in such a manner as to lead the defendant reasonably to believe that the complainant consented; or
- (4) Is evidence of sexual behavior offered as the basis of expert psychological or psychiatric opinion that the complainant fantasized or invented the act or acts charged.

Id.

more specific showing of relevance before such behavior can be proved.'"⁴² These decisions reflect a concerted effort by the courts to protect rape victims from defense counsel's arguably prejudicial⁴³ attempts at uncovering prior sexual exploits.

At the same time that the state legislatures began to enact rape shield statutes, two researchers published a report on an acute stress reaction observed in rape victims following the event.⁴⁴ Labelling this condition rape trauma syndrome (RTS),⁴⁵ Burgess and Holmstrom divided the stress reaction into two phases. In the initial "acute phase" the victim experiences both physical and emotional reactions. The physical effects include shock, muscle tension, and gastrointestinal irritability, while the emotional effects range from fear, anger, humiliation, and a desire for revenge to an outwardly calm presence that masks inner torment.⁴⁶ During this phase the victim most acutely feels the impact of the rape, although her expressions can range a wide gamut.⁴⁷ In the subsequent reorganization phase the victim undergoes a long-term process of redefinition.⁴⁸ The symptoms of this phase include change in lifestyle and residence, nightmares, and phobic reactions.⁴⁹ Although these symptoms vary in degree and sequence, few victims report an absence of any symptoms.⁵⁰

A number of researchers questioned the methodology of the Burgess and Holmstrom study and other, similar studies that followed it.⁵¹ Specifically, they criticized the lack of a control group, the selective nature of the sample group, and the lack of long-term assessment of the subjects.⁵² Furthermore, several

42. *State v. Fortney*, 301 N.C. 31, 37, 269 S.E.2d 110, 113 (1980) (quoting DETAILED COMMENTS ON DRAFT LAW, LEGISLATIVE RESEARCH COMMISSION, REPORT TO THE GEN. ASSEMBLY OF NORTH CAROLINA: SEXUAL ASSAULTS 92 (1977)); accord *State v. Herring*, 322 N.C. 733, 743-44, 370 S.E.2d 363, 370 (1988).

43. See H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* 249-57 (1966). But see Tanford & Bocchino, *Rape Victim Shield Laws and the Sixth Amendment*, 128 U. PA. L. REV. 544, 544 (1980). Tanford and Bocchino argue that the rape shield laws are ineffective in reducing jury prejudice and violate the sixth amendment guarantee of the right to introduce evidence when the probative value outweighs prejudicial effects. Tanford & Bocchino, *supra* at 545, 572-78.

44. Burgess & Holmstrom, *Rape Trauma Syndrome*, *supra* note 5. The researchers conducted a one-year study of 92 adult victims of forcible rape. These victims were selected from 146 patients admitted to the emergency ward of Boston City Hospital. *Id.* at 981.

45. *Id.* at 981. DSM III recognizes a finite number of other psychologically traumatic events as causes of PTSD, including military combat, natural disasters, accidental disasters (such as severe car accidents or airplane disasters), fires, collapsed buildings, and deliberately caused traumas (such as torture, death camps, and bombing). Common experiences such as loss of loved ones, illnesses, business losses, and marital conflict do not cause PTSD. AMERICAN PSYCHIATRIC ASSOCIATION, *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS* 247-48 (3d ed. rev. 1987) [hereinafter DSM III-R].

46. Burgess & Holmstrom, *Rape Trauma Syndrome*, *supra* note 5, at 982.

47. *Id.* at 982-83.

48. *Id.* at 983-84.

49. *Id.*

50. *Id.* at 983.

51. See Ruch & Leon, *Type of Sexual Assault Trauma: A Multidimensional Analysis of a Short-term Panel*, 8 VICTIMOLOGY: AN INT'L J. 237, 238-39 (1983); S. KATZ & M. MAZUR, *supra* note 5, at 20-27.

52. See, e.g., Ruch & Leon, *supra* note 51, at 238-39; Note, *supra* note 3, at 1670-75; Note, *Expert Testimony On Rape Trauma Syndrome: An Argument For Limited Admissibility*—*State v. Black*, 63 WASH. L. REV. 1063, 1065 n.9 (1988) [hereinafter Note, *Expert Testimony*].

researchers argued that the studies failed to identify any reactions specific to rape.⁵³

In response to these criticisms researchers recently have performed a number of studies designed to correct the methodological flaws of the Burgess and Holmstrom research.⁵⁴ Most of these studies have used control groups of non-rape victims, a battery of standardized psychological tests, and a wider sampling of subjects.⁵⁵ The results of these more scientific tests did not contradict Burgess and Holmstrom's findings.⁵⁶ On the contrary, they reinforced and extended the conclusion that rape victims consistently suffer higher levels of fear, anxiety, depression, and anger than nonvictims.⁵⁷ As a result of the original and subsequent studies confirming the trauma associated with rape, the American Psychiatric Association, in the 1980 edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM III), recognized rape as one of a small number of causes of PTSD.⁵⁸ RTS represents one of several sub-categories of PTSD, which a number of especially traumatic stressors can cause.⁵⁹ PTSD generally causes the sufferer to re-experience the traumatic event,⁶⁰ feel estranged from others,⁶¹ exhibit fear and avoidance of the stimuli associated with the event,⁶² and display and feel a wide range of emotions.⁶³ Not only have the recent studies confirmed rape victims' symptoms that correspond with PTSD,⁶⁴ but at least one researcher has contended that rape victims as a group exhibit

53. In fact, the original study by Burgess and Holmstrom found similar reactions among rape and attempted rape victims. Burgess & Holmstrom, *Rape Trauma Syndrome*, *supra* note 5, at 982. Another study concluded that RTS reactions could result from any stressful sexual situation. Notman & Nadelson, *The Rape Victim: Psychodynamic Considerations*, 133 AM. J. PSYCHIATRY 408, 408 (1976). Some courts later cited these studies as evidence that PTSD or RTS represented a therapeutic technique and could not be used as a fact-finding tool. See *People v. Bledsoe*, 36 Cal. 3d 236, 251, 681 P.2d 291, 301, 203 Cal. Rptr. 450, 460 (1984).

Even more disturbing were results compiled by Symonds, which indicated that similar trauma could result from any violent crime. Specifically, Symonds compared rape and robbery victims and found similar victim responses. Symonds, *The Rape Victim: Psychological Patterns of Response*, 36 AM. J. PSYCHOANALYSIS 27 (1976).

54. See, e.g., Becker, Skinner & Able, *Sequelae of Sexual Assault: The Survivor's Perspective*, in *THE SEXUAL AGGRESSOR: CURRENT PERSPECTIVES ON TREATMENT* 240 (1983); Ellis, *A Review of Empirical Rape Research: Victim Reactions and Response to Treatment*, 3 CLINICAL PSYCHOLOGY REV. 473 (1983); Veronen, Kilpatrick & Resick, *supra* note 5.

55. See, e.g., studies cited *supra* note 54.

56. *Id.*

57. See, e.g., Becker, Skinner & Abel, *supra* note 54, at 249, 253; Ellis, *supra* note 54, at 477.

58. DSM III, *supra* note 23, at 236. The revised third edition of the DSM III continues to recognize rape as a cause of PTSD. DSM III-R, *supra* note 45, at 247.

59. See *supra* note 45.

60. DSM III-R, *supra* note 45, at 247.

61. *Id.* at 248.

62. *Id.* This avoidance of anything associated with the event can result in partial amnesia with respect to the rape, a sudden change in lifestyle to avoid reminders of the incident (such as a change in residence if the rape took place in the victim's home), and even a decreased response to all stimuli known as "psychic numbing." *Id.* at 248-49.

63. *Id.* at 247-49. As shown in the original Burgess and Holmstrom study, see Burgess & Holmstrom, *Rape Trauma Syndrome*, *supra* note 5, these emotional symptoms can range from open displays of terror, depression, and anger to a stoic exterior that hides the inner torment of the victim. DSM III-R, *supra* note 45, at 247-49.

64. See Burgess & Holmstrom, *RTS and PTSD*, *supra* note 5, at 46; Martin, Warfield & Braen, *Physician's Management of the Psychological Aspects of Rape*, 249 J. A.M.A. 501 (1983).

symptoms more severe than those shown by almost any other group of PTSD sufferers.⁶⁵

As the documentation of PTSD (and its sub-category RTS) emerged in support of Burgess and Holmstrom's original conclusions,⁶⁶ the specter of its possible use in rape trials began to loom largely. Properly classified as scientific testimony, PTSD evidence must meet both the general requirements of relevance⁶⁷ and the test for admitting expert testimony.⁶⁸ This testimony may "embrace an ultimate issue to be decided by the trier of fact,"⁶⁹ but it cannot invade the province of the jury by simply telling the jury how to decide an issue.⁷⁰ In addition, evidence of PTSD must meet the jurisdiction's admissibility test for novel scientific theories. Many states apply the so-called *Frye* test,⁷¹ which requires that the scientific community "generally accept" the theory before courts may sanction its use.⁷² This standard represents a strict and unyielding approach towards novel scientific theories, and many argue it goes too far by barring highly reliable and relevant testimony.⁷³ Not only must the PTSD evidence cross the *Frye* barrier, but it must also withstand the skepticism surrounding psychological testimony in general.⁷⁴

Many states, however, have rejected the *Frye* test for novel scientific techniques and instead admit such techniques if they are reliable.⁷⁵ North Carolina courts, in particular, use this "reliability standard."⁷⁶ Although the technique

65. Wilson, Smith & Johnson, *supra* note 5, at 157 (determining that only PTSD caused by combat in Vietnam surpasses the suffering rape victims endure).

66. See *supra* notes 54-65 and accompanying text.

67. Relevancy consists of evidence "having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." N.C.R. EVID. 401.

68. Because PTSD evidence relies on an expert's diagnosis and testimony, it must meet the requirements of rule 702. Thus, the witness must qualify as an expert and his or her testimony must assist the trier of fact in making its conclusions. N.C.R. EVID. 702.

69. N.C.R. EVID. 704.

70. FED. R. EVID. 704 advisory committee notes (The Federal Rule is identical to the North Carolina version of 704.). In the context of PTSD, this means that the expert may explain PTSD and, depending on the jurisdiction, testify that the victim suffers from PTSD. Neither the expert nor the court, however, may order the jury to connect the PTSD to the rape or even to accept the expert's opinion.

71. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

72. *Id.* at 1014. General acceptance among the scientific community can be shown through use of the theory in other cases, and discussion in law review articles and other scholarly or authoritative materials. Massaro, *supra* note 29, at 434-35.

73. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 COLUM. L. REV. 1197, 1223-24 (1980).

74. The fear surrounding psychological testimony is that it relies too much on theory, cannot be validated, and draws impermissible conclusions for the trier of fact. Morse, *Crazy Behavior, Morals, and Science: An Analysis Of Mental Health Law*, 51 S. CAL. L. REV. 527, 618 (1978). *Contra*, Bonnie & Slobogin, *The Role of Mental Health Professionals in the Criminal Process: The Case For Informed Speculation*, 66 VA. L. REV. 427, 463-66 (1980) (arguing that any shortcomings in the psychological testimony may be countered by cross-examination and limiting instructions and, furthermore, that psychological testimony is highly relevant and accurate).

75. Massaro, *supra* note 29, at 435. Of course, general acceptance can bolster a claim for reliability. *Id.*

76. See *State v. Bullard*, 312 N.C. 129, 147, 322 S.E.2d 370, 380 (1984); *State v. Temple*, 302 N.C. 1, 12, 273 S.E.2d 273, 280 (1981) (citing *State v. Powell*, 264 N.C. 73, 74, 140 S.E.2d 705, 706 (1965)).

must be established and recognized, the North Carolina Supreme Court in *State v. Bullard* specifically rejected the *Frye* test as the exclusive standard for admissibility, and recognized reliability as the chief test.⁷⁷

The increasing presence of PTSD as an established and documented technique and North Carolina's "reliability standard" for novel scientific techniques set the stage for an attempt to use evidence of the syndrome in a North Carolina rape trial. In 1986 the inevitable occurred in *State v. Stafford*,⁷⁸ when the State attempted to offer testimony showing the complainant suffered from RTS.⁷⁹ The victim in *Stafford* was a thirteen year-old girl who accused her uncle of raping her.⁸⁰ At trial the pediatrician who examined the victim, when asked if the victim suffered from RTS, testified that "I can't make any conclusions whether or not this means she was raped. I can just say she fulfills some of the criteria for the syndrome that has been defined, and that's all I can say."⁸¹ After the jury convicted the defendant, he appealed, arguing that RTS testimony is inadmissible under rule 702.⁸² The North Carolina Court of Appeals, however, avoided the broad issue of whether expert testimony on RTS (or PTSD) is admissible in rape trials, and determined instead that this particular testimony was inadmissible as hearsay under rule 802.⁸³ The supreme court affirmed on the hearsay grounds and reiterated, "we do not deem it necessary to reach on this record the question whether in a proper case testimony about rape trauma syndrome will be admissible in the courts of this state."⁸⁴ Although the majority on both the court of appeals and the supreme court dodged the RTS issue, dissenters in both cases argued that RTS testimony "as to the symptoms of the syndrome and its existence, is admissible to assist jurors in understanding the evidence and in drawing appropriate conclusions therefrom."⁸⁵ The majority's position, in contrast, remained a mystery.⁸⁶

While North Carolina danced around the PTSD admissibility question,

77. *Bullard*, 312 N.C. at 147, 322 S.E.2d at 380 (holding that, while the general acceptance of a scientific technique does strengthen the likelihood of its admission into evidence, "[p]lainly, our Court does not adhere exclusively to the *Frye* formula").

78. 317 N.C. 568, 346 S.E.2d 463 (1986). At the time *Stafford* came before the court, at least four other states already had faced the PTSD question. See *infra* note 87.

79. *Stafford*, 317 N.C. at 571, 346 S.E.2d at 465.

80. *Id.* at 569, 346 S.E.2d at 464.

81. *Id.* at 571, 346 S.E.2d at 465-66. The doctor subsequently testified on the nature of RTS and its causes. *Id.* at 571-72, 387 S.E.2d at 466.

82. See *State v. Stafford*, 77 N.C. App. 19, 20, 334 S.E.2d 799, 800 (1985), *aff'd*, 317 N.C. 568, 346 S.E.2d 463 (1986).

83. *Id.* at 21, 334 S.E.2d at 800. The court ruled that the statements given to the doctor by the victim constituted inadmissible hearsay and did not fall under the exception to hearsay in North Carolina Rules of Evidence 803(4), which renders admissible statements made for the purpose of medical diagnosis or treatment. *Id.* at 21, 331 S.E.2d at 801. For an argument that the court misapplied the hearsay rules to avoid the PTSD question, see Note, *State v. Stafford: Rape Trauma Syndrome and the Admissibility of Statements Made by Rape Victims*, 64 N.C.L. REV. 1364, 1370 (1986).

84. *Stafford*, 317 N.C. at 575, 346 S.E.2d at 468.

85. *Stafford*, 77 N.C. App. at 26, 334 S.E.2d at 803 (Martin, J., dissenting); *Stafford*, 317 N.C. at 576, 346 S.E.2d at 468 (Martin, J., dissenting). Judge Becton, in contrast, concurred in the court of appeals decision and argued that RTS had not gained sufficient scientific reliability to be admissible. *Stafford*, 77 N.C. App. at 22, 334 S.E.2d at 801 (Becton, J., concurring).

86. Later, in *State v. Goodwin*, the court again avoided the PTSD controversy by holding that

other states squarely confronted it. At the time of *Stafford* four other state supreme courts had ruled directly on the use of PTSD in rape trials.⁸⁷ Of these, three either severely limited or completely excluded the evidence.⁸⁸ These courts, and those that later followed their decisions,⁸⁹ relied on three major arguments to justify their holdings: reliability concerns, lack of helpfulness to the jury, and the danger of prejudice.

First, these courts concluded that PTSD was not sufficient to meet the standards required for the admissibility of scientific expert testimony. In addition to the methodological shortcomings of the early studies, which cast "grave doubt" on the reliability of PTSD (and RTS),⁹⁰ the courts rejecting PTSD testimony contended that PTSD was devised to assist victims in overcoming their trauma rather than to test the accuracy of the victims' contentions.⁹¹ One court concluded that "[r]ape trauma syndrome is not a fact-finding tool, but a therapeutic tool useful in counseling."⁹² In effect, this argument did not dispute the general acceptance of PTSD as a counseling technique, but viewed PTSD as an inappropriate fact-finding method. Furthermore, some courts decided that the difficulty of determining the cause of the syndrome makes evidence of RTS or PTSD inherently unreliable.⁹³ Because other stressors might represent the real cause of PTSD, reliability problems present a serious danger.⁹⁴

The second major reason some courts rejected PTSD and RTS testimony was doubt as to whether its introduction assists the jury enough to merit its admissibility.⁹⁵ Essentially this contention rests on the notion that the jury's

the expert offering the PTSD testimony was not qualified to testify as an expert. 320 N.C. 147, 151, 357 S.E.2d 639, 641 (1987).

87. See *People v. Bledsoe*, 36 Cal. 3d 236, 251, 681 P.2d 291, 301, 203 Cal. Rptr. 450, 460 (1984) (rejecting testimony of a rape counselor that indicated a rape had occurred); *State v. Marks*, 231 Kan. 645, 653, 647 P.2d 1292, 1299 (1982) (admitting RTS testimony to show nonconsent to intercourse); *State v. Saldana*, 324 N.W.2d 227, 232 (Minn. 1982) (rejecting RTS testimony where it alleged the victim was a victim of rape and had not fantasized the event); *State v. Taylor*, 663 S.W.2d 235, 241-42 (Mo. 1984) (en banc) (rejecting RTS testimony offered on the issue of consent).

88. See *Bledsoe*, 36 Cal. 3d at 251, 681 P.2d at 301, 203 Cal. Rptr. at 460; *Saldana*, 324 N.W.2d at 232; *Taylor*, 663 S.W.2d at 241-42.

89. See, e.g., *People v. Taylor*, 75 N.Y.2d 277, 289-90, 552 N.E.2d 131, 136-38, 552 N.Y.S.2d 883, 888-89 (1990); *Commonwealth v. Gallagher*, 519 Pa. 291, 297, 547 A.2d 355, 358-59 (1988); *State v. Black*, 109 Wash. 2d 336, 347-50, 745 P.2d 12, 18-19 (1987).

90. *Black*, 109 Wash. 2d at 345, 745 P.2d at 17; see *supra* text accompanying notes 51-53. For a description of later studies that purported to correct the methodological shortcomings, see *supra* text accompanying notes 54-65.

91. See *Bledsoe*, 36 Cal. 3d at 251, 681 P.2d at 301, 203 Cal. Rptr. at 460. The court stated:

We emphasize that our conclusion in this regard is not intended to suggest that rape trauma syndrome is not generally recognized or used in the general scientific community [from] which it arose, but only that it is not relied on in that community for the purpose for which the prosecution sought to use it in this case, namely, to prove that a rape occurred.

Id.

92. *Saldana*, 324 N.W.2d at 230.

93. See *id.* at 229; *State v. Taylor*, 663 S.W.2d 235, 238 (Mo. 1984) (en banc); *Black*, 109 Wash. 2d at 344, 745 P.2d at 16-17. "Rape trauma syndrome is not the type of scientific test that accurately and reliably determines whether a rape has occurred. The characteristic symptoms may follow any psychologically traumatic event." *Saldana*, 324 N.W.2d at 229. For the contrary view that only a finite number of traumatic events may cause PTSD, see *supra* note 45.

94. See Note, *supra* note 3, at 1699.

95. See, e.g., *Saldana*, 324 N.W.2d at 230; *Taylor*, 663 S.W.2d at 238.

common sense suffices to determine whether a rape occurred.⁹⁶ Due to what they perceived as the questionable validity of the technique and the inherent dangers of psychological testimony,⁹⁷ these courts found a decision based on the facts alone more desirable.⁹⁸

A concern troubling the courts that is related to this second line of reasoning was that admitting PTSD might result in a satellite "battle of the experts." One court argued that "[t]o allow such [PTSD] testimony would inevitably lead to a battle of experts that would invade the jury's province of fact-finding and add confusion rather than clarity."⁹⁹ The courts feared that instead of assisting the jury, PTSD testimony would waste valuable time and only cloud the issue.¹⁰⁰ An even larger problem introduced by a "battle of the experts" is the potential to reawaken the traditional focus on the victim.¹⁰¹ Logically, if the prosecution introduces evidence of PTSD, the defendant should maintain the ability to cross-examine the witness regarding all possible causes of the victim's trauma.¹⁰² Thus, any sexually stressful or otherwise painful event becomes relevant as a possible alternative to rape as the cause of the PTSD. Similarly, the defendant could argue that the court should allow his expert not only to testify regarding PTSD's reliability, but also to examine the victim to determine if, in the expert's estimation, rape or some other trauma caused the victim's symptoms.¹⁰³ This "battle of the experts" could cost the legal system time, confuse the jury, and focus the jury's attention once again upon the victim.

Lurking behind both the reliability concerns and the contention that PTSD will not assist the jury is the third overriding concern of courts that have rejected PTSD and RTS: the fear of undue prejudice.¹⁰⁴ The first courts to address the admissibility of PTSD testimony believed that the danger of undue prejudice far outweighs the evidence's marginal reliability and only slightly as-

96. In refusing to admit RTS testimony the *Taylor* court held that expert testimony " 'should never be admitted unless it is clear that the jurors themselves are not capable, for want of experience or knowledge of the subject, to draw correct conclusions from the facts proved.' " *Taylor*, 663 S.W.2d at 239 (quoting *Sampson v. Missouri Pac. R.R.*, 560 S.W.2d 573, 586 (Mo. 1978) (en banc)).

97. See *supra* notes 51-53 and accompanying text.

98. See *Saldana*, 324 N.W.2d at 230; *Taylor*, 663 N.W.2d at 238-39; *State v. Black*, 109 Wash. 2d 336, 350, 745 P.2d 12, 19 (1987).

99. *Saldana*, 324 N.W.2d at 230.

100. Frazier & Borgida, *Juror Common Understanding and the Admissibility of Rape Trauma Syndrome Evidence in Court*, 12 LAW AND HUMAN BEHAVIOR 101, 116-17 (1988).

101. See Note, *supra* note 3, at 1703-04. The sixth amendment guarantees the defendant the right to probe the victim's psychological or sexual history if she offers PTSD evidence. *Id.* at 1704.

102. See, e.g., *State v. Allewalt*, 308 Md. 89, 109, 517 A.2d 741, 751 (1986); *State v. Liddell*, 211 Mont. 180, 188, 685 P.2d 918, 923 (1984).

103. One court noted:

Cross-examination can include not only cross-examining the expert about PTSD in general, but also cross-examining the expert and the prosecutrix about possible causes of the disorder other than the assault charged in the criminal case. In addition, we can foresee cases where the defendant will seek to counter the State's PTSD evidence with his own expert testimony. That can, in turn, lead to issues concerning compulsory psychiatric examination of the complainant by an expert for the defense.

Allewalt, 308 Md. at 109, 517 A.2d at 751.

104. See, e.g., *State v. Saldana*, 324 N.W.2d 227, 230-31 (Minn. 1982); *State v. Taylor*, 663 S.W.2d 235, 238 (Mo. 1984) (en banc); *State v. Black*, 109 Wash. 2d 336, 348, 745 P.2d 12, 18 (1987).

sists the jury.¹⁰⁵ Although in many later decisions the courts' focus shifted away from reliability concerns,¹⁰⁶ the courts still restricted or prohibited PTSD and RTS testimony on the basis that the prejudicial effects of the testimony outweigh its probative value.¹⁰⁷ At the core of this view is the belief that expert testimony that the victim suffers from PTSD automatically surrounds the victim's story with an impenetrable shield of credibility.¹⁰⁸ This shield protects the victim regardless of whether the expert actually comments on the credibility of the victim,¹⁰⁹ because, as commentators have suggested, the jury attaches "mythic infallibility" to the expert.¹¹⁰ Moreover, at least one court has argued that allowing the expert to indicate, either expressly or implicitly, that the victim suffers from PTSD invades the jury's fact-finding province, thereby relieving it from deciding whether a rape actually occurred.¹¹¹ The majority of the early decisions, therefore, either banned PTSD testimony outright or severely limited its scope.

As more state courts tackled the PTSD issue, however, this early majority position gradually became a minority viewpoint.¹¹² *State v. Marks*¹¹³ was the

105. "[T]he danger of unfair prejudice outweighs any probative value [of RTS testimony.]" *Saldana*, 324 N.W.2d at 230. See, e.g., *People v. Bledsoe*, 36 Cal. 3d 236, 251, 681 P.2d 291, 301, 203 Cal. Rptr. 450, 460 (1984); *Taylor*, 663 S.W.2d at 238.

106. At least one court that completely precluded PTSD or RTS testimony in rape trials, however, focused on the earlier concerns of the technique's inherent unreliability as a fact-finding device. See *Black*, 109 Wash. 2d at 345-46, 745 P.2d at 17.

107. Courts recently deciding this issue have viewed the danger of prejudice as greatly increased when the prosecution uses PTSD or RTS testimony with respect to the issue of consent or to bolster the victim's credibility, as opposed to offering it merely to educate the jury on the trauma associated with rape. See, e.g., *People v. Taylor*, 75 N.Y.2d 277, 293, 552 N.E.2d 131, 139, 552 N.Y.S.2d 883, 891 (1990) (The court rejected the use of RTS testimony introduced to prove a rape occurred because "where it is introduced to prove the crime took place, its helpfulness is outweighed by the possibility of undue prejudice."); *Commonwealth v. Gallagher*, 519 Pa. 291, 297, 547 A.2d 355, 358-59 (1988) (reversing the appellate court's ruling that RTS testimony was admissible when offered to bolster the credibility of the victim's identification of the rapist). These courts recognized PTSD and RTS testimony as valid and reliable scientific tools and did not bar such testimony given to explain the victim's behavior, but ruled that the prejudice associated with the expert's validating the victim's claim that a rape occurred outweighed any probative value of expert testimony on the issue of consent. *Taylor*, 75 N.Y.2d at 293, 552 N.E.2d at 138-39, 552 N.Y.S.2d at 891.

The argument over the prejudicial effects of RTS and PTSD testimony forms the center of the continuing controversy surrounding PTSD testimony. Most modern courts recognize the reliability of the technique, but split when this testimony appears on the issue of consent. One court facing the PTSD question has remarked that

[the admission of PTSD or RTS testimony] is an almost unanimous uniform rule when the expert neither uses the term "rape trauma syndrome" nor offers an opinion on whether the victim had been raped. However, if the expert testifies on these two matters, the courts are approximately evenly split on the admissibility of such evidence.

State v. Gettier, 438 N.W.2d 1, 5-6 (Iowa 1989). Even courts that admit PTSD or RTS testimony on the consent issue remain sensitive to the dangers of prejudice and attempt to neutralize them. See *infra* notes 139-44 and accompanying text.

108. The *Saldana* court contended that the expert testimony "'gave a stamp of scientific legitimacy to the truth of the complaining witness's factual testimony.'" *Saldana*, 324 N.W.2d at 231 (quoting *People v. Izzo*, 90 Mich. App. 727, 730, 282 N.W.2d 10, 11 (1979)).

109. Of course, if the expert comments that the victim is not fantasizing the rape and her story is credible, the danger of undue prejudice becomes far more imminent. *Id.*

110. See Note, *supra* note 3, at 1702.

111. See, e.g., *State v. Black*, 109 Wash. 2d 336, 348, 745 P.2d 12, 18 (1987).

112. For a list of the states admitting PTSD or RTS to show the victim's behavior was similar to other PTSD sufferers, see *Strickland*, 96 N.C. App. at 647, 387 S.E.2d at 65.

first case to uphold the admissibility of RTS or PTSD.¹¹⁴ Contrary to the decisions that questioned the technique's validity,¹¹⁵ the Kansas Supreme Court firmly rooted its decision in the belief that PTSD is a reliable diagnostic technique.¹¹⁶ Other courts admitting PTSD testimony also accepted it as reliable and accurate,¹¹⁷ regardless of its therapeutic origins.¹¹⁸

Courts admitting PTSD evidence responded to the fear that the syndrome can appear as a result of a number of other stressors¹¹⁹ by pointing to the narrow range of stressors that can cause PTSD¹²⁰ and arguing that cross-examination offers the defendant an opportunity to suggest any other possible cause.¹²¹ Of course this contention raises fears that defendants will cross-examine to circumvent rape shield laws that limit inquiry into the victim's past.¹²² In *State v. McQuillen* the Kansas Supreme Court encountered this argument, but dismissed it stating "[s]afeguards are still contained within the statute to protect the victim. A showing of relevancy is still necessary before the complaining witness' prior sexual conduct may be admitted into evidence on behalf of the defendant."¹²³ Thus, since only possible causes of PTSD are relevant to the causation issue, the court can still prevent attempts to examine recklessly every facet of the victim's life.

Additionally, evidence of trauma recently has been allowed in cases involving battered children,¹²⁴ battered women,¹²⁵ and insanity defenses.¹²⁶ Hence,

113. 231 Kan. 645, 647 P.2d 1292 (1982).

114. *Id.* at 653, 647 P.2d at 1299.

115. See *supra* notes 90-94 and accompanying text.

116. *Marks*, 231 Kan. at 654, 647 P.2d at 1299. Most commentators, in light of later studies validating the syndrome, agree with this perspective. One study shows that the scientific community overwhelmingly accepts the reliability of the PTSD concept and supports its use in court. See *Frazier & Borgida, supra* note 100, at 111; see also *supra* text accompanying notes 54-65 (examining studies that have documented the intense emotional reactions consistently suffered by rape victims).

117. See, e.g., *Kruse v. State*, 483 So. 2d 1383, 1387 (Fla. Dist. Ct. App. 1986); *State v. Allewalt*, 308 Md. 89, 107, 517 A.2d 741, 751 (1986) (quoting *Marks*, 231 Kan. at 654, 647 P.2d at 1299); *State v. Liddell*, 211 Mont. 180, 188, 685 P.2d 918, 923 (1984).

118. While some courts have held that the therapeutic origins of PTSD and RTS severely diminished their value as scientifically reliable techniques, recent commentators have argued instead that the bases of PTSD and RTS have no effect on their reliability and accuracy as evidentiary tools. See, e.g., Note, *Expert Testimony, supra* note 52, at 1075-76.

119. See *supra* note 93.

120. Only a narrow range of abnormal traumatic events can cause PTSD. See *supra* note 45. Absent a showing of one of these events, the victim's past is irrelevant.

121. See, e.g., *Allewalt*, 308 Md. at 109, 517 A.2d at 751; *Liddell*, 211 Mont. at 188-89, 685 P.2d at 923.

122. In *State v. McQuillen*, the defense argued that admitting RTS testimony "provides a legal method by which the defendant can evade [the rape shield statute]." 236 Kan. 161, 172, 689 P.2d 822, 830 (1984). For an explanation of the protections offered by rape shield statutes, see *supra* notes 40-43 and accompanying text.

123. *McQuillen*, 236 Kan. at 172, 689 P.2d at 830.

124. In *People v. Bledsoe* the court recognized the admissibility of "battered child syndrome," which indicates that a child with severe injuries has not sustained these injuries by accidental means. 36 Cal. 3d 236, 249, 681 P.2d 291, 299-300, 203 Cal. Rptr. 450, 458-59 (1984). However, the court went on to distinguish battered child syndrome as medical testimony on the cause of a particular physical injury, whereas RTS represented an opinion about the psychological state of the victim. *Id.*

125. A number of courts have recently allowed testimony on battered woman syndrome. See Note, *State v. Norman: Self-Defense Unavailable to Battered Women who Kill Passive Abusers*, 68 N.C.L. REV. 1159, 1165 n.44 (1990) [hereinafter Note, *Norman*]; Note, *A Trend Emerges: A State*

courts have admitted PTSD to maintain consistency with their approach to other psychological traumas.¹²⁷

Courts that allow PTSD testimony view the evidence as potentially very helpful in explaining to the jury the victim's reactions to rape.¹²⁸ Specifically, they have pointed to documented misconceptions jurors often hold regarding rape and its victims.¹²⁹ Thus, PTSD evidence, for example, may explain effectively the victim's reluctance to report the crime or other behavior following the rape.¹³⁰

Some courts also view PTSD as relevant to whether the victim actually consented.¹³¹ To meet the criteria for relevance, evidence must tend to make a fact at issue more or less probable than it would be otherwise.¹³² On this particular subject, Professor Massaro has commented:

That definition makes psychological "bruises" as relevant as physical bruises in a consent-rape trial. Both certainly may result from many causes. Neither the physician who testifies that a woman has physical bruises nor a psychologist who testifies that a woman suffers from RTS can state unequivocally that the condition was caused by a specific incident of non-consensual intercourse, yet the evidence of a victim's physical injuries is deemed clearly relevant in a rape case and admissibility of this is beyond doubt.¹³³

PTSD evidence, then, makes it more probable that a rape did in fact occur,

Survey on the Admissibility of Expert Testimony Concerning the Battered Woman Syndrome, 25 J. FAM. L. 373 (1986). Battered woman syndrome results from the interaction of a series of abusive actions. As a result the woman becomes unable to leave the relationship, notify the authorities, or solicit help from friends or family. See L. WALKER, *THE BATTERED WOMAN* 55-70 (1979) (explaining battered woman syndrome as a cycle of violence consisting of three stages: tension building, acute battering, and contrition); Note, *Norman, supra*, at 1165-68.

126. In comparing RTS to insanity defense cases, one court held that "experts have traditionally been allowed to testify on numerous issues ultimately to be decided by the jury." *Kruse v. State*, 483 So. 2d 1383, 1387 (Fla. Dist. Ct. App. 1986).

127. Some commentators argue, however, that insanity and battered women syndrome expert testimony is completely different from RTS testimony because only the defendant may initiate it. Thus, the defendant's control over the admission of this testimony serves as some protection, whereas no similar protection for the defendant exists in regard to RTS. Courts therefore should not analogize between the admissibility of RTS or PTSD and other admissible syndromes. See Note, *supra* note 3, at 1693-95.

128. Because of its helpfulness to the jury the testimony fulfills the evidentiary requirements for admissibility. N.C.R. EVID. 702.

129. See, e.g., *People v. Bledsoe*, 36 Cal. 3d 236, 247-48, 681 P.2d 291, 298, 203 Cal. Rptr. 450, 457 (1984) (reserving the possibility that RTS could be used to educate the jurors on the subject of rape and its victims, even though RTS was not allowed on the issue of consent). In a study of average citizens' understanding of rape based upon a fourteen-question quiz, "the scores showed that, in general, most people knew very little about the facts regarding rape. The average score of the respondents on the fourteen-item test was less than four items correct." H. FEILD & L. BIENEN, *JURORS AND RAPE: A STUDY IN PSYCHOLOGY AND LAW* 89 (1980).

130. See *Bledsoe*, 36 Cal. 3d at 251, 681 P.2d at 301, 203 Cal. Rptr. at 460; *People v. Taylor*, 75 N.Y.2d 277, 286, 552 N.E.2d 131, 138-39, 552 N.Y.S.2d 883, 891 (1990); *Scadden v. State*, 732 P.2d 1036, 1039 (Wyo. 1987).

131. See, e.g., *State v. McQuillen*, 236 Kan. 161, 172, 689 P.2d 822, 829 (1984); *State v. Al-lewalt*, 308 Md. 89, 109, 517 A.2d 741, 751 (1986); *State v. Liddell*, 211 Mont. 180, 188, 685 P.2d 918, 923 (1984).

132. N.C.R. EVID. 401.

133. Massaro, *supra* note 29, at 440-41 (footnote omitted).

according to courts that accept its reliability as a diagnostic tool.¹³⁴

Despite its use as a diagnostic tool, most courts have decided to admit PTSD testimony because "although the admission of evidence of this nature has a prejudicial impact on the defendant's claim of innocence, the probative value of this testimony *clearly outweighs* the prejudicial value."¹³⁵ Central to this view is the notion that juries do *not* rely on experts as "mythically infallible," but weigh expert testimony just as they would any other evidence.¹³⁶ Moreover, these courts have reasoned that expert testimony on RTS or PTSD does not invade the province of the jury, but only assists the jury in deciding the ultimate issues.¹³⁷ Cross-examination allows the defense to point out any inherent deficiencies in the testimony and protect against any danger of undue prejudice.¹³⁸

In recognition of the inherent prejudicial dangers of per se admissibility of PTSD testimony,¹³⁹ few courts that admit PTSD testimony do so unconditionally.¹⁴⁰ Some courts require specific jury instructions stressing to the jury that the expert testimony is not a legal conclusion and carries only the same weight as other evidence.¹⁴¹ Many courts allow the expert to explain the symptoms of PTSD and to state that the victim exhibits similar symptoms, but do not allow the expert to address the credibility of the victim.¹⁴² Still other courts allow PTSD testimony only to explain the victim's post-rape behavior, which the jury might find nonsensical or unusual (such as failing to report the rape immediately).¹⁴³ Finally, a few courts continue to forbid completely PTSD testimony in rape trials.¹⁴⁴

As the battle over PTSD and RTS admissibility continued in other state courts, North Carolina courts increasingly began to indicate that they would

134. See *Allewalt*, 308 Md. at 109, 517 A.2d at 751.

135. *State v. Whitman*, 16 Ohio App. 3d 246, 247, 475 N.E.2d 486, 488 (1984) (allowing RTS evidence to explain a child victim's behavior). For a holding that the probative value outweighs possible prejudice when RTS evidence alleges lack of consent, see *State v. Huey*, 145 Ariz. 59, 63, 699 P.2d 1290, 1294 (1985).

136. *Allewalt*, 308 Md. at 102, 517 A.2d at 748.

137. See, e.g., *Huey*, 145 Ariz. at 63, 699 P.2d at 1294; *State v. Marks*, 231 Kan. 645, 654, 647 P.2d 1292, 1299 (1982).

138. See *Allewalt*, 308 Md. at 109, 517 A.2d at 751; *State v. Liddell*, 211 Mont. 180, 188, 685 P.2d 918, 923 (1984).

139. See *Huey*, 145 Ariz. at 63, 699 P.2d at 1294.

140. The list of states that allow per se admissibility is fairly short. See, e.g., *id.*; *People v. Douglas*, 183 Ill. App. 3d 241, 256-57, 538 N.E.2d 1335, 1344 (1989), *cert. denied*, 127 Ill. 2d 625, 546 N.E.2d 1141 (1989); *Marks*, 231 Kan. at 654, 647 P.2d at 1299; *Liddell*, 211 Mont. at 188, 685 P.2d at 923.

141. "[B]y proper jury instructions . . . the trial court can prevent any impression that the psychiatric opinion is like a chemical reaction." *Allewalt*, 308 Md. at 109, 517 A.2d at 751.

142. See, e.g., *Kruse v. State*, 483 So. 2d 1383, 1387-88 (Fla. Dist. Ct. App. 1986); *Simmons v. State*, 504 N.E.2d 575, 579 (Ind. 1987); *Allewalt*, 308 Md. at 108-09, 517 A.2d at 751; *State v. McCoy*, 366 S.E.2d 731, 736-37 (W. Va. 1988).

143. See *People v. Bledsoe*, 36 Cal. 3d 236, 251, 681 P.2d 291, 301, 203 Cal. Rptr. 450, 460 (1984) (refusing to allow RTS testimony to disprove consent, but expressly reserving the possibility that RTS is admissible to dispel the juror's misconceptions on rape victims); *People v. Taylor*, 75 N.Y.2d 277, 293, 552 N.E.2d 131, 138-39, 552 N.Y.S.2d 883, 891 (1990) (same); *Scadden v. State*, 732 P.2d 1036, 1039 (Wyo. 1987) (expert testimony allowed to explain the victim's delay in reporting the rape).

144. See, e.g., *State v. Saldana*, 324 N.W.2d 227, 232 (Minn. 1982); *State v. Taylor*, 663 S.W.2d 235, 241 (Mo. 1984) (en banc); *State v. Black*, 109 Wash. 2d 336, 349, 745 P.2d 12, 19 (1987).

allow such testimony. *State v. Clemmons*, decided in 1987, involved a rape occurring in the victim's home.¹⁴⁵ On appeal of his conviction, the defendant contended that the trial court improperly admitted evidence of his prior misconduct towards another female.¹⁴⁶ Although recognizing the trial court's error in admitting this evidence, the supreme court refused to overturn the decision because of the overwhelming evidence against the defendant.¹⁴⁷ The court relied "particularly [on] the medical evidence of the victim's severe post-traumatic stress disorder for a lengthy period immediately following the incident."¹⁴⁸ While the PTSD evidence was not at issue in *Clemmons*, the court's reliance on it indicated its willingness to accept such testimony, at least in some capacity, as evidence in rape trials.

*State v. Teeter*¹⁴⁹ followed on the heels of the *Clemmons* ruling and involved expert testimony that the victim suffered from symptoms consistent with those found in other rape victims.¹⁵⁰ Although the expert who testified on behalf of the State never expressly used the term PTSD or RTS, the symptoms he described closely mirrored PTSD evidence.¹⁵¹ The court, rejecting the defendant's characterization of the testimony as going to the credibility of the victim's story,¹⁵² allowed the testimony. The court stated that "Dr. Short never testified that the sexual acts related by [the victim] were committed by any particular person, nor did he purport to express an opinion as to defendant's guilt or innocence."¹⁵³ Again this decision, and others like it,¹⁵⁴ indicated the court's willingness to admit testimony similar to PTSD evidence.¹⁵⁵

In *Strickland* the North Carolina Court of Appeals took the decisions in *Clemmons* and *Teeter* to a new level by expressly allowing an expert to testify that the victim suffered from PTSD.¹⁵⁶ The court firmly rooted its decision in the widespread acceptance of PTSD.¹⁵⁷ Not only is PTSD a reliable scientific

145. *State v. Clemmons*, 319 N.C. 192, 194, 353 S.E.2d 209, 210-11 (1987).

146. In an attempt to impeach the defendant's credibility the prosecution cross-examined the defendant about a prior incident. Allegedly the defendant had entered a woman's apartment, made advances towards her, thrown her on a bed, got on top of her, and stopped only when the woman managed to kick a ringing phone off the hook. *Id.* at 196-97, 353 S.E.2d at 212.

147. *Id.* at 199-200, 353 S.E.2d at 213-14.

148. *Id.* at 199, 353 S.E.2d at 213.

149. 85 N.C. App. 624, 355 S.E.2d 804, *disc. rev. denied*, 320 N.C. 175, 358 S.E.2d 67 (1987).

150. The expert, Dr. Short, testified that the victim's behavior following the rape resembled that of other rape victims in that she suffered nightmares, appeared nervous, and expressed fear of her alleged abuser. *Id.* at 630, 355 S.E.2d at 808.

151. *Id.*

152. *Id.*

153. *Id.*

154. See, e.g., *State v. Kennedy*, 320 N.C. 20, 31-32, 357 S.E.2d 359, 366-67 (1987) (expert testified that the victim's symptoms were characteristic of those exhibited by children who have experienced sexual abuse); *State v. Bailey*, 89 N.C. App. 212, 220-22, 365 S.E.2d 651, 656-57 (1988) (expert testified about the victim's behavior during his examination).

155. In its brief to the court of appeals, the State contended, with regard to *Teeter* and other similar cases, that "[b]oth this court and the North Carolina Supreme Court have held that an expert witness may testify to general symptoms and characteristics of victims of sexual abuse and to state an opinion that the behavior exhibited by the victim is consistent with . . . other victims of sexual abuse." Brief for the State at 20, *Strickland* (No. 8910SC411).

156. *Strickland*, 96 N.C. App. at 648-49, 387 S.E.2d at 66.

157. *Id.* at 646-47, 387 S.E.2d at 65.

method, but it also is generally accepted among the medical community,¹⁵⁸ among legal commentators,¹⁵⁹ and by a majority of the state courts.¹⁶⁰ This broader acceptance of PTSD testimony prevents the inconsistency that would result if courts allowed experts to testify that a victim exhibited symptoms similar to those of other rape victims, but refused to allow the expert to mention the medical symptoms suffered by rape victims. The mere label "PTSD" or "RTS" should not invoke different admissibility rules.¹⁶¹ Further, admission of PTSD testimony serves to educate the jury concerning the popular misconceptions surrounding rape and its victims.¹⁶²

Using PTSD testimony to educate the jury is, however, a far cry from using it to disprove consent, as the *Strickland* court allowed.¹⁶³ The level of prejudice rises dramatically when the expert discusses whether the sexual relations were voluntary or forced, as opposed to merely explaining the victim's behavior or commenting on the general severity of rape.¹⁶⁴ Surprisingly, the court in *Strickland* made no mention of the possible prejudices surrounding use of PTSD testimony for this purpose. Virtually every other court ruling on PTSD or RTS has considered the impact the testimony might have on the defendant and has provided some means to limit these dangers, or at least to explain how the relevance of PTSD outweighs the prejudicial effects.¹⁶⁵ Judge Lewis' brief opinion in *Strickland*, in contrast, does not raise, let alone address, these concerns.¹⁶⁶ The court simply cited a number of cases as supporting the admissibility of

158. Frazier & Borgida, *supra* note 100, at 111.

159. See Massaro, *supra* note 29, at 460-70; Note, *Expert Testimony*, *supra* note 52, at 1074-86.

160. See *Strickland*, 96 N.C. App. at 647, 387 S.E.2d at 65 (listing the state decisions allowing PTSD and RTS testimony).

161. See Brief for the State at 20.

162. See *supra* text accompanying notes 29-34. People v. Bledsoe, 36 Cal. 3d 236, 681 P.2d 291, 203 Cal. Rptr. 450 (1984). The Bledsoe court stated, "We hasten to add that nothing in this opinion is intended to imply that evidence of the emotional and psychological trauma that a complaining witness suffers after an alleged rape is inadmissible in a rape prosecution." *Id.* at 251, 681 P.2d at 301, 203 Cal. Rptr. at 460. The *Strickland* court recognized Judge Martin's dissent in *State v. Stafford*, which pointed out PTSD testimony's ability to assist the jury in understanding the plight of rape victims:

"There is recognized scientific authority for the medical conclusion that there exists a complex and unique number of physical and emotional symptoms exhibited by victims of rape, which are similar, but not identical, to other post-traumatic stress disorder symptoms. An understanding of those symptoms, the unique reactions of victims of rape, is not within the common knowledge or experience of most persons called upon to serve as jurors. Therefore, expert testimony as to the symptoms of the syndrome and its existence, is admissible to assist jurors in understanding the evidence and in drawing appropriate conclusions therefrom."

Strickland, 96 N.C. App. at 648, 387 S.E.2d at 66 (Martin, J., dissenting) (quoting *State v. Stafford*, 77 N.C. App. 19, 26, 334 S.E.2d 799, 803 (1985), *aff'd* 317 N.C. 568, 346 S.E.2d 463 (1986)).

163. See *Strickland*, 96 N.C. App. at 645-46, 387 S.E.2d at 65.

164. When the expert offers his opinion that the victim did not consent or that a rape occurred, the danger of the jury blindly accepting the victim's story is much greater than when the expert merely talks about the nature of PTSD to educate the jury. See *supra* notes 104-11 and accompanying text. In fact, some courts draw a sharp line between PTSD testimony on the issue of consent and testimony to dispel the jury's misconceptions or explain the victim's behavior. See *Bledsoe*, 36 Cal. 3d at 251, 681 P.2d at 301, 203 Cal. Rptr. at 460; People v. Taylor, 75 N.Y.2d 277, 293, 552 N.E.2d 131, 138-39, 552 N.Y.S.2d 883, 890-91 (1990); Scadden v. State, 732 P.2d 1036, 1047 (Wyo. 1987).

165. See *supra* notes 139-44 and accompanying text.

166. See *Strickland*, 96 N.C. App. at 644-49, 387 S.E.2d at 63-66.

PTSD or RTS,¹⁶⁷ failing to acknowledge that these decisions comprised a wide variety of specific restrictions on the testimony itself and the context in which the testimony was used.¹⁶⁸ The *Strickland* opinion left open the question whether PTSD and RTS testimony is per se admissible or whether its admissibility is limited to situations similar to the facts of *Strickland*. In *Strickland* the State's expert testimony that the victim displayed symptoms consistent with PTSD was offered to refute the defendant's contention that the victim consented.¹⁶⁹ The expert neither explicitly addressed the victim's credibility nor stated that she thought a rape had occurred;¹⁷⁰ therefore, the court did not discuss whether an expert could testify directly on the credibility issue. Language in the opinion, however, suggests that the court will confine expert testimony to a more general realm.¹⁷¹ The court, when citing the many state decisions allowing PTSD, stated "[m]ost jurisdictions allow such testimony on PTSD, or on rape trauma syndrome, or expert testimony regarding reactions or behavior consistent with other victims of sexual assault."¹⁷² This comment indicates that the court recognized the problems associated with PTSD testimony and might admit it only to show that the victim displayed symptoms consistent with PTSD. This view represents the best approach to PTSD, because it lessens the possibility that the jury might abdicate its fact-finding duties by simply accepting the expert's opinion that a rape did occur, and greatly reduces the likelihood of undue prejudice resulting from PTSD testimony. The court's opinion in *Strickland* never expressly discussed any boundaries to PTSD testimony, and its approach to the prejudice issue remains a mystery with few clues.

An equally disturbing aspect of the *Strickland* opinion is the court's failure to consider the scope of a defendant's cross-examination after the introduction of PTSD testimony.¹⁷³ The court never discussed whether a defendant will be able to pry indiscriminately into the victim's past for an alternative cause of the PTSD. Other courts have maintained that the defendant cannot escape rape shield laws once PTSD testimony enters the case.¹⁷⁴ By placing boundaries on just what the defendant can explore as he tries to discredit the notion that rape, rather than some other trauma, caused the victim's PTSD, these courts attempt to prevent a return to a trial that focuses on the victim. Of course the court must honor the defendant's constitutional right to confront his accusers,¹⁷⁵ so

167. See *id.* at 647, 387 S.E.2d at 65.

168. See *supra* notes 139-44 and accompanying text. In fact, one of the decisions cited by the *Strickland* court was later overturned by the Pennsylvania Supreme Court. *Commonwealth v. Gallagher*, 519 Pa. 291, 297, 547 A.2d 355, 359 (1988). In its haste the *Strickland* court mistakenly cited the appellate decision that admitted RTS testimony. *Strickland*, 96 N.C. App. at 647, 387 S.E.2d at 65 (citing *Commonwealth v. Gallagher*, 353 Pa. Super. 426, 510 A.2d 735 (1986), *rev'd*, 519 Pa. 291, 547 A.2d 355 (1988)).

169. *Strickland*, 96 N.C. App. at 645-46, 387 S.E.2d at 64-65.

170. *Id.* at 646, 387 S.E.2d at 65.

171. *Id.*

172. *Id.*

173. *Id.* at 644-49, 387 S.E.2d at 63-66.

174. See *State v. McQuillen*, 236 Kan. 161, 172, 689 P.2d 822, 830 (1984); *supra* text accompanying notes 122-23.

175. U.S. CONST. amend. VI.

the right to cross-examine cannot be overly circumscribed. The court, however, can limit the defendant's inquiry to *relevant* events without infringing on the defendant's constitutional rights.¹⁷⁶

Closely related to the danger of a defendant using cross examination to uncover skeletons from the victim's past is the question of mandatory psychological examination of the victim.¹⁷⁷ This question could arise in two situations.¹⁷⁸ First, the defendant might seek psychological examination evidence in response to the prosecution's PTSD or RTS evidence. Should the defendant have access to the medical records of the expert, or can the defendant demand to have his own expert examine the victim?¹⁷⁹ At present, North Carolina courts do not allow mandatory psychological exams,¹⁸⁰ but commentators have argued that if PTSD is available to disprove consent, the accused should have an equal opportunity to have his own expert examine the victim.¹⁸¹ More dangerous is the second possibility, that the defendant might someday attempt to use PTSD to prove consent. For example, the defendant might demand an expert examination of the victim, and if she exhibits no PTSD symptoms he might then contend that no rape occurred. Thus, PTSD could become a subterranean essential element of rape, putting the onus on the victim to prove she suffered from PTSD or face the contention that the absence of PTSD symptoms proves her consent to sexual intercourse. To avoid these potential problems, courts must refuse the defendant access to medical records unless the victim plans to introduce PTSD evidence and must prevent defendants from initiating discussion of PTSD.¹⁸² The victim would maintain control over the PTSD issue, thereby preventing PTSD from becoming a weapon of the defense.¹⁸³ The North Carolina Court of

176. One court has taken this approach already as a means to limit the defendant's inquiry into the victim's past. *McQuillen*, 236 Kan. at 172, 689 P.2d at 830 (1984).

177. *See State v. Allewalt*, 308 Md. 89, 109-10, 517 A.2d 741, 751 (1986).

178. The *Allewalt* court foresaw both possibilities:

When a trial judge admits PTSD evidence . . . the ruling necessarily carries certain baggage with it. . . . [W]e can foresee cases where the defendant will seek to counter the State's PTSD evidence with his own expert testimony. That can, in turn, lead to issues concerning compulsory psychiatric examination of the complainant by an expert for the defense. Lurking in the background is the nice question of whether the absence of PTSD is provable by the accused in defense of a rape charge, as tending to prove that there was consent.

Id.

179. This issue recently came before the Ohio Court of Appeals when a defendant contended that his counsel provided him an inadequate defense by failing to produce an expert witness to testify that the victim did not suffer from RTS and, therefore, had consented to the defendant's sexual advances. *State v. Rose*, No. 57573 (Ohio Ct. App. Oct. 4, 1990) (LEXIS, States library, Ohio file). Although the court refused to grant the defendant's motion for retrial because the decision not to call an expert was a non-reviewable tactical decision, it recognized the possibility of defendants using RTS testimony to prove consent. *Id.*

180. *State v. Clontz*, 305 N.C. 116, 123-24, 286 S.E.2d 793, 797 (1982) (holding that a trial judge does not have the discretionary power to compel an unwilling witness to submit to a psychiatric examination).

181. *See Massaro, supra* note 29, at 453-60; Comment, *Expert Testimony on Rape Trauma Syndrome: Admissibility and Effective Use in Criminal Rape Prosecution*, 33 AM. U.L. REV. 417, 459-60 (1984).

182. Massaro, *supra* note 29, at 453-60.

183. This would create a situation similar to the use of the battered woman syndrome and the insanity defense, where the defendant controls the evidence's admissibility. *See Note, supra* note 3, at 1694-95.

Appeals failed to consider these questions in its *Strickland* ruling, but assuredly it will be forced to face them in future cases.

The failure of the court of appeals to address or provide for the major problems associated with PTSD testimony, and the subsequent supreme court decision to deny review, leaves uncertain the exact limits surrounding PTSD testimony.¹⁸⁴ Boundaries must be placed on the use of PTSD, either now or in subsequent cases. Rather than articulating these limits on a case-by-case basis, efficiency mandates forming a blueprint for handling future difficulties. Virtually every court faced with the PTSD dilemma has recognized its inherent dangers and established concrete limitations or, at least, discussed why the benefits outweigh the risks of this form of testimony.¹⁸⁵ *Strickland* offered the perfect opportunity to form an effective PTSD strategy, but neither court seized it. Thus, the difficult problems surrounding this issue will continue to appear until the court establishes a consistent approach to PTSD and RTS testimony.¹⁸⁶ Although the court made the correct preliminary decision to admit PTSD testimony, it failed to lay the proper groundwork for consistent and fair admission of the evidence by the trial courts.

LOUIS A. TROSCH, JR.

184. *State v. Strickland*, 326 N.C. 486, 392 S.E.2d 100 (1990).

185. See *supra* text accompanying notes 139-44.

186. *State v. Hall* presents the perfect chance for the court to establish this consistent approach, but the supreme court must go further than the lower courts, which declared only that PTSD is admissible to prove consent. 98 N.C. App. 1, 8, 390 S.E.2d 169, 172-73, *disc. rev. allowed*, 327 N.C. 486, 397 S.E.2d 228 (1990). For a full discussion of *Hall*, see *supra* note 11.

***Duke University v. St. Paul Fire & Marine Insurance Co.:* The Court of Appeals Brings Ambiguity to the Interpretation of "Professional Services"**

The extraordinary development of insurance, and its necessary adaptation to the varying and complicated business relations of a progressive age tax the utmost ability of the courts. . . . While we should protect the companies against all unjust claims and enforce all reasonable regulations necessary for their protection, we must not forget that the primary object of all insurance *is to insure*. . . . We can not permit insurance companies by unreasonable stipulations to evade the payment of such indemnity when justly due, and thus defeat the very object of their existence.¹

Over ninety years ago, the North Carolina Supreme Court, revealing more than a hint of consternation about the increasing complexity of insurance contracts, joined other American jurisdictions in concluding that ambiguous clauses in contracts of insurance should be interpreted in favor of the insured. This canon of construction, which since has proved inherently flexible in both application and outcome,² has become a favorite of modern jurists dealing in the complex and technical language of insurance contracts.³

Most recently, the North Carolina Court of Appeals applied the doctrine of strict construction in *Duke University v. St. Paul Fire & Marine Insurance Co.*⁴ to give narrow interpretation to a "professional services" exclusion in the general liability insurance policy of a North Carolina hospital.⁵ This ruling places North Carolina among a growing number of states giving narrow construction to "professional services" exclusions.⁶ This Note analyzes the court's holding in *Duke University* and concludes that while the strict construction of professional services exclusions is consistent with North Carolina courts' long term deference to policyholders, the court's rationale creates a confusing area of overlap between general liability insurance policies and professional service insurance policies. The Note suggests that the court could have minimized this overlap by

1. *Grabbs v. Farmers' Mut. Fire Ins. Ass'n*, 125 N.C. 389, 399, 34 S.E. 503, 506 (1899).

2. See *infra* notes 41-61 and accompanying text.

3. 2 M. RHODES, *COUCH CYCLOPEDIA OF INSURANCE LAW* § 15:74, at 334, 341 (rev. 2d ed. 1984). "The words, 'the contract is to be construed against the insurer' comprise the most familiar expression in the reports of insurance cases." *Id.* at 334. For a list of cases applying the rule that ambiguous language is to be construed against the insurer, see *id.* at 341 n.10.

4. 96 N.C. App. 635, 386 S.E.2d 762, *disc. rev. denied*, 326 N.C. 595, 393 S.E.2d 876 (1990).

5. *Id.* at 641, 386 S.E.2d at 766.

6. See, e.g., *Gulf Ins. Co. v. Gold Cross Ambulance Serv. Co.*, 327 F. Supp. 149, 154-55 (W.D. Okla. 1971); *Keepes v. Doctors Convalescent Center, Inc.*, 89 Ill. App. 2d 36, 40, 231 N.E.2d 274, 276 (1967); *American Casualty Co. v. Hartford Ins. Co.*, 479 So. 2d 577, 579 (La. App. 1985); *D'Antoni v. Sara Mayo Hosp.*, 144 So. 2d 643, 646-47 (La. App. 1962); *Lady Beautiful, Inc. v. Zurich Ins. Co.*, 16 Ohio Misc. 169, 172, 240 N.E.2d 894, 896 (1968). But see *Ocean Accident & Guarantee Corp. v. Herzberg's Inc.*, 100 F.2d 171, 173 (8th Cir. 1938), *cert. denied*, 306 U.S. 645 (1939); *Antles v. Aetna Casualty & Sur. Co.*, 221 Cal. App. 2d 438, 443, 34 Cal. Rptr. 508, 511 (1963); *Brockbank v. Travelers Ins. Co.*, 12 A.D.2d 691, 691, 207 N.Y.S.2d 723, 724 (1960), *appeal denied*, 9 N.Y.2d 609, 172 N.E.2d 293, 210 N.Y.S.2d 1025 (1961); *Harris v. Fireman's Fund Indem. Co.*, 42 Wash. 2d 655, 660-61, 257 P.2d 221, 227 (1953).

considering the intent of the contracting parties and the economic considerations driving insurance contracts. In particular, the Note submits that a proper analysis would have considered that insurance policies are designed to classify insureds according to commonality of risk, and that these classifications, when successfully applied, serve the important public policy goal of placing the burden of insuring a particular type of risk on those persons most able to protect against it.

At issue in *Duke University* was a general liability insurance policy insuring Duke University and Duke University Medical Center against claims of injury arising from accidents on their premises.⁷ The policy specifically excluded "claims arising out of the providing or failure to provide professional services" in hospital operations.⁸ In June of 1986, an elderly patient at the Medical Center's outpatient dialysis center was injured when attendants failed to stabilize her dialysis chair as she attempted to rise after treatment.⁹ The patient later died from injuries sustained in the fall and a wrongful death action ensued.¹⁰ St. Paul Fire and Marine Insurance Company, the underwriter of the university's general liability policy, refused to defend the wrongful death action on grounds that the claim was within the policy's professional services exclusion.¹¹ Following settlement of the parent action, the university brought suit against the insurer to recover its costs.¹² The trial court ordered summary judgment for the university and the insurer appealed.¹³ The court of appeals unanimously affirmed.¹⁴

The interpretation of professional services exclusions in insurance policies was an issue of first impression for North Carolina appellate courts.¹⁵ The court of appeals thus began its analysis of the exclusion clause by briefly reviewing North Carolina's well-settled rules of construction for contracts of insurance. Specifically, the court noted that "[p]rovisions which exclude liability coverage are not favored . . . and any ambiguities must be construed against the insurer and in favor of the insured."¹⁶

After quickly concluding that the claim at issue could not be excluded on causal grounds alone,¹⁷ the court turned to the task of construing the term "pro-

7. *Duke University*, 96 N.C. App. at 636, 386 S.E.2d at 763; see Record at 21, *Duke University* (No. 8914SC33).

8. *Duke University*, 96 N.C. App. at 637, 386 S.E.2d at 764; Record at 29.

9. *Duke University*, 96 N.C. App. at 638, 386 S.E.2d at 764.

10. *Id.* at 636, 638, 386 S.E.2d at 763, 764; see Record at 4.

11. *Duke University*, 96 N.C. App. at 636, 386 S.E.2d at 763.

12. *Id.*

13. *Id.*

14. *Id.* at 641, 386 S.E.2d at 766.

15. *Id.* at 638, 386 S.E.2d at 764.

16. *Id.*

17. See *id.* at 638-39, 386 S.E.2d at 764. The exclusion clause provided that claims "arising out of" professional services should be excluded. *Id.* at 638, 386 S.E.2d at 764. The court noted that the "arising out of language" must be construed narrowly "to require that the excluded cause be the sole proximate cause of the injury." *Id.* at 638-39, 386 S.E.2d at 764 (citing *State Capital Ins. Co. v. Nationwide Mut. Ins. Co.*, 318 N.C. 534, 547, 350 S.E.2d 66, 74 (1986)). Thus the court held: "[C]overage [in this case] is excluded only if any negligence with respect to assisting decedent out of

essional services." Two principles guided its inquiry. First, the court noted that "a 'professional service' generally is defined as one arising out of a vocation or occupation involving specialized knowledge or skills, and the skills are mental as opposed to manual."¹⁸ Second, the court observed, acts or omissions should be classified as professional services on the basis of the nature of the acts or omissions themselves, rather than the "position of the person responsible for the act or omission."¹⁹ Applying these postulates and the rule that courts should construe ambiguous policy exclusions in favor of the insured, the court concluded that the actions at issue in *Duke University* could not be classified as professional services for policy exclusion purposes.²⁰ Professional service policy exclusions, the court held, apply "only [to] those services for which professional training is a prerequisite to performance."²¹ Stabilizing the dialysis chair and assisting the injured party required no special skill or training and thus could not be viewed as falling within the professional services exclusion.²²

The court was not persuaded by authorities from other jurisdictions that had given professional service exclusions broader sweep.²³ Jurisdictions that had so ruled, the court noted, "did not employ the strict rule of construction against the insurer that we must follow in this case."²⁴ Further, the court held that while its interpretation of the exclusion language implied that the same claim might be insured under both professional malpractice policies and general liability policies with professional service exclusions, this potential for overlapping coverage would not affect its decision.²⁵

The notion that insurance policy exclusions should be disfavored and narrowly construed against the insurer is but one element in a group of constructional principles North Carolina courts apply when interpreting contracts of

the chair was a providing [sic] or failure to provide professional services. In order to resolve this issue, we must construe the term 'professional services.' " *Id.* at 639, 386 S.E.2d at 764.

18. *Id.* at 639, 386 S.E.2d at 765 (citing *Smith v. Keator*, 21 N.C. App. 102, 105-06, 203 S.E.2d 411, 415, *aff'd*, 285 N.C. 530, 206 S.E.2d 203, *appeal dismissed*, 419 U.S. 1043 (1974)); *see infra* notes 74-80 and accompanying text.

19. *Duke University*, 96 N.C. App. at 639, 386 S.E.2d at 765 (citing *Gulf Ins. Co. v. Gold Cross Ambulance Serv. Co.*, 327 F. Supp. 149, 152 (W.D. Okla. 1971)); *see infra* notes 88-89 and accompanying text.

20. *Duke University*, 96 N.C. App. at 641, 386 S.E.2d at 766.

21. *Id.*

22. *Id.* Further, the court concluded that:

"[a]lthough the dialysis chair was a specialized piece of equipment, the injury was not related to any special function of the chair but merely resulted from the presence of casters on the chair which enable it to be easily moved. . . . [N]o special training is required for a person to know that a chair with casters may move when someone attempts to rise from it."

Id.

23. *See supra* note 6.

24. *Duke University*, 96 N.C. App. at 640, 386 S.E.2d at 765. Presumably the court was referring to the rule that "provisions which exclude liability coverage are not favored," *id.* at 638, 386 S.E.2d at 764, since virtually all jurisdictions employ the general rule of construction against the insurer. *See supra* note 3. Even the rule disfavoring exclusions may not distinguish adequately many of the holdings in jurisdictions that give broader interpretation to the exclusion language. *See infra* note 37.

25. *Duke University*, 96 N.C. App. at 640, 386 S.E.2d at 765. The court did not explain this conclusion.

insurance. The North Carolina Supreme Court, in *Maddox v. Colonial Life & Accident Insurance Co.*, summed up these constructional rules:²⁶

In interpreting the relevant provisions of [insurance policies], we are guided by the general rule that in the construction of insurance contracts, any ambiguity in the meaning of a particular provision will be resolved in favor of the insured and against the insurance company. Exclusions from and exceptions to undertakings by the company are not favored, and are to be strictly construed to provide the coverage which would otherwise be afforded by the policy. The various clauses are to be harmoniously construed, if possible, and every provision given effect. An ambiguity exists where, in the opinion of the court, the language of the policy is fairly and reasonably susceptible to either of the constructions asserted by the parties.²⁷

Three principles underlie this aggregated rule of construction. First, in construing ambiguous policy provisions, courts must favor the insured.²⁸ Second, absent ambiguity, courts must construe an insurance policy in accordance with its express terms.²⁹ Finally, ambiguous terms must be construed in harmony with the express provisions of the policy.³⁰

Although frequently applied mechanically by modern courts,³¹ there are important theoretical underpinnings for the rule that insurance policies should be construed in favor of the insured. As the quotation that opens this Note suggests, North Carolina courts struggling with the rapid development of new forms of insurance during the early twentieth century sought to protect relatively unschooled laypersons ill-equipped to analyze and interpret their policies.³² Although this rule is similar to the admonition *contra proferentem*,³³ it has independent policy justifications. As one court has pointed out, "[insurance policies] are unipartite. . . . In general, the insured never sees the policy until after he contracts and pays his premium, and he then most frequently receives it from a distance, when it is too late for him to obtain explanations or

26. 303 N.C. 648, 280 S.E.2d 907 (1981) (six of seven justices concurring in the result but not in the opinion), quoted with approval in *Nationwide Mut. Fire Ins. Co. v. Allen*, 68 N.C. App. 184, 188, 314 S.E.2d 552, 554-55, disc. rev. denied, 311 N.C. 761, 321 S.E.2d 142 (1984).

27. *Maddox*, 303 N.C. at 650, 280 S.E.2d at 908 (citations omitted).

28. See *infra* notes 31-40 and accompanying text.

29. See *infra* notes 41-61 and accompanying text.

30. See *infra* notes 62-73 and accompanying text.

31. See, e.g., *State Capital Ins. Co. v. Nationwide Mut. Ins. Co.* 318 N.C. 534, 541, 547, 350 S.E.2d 66, 70, 73-74 (1986); *Waste Management, Inc. v. Peerless Ins. Co.*, 315 N.C. 688, 693-94, 340 S.E.2d 374, 378-79 (1986); *Maddox v. Colonial Life & Accident Ins. Co.*, 303 N.C. 648, 653-54, 280 S.E.2d 907, 910 (1981).

32. See, e.g., *Roberts v. American Alliance Ins. Co.*, 212 N.C. 1, 4, 192 S.E. 873, 875 (1937); *East Carolina Ry. v. Maryland Casualty Co.*, 145 N.C. 114, 117, 58 S.E. 906, 907 (1907) (quoting *Bray v. Virginia Fire & Marine Ins. Co.*, 139 N.C. 390, 393, 51 S.E. 922, 923 (1905)).

33. See 2 M. RHODES, *supra* note 3, § 15:74, at 334, 341. The constructional principle of *contra proferentem* suggests that "if language supplied by one party is reasonably susceptible to two interpretations, one of which favors each party, the one that is less favorable to the party that supplied the language is preferred." E. FARNSWORTH, *CONTRACTS* § 7.11, at 518 (2d ed. 1990). The rule frequently is applied in the construction of standard contracts and "often operates against a party that is at a distinct advantage in bargaining." *Id.*

modifications."³⁴

In contrast to the clearly articulated considerations supporting the general rule of construction in favor of insureds, North Carolina courts adopted the principle that exclusions are disfavored and thus strictly construed against the insurer with virtually no independent explanation. The exclusion rule first was applied in North Carolina in *Allstate Insurance Co. v. Shelby Mutual Insurance Co.*³⁵ in 1967. Justice Lake, who wrote frequently for the court regarding insurance matters, stated simply that "[e]xclusions from and exceptions to undertakings by the company are not favored."³⁶ Presumably, the court considered this unexplained statement an obvious corollary to the general rule that courts should construe ambiguities in policies of insurance in favor of the insured. Since exclusions by their very nature must diminish coverage, and hence disfavor the insured, any ambiguity in such clauses necessarily would cause a court to construe the limiting language narrowly.³⁷ Additionally, the *Allstate* court may well have concluded that policyholders unfamiliar with the structure and terminology of insurance documents might overlook or misunderstand exclusion clauses, and that disfavoring these exclusions would protect unsuspecting insurance purchasers.³⁸ Whatever rationale prompted the rule, it has found wide

34. *Roberts*, 212 N.C. at 4, 192 S.E. at 875.

35. 269 N.C. 341, 152 S.E.2d 436 (1967).

36. *Id.* at 346, 152 S.E.2d at 440. Nor do the authorities cited for the proposition explain why exclusions should be disfavored beyond any other language purporting to define the nature and extent of coverage. See *Jamestown Mut. Ins. Co. v. Nationwide Mut. Ins. Co.*, 266 N.C. 430, 435, 146 S.E.2d 410, 414 (1966); *Anderson v. Allstate Ins. Co.*, 266 N.C. 309, 312, 145 S.E.2d 845, 848 (1966).

Whatever the reasons for disfavoring exclusions, courts have held that the burden of proving facts within an exclusion rests with the insurer. See, e.g., *Brevard v. State Farm Mut. Auto. Ins. Co.*, 262 N.C. 458, 461, 137 S.E.2d 837, 839 (1964) ("In an action to recover under an insurance policy, the burden is on the plaintiff to allege and prove coverage. On the other hand, the burden of showing an exclusion from coverage in [sic] on the insurer."); *Barclays American/Leasing, Inc. v. North Carolina Ins. Guar. Ass'n*, 99 N.C. App. 290, 294, 392 S.E.2d 772, 774 (1990); *Reliance Ins. Co. v. Morrison*, 59 N.C. App. 524, 525, 297 S.E.2d 187, 188 (1982).

37. 2 M. RHODES, *supra* note 3, states simply: "In accord with general principles, policy exceptions are strictly construed and do not reduce the coverage beyond their express terms. . . . [A]ny ambiguity in an exception clause is construed in favor of the insured." 2 M. RHODES, *supra* note 3, § 44A:3, at 7.

If this is the only theoretical justification for North Carolina courts' disfavoring of exclusions, then it would seem there is really *no* reason to disfavor such clauses more than any other limitation in coverage. If so, the *Duke University* court's attempt to distinguish jurisdictions that have construed exclusions more broadly is seriously flawed. See *Duke University*, 96 N.C. App. at 639-40, 386 S.E.2d at 765; *supra* note 24 and accompanying text. Each of the conflicting jurisdictions cited by the court has well-developed rules requiring strict construction of ambiguous policy limitations. See *New York Life Ins. Co. v. Hollender*, 38 Cal. 2d 73, 81, 237 P.2d 510, 514 (1951); *Tonkin v. California Ins. Co.*, 294 N.Y. 326, 328-29, 62 N.E.2d 215, 216 (1945); *Jack v. Standard Marine Ins. Co.*, 33 Wash. 2d 265, 271, 205 P.2d 351, 354 (1949) (quoting *Guaranty Trust Co. v. Continental Life Ins. Co.*, 159 Wash. 683, 688, 294 P. 585, 587 (1930)).

38. This supposition is supported by rather colorful language used by Justice Lake in *Jamestown Mut. Ins. Co. v. Nationwide Mut. Ins. Co.*, 266 N.C. 430, 146 S.E.2d 410 (1966):

When an insurance company, in drafting its policy of insurance, uses a "slippery" word to mark out and designate those who are insured by the policy, it is not the function of the court to sprinkle sand upon the ice by strict construction of the term. All who may, by any reasonable construction of the word, be included within the coverage afforded by the policy should be given its protection. If, in the application of this principle of construction, the limits of coverage slide across the slippery area and the company falls into a

application in recent North Carolina insurance holdings.³⁹ Like the general principle of strict construction, however, North Carolina courts employ the doctrine quite mechanically, with little if any consideration for its theoretical basis.⁴⁰

Before a court is permitted to apply any of the rules of strict construction outlined above, it first must find a provision ambiguous.⁴¹ The presence of ambiguity is a determination for the court.⁴² In cases involving exclusion clauses, courts have found ambiguity in policy provisions that are substantially contradictory,⁴³ as well as in provisions that are susceptible to more than one interpretation.⁴⁴ In the case of contradictory provisions, courts generally have denied effect to exclusionary language.⁴⁵ Clauses susceptible to multiple interpretations, however, have presented a host of difficult cases in which courts have acted largely on the equities of particular situations. In *York Industrial Center, Inc. v. Michigan Mutual Liability Co.*,⁴⁶ for example, the court held that an insurance policy that specifically exempted intentional destruction covered liability in trespass incurred by a developer who cleared erroneously surveyed property.⁴⁷ The court concluded that although trespass, by definition, requires

coverage somewhat more extensive than it contemplated, the fault lies in its own selection of the words by which it chose to be bound.

Id. at 437, 146 S.E.2d at 416.

39. *E.g.*, *State Capital Ins. Co. v. Nationwide Mut. Ins. Co.*, 318 N.C. 534, 542-43, 350 S.E.2d 66, 71 (1986); *Maddox v. Colonial Life and Accident Ins. Co.*, 303 N.C. 648, 650, 280 S.E.2d 907, 908 (1981); *Wilkins v. American Motorists Ins. Co.*, 97 N.C. App. 266, 270, 388 S.E.2d 191, 193 (1990); *W & J Rives, Inc. v. Kemper Ins. Group*, 92 N.C. App. 313, 317, 374 S.E.2d 430, 433 (1988), *disc. rev. denied*, 324 N.C. 342, 378 S.E.2d 809 (1989); *Western World Ins. Co. v. Carrington*, 90 N.C. App. 520, 523, 369 S.E.2d 128, 129 (1988); *Graham v. James F. Jackson Assocs., Inc.*, 84 N.C. App. 427, 430, 352 S.E.2d 878, 881, *disc. rev. granted*, 319 N.C. 458, 356 S.E.2d 4, *disc. rev. withdrawn*, 321 N.C. 295, 362 S.E.2d 277 (1987).

40. *See supra* note 39.

41. *Wachovia Bank & Trust Co. v. Westchester Fire Ins. Co.*, 276 N.C. 348, 354, 172 S.E.2d 518, 522 (1970). If there is no ambiguity, "the court must enforce the contract as the parties have made it and may not, under the guise of interpreting an ambiguous provision, remake the contract and impose liability upon the company which it did not assume and for which the policyholder did not pay." *Id.* For two recent cases that follow this rule, see *Wilkins*, 97 N.C. App. at 272, 388 S.E.2d at 195 and *Mastrom, Inc. v. Continental Casualty Co.*, 78 N.C. App. 483, 484, 337 S.E.2d 162, 163 (1985).

42. *Waste Management, Inc. v. Peerless Ins. Co.*, 315 N.C. 688, 694, 340 S.E.2d 374, 379 (1986). "No ambiguity . . . exists unless, in the opinion of the court, the language of the policy is fairly and reasonably susceptible to either of the constructions for which the parties contend." *Id.* (quoting *Wachovia Bank & Trust Co. v. Westchester Fire Ins. Co.*, 276 N.C. 348, 354, 172 S.E.2d 518, 522 (1970)).

43. *See, e.g.*, *Graham*, 84 N.C. App. at 430-31, 352 S.E.2d at 881 (insurer liable for coverage when policy insured against injury resulting from assault but excluded claims arising out of any criminal act); *Southeast Airmotive Corp. v. United States Fire Ins. Co.*, 78 N.C. App. 418, 418-20, 337 S.E.2d 167, 168-69 (1985) (insurer liable for coverage when policy covered damage to property arising from use of aircraft but excluded damage to property carried on an aircraft), *disc. rev. denied*, 316 N.C. 196, 341 S.E.2d 583 (1986); *Stanback v. Westchester Fire Ins. Co.*, 68 N.C. App. 107, 111-15, 314 S.E.2d 775, 778-80 (1984) (insurer liable for coverage when policy covered personal injury, including false arrest and false imprisonment, but excluded acts "committed . . . with intent to cause personal injury").

44. *See infra* notes 46-61 and accompanying text.

45. *See supra* note 43.

46. 271 N.C. 158, 155 S.E.2d 501 (1967).

47. *Id.* at 159-61, 163, 155 S.E.2d at 503-04.

an intentional act, the clearing of adjacent property did not constitute intentional destruction for the purposes of the policy exemption.⁴⁸ Similarly, the court in *W & J Rives, Inc. v. Kemper Insurance Group*⁴⁹ construed ambiguous policy language in favor of the insured. In that case, Rives, a manufacturer of sportswear, had contracted with Aetna Casualty and Insurance Company for an excess umbrella policy.⁵⁰ The Aetna policy contained an exemption for property damage "to the extent that the insured has agreed to provide [other] insurance therefor."⁵¹ When a shipment of Rives' goods was stolen in 1984, the company claimed against Aetna for the value of loss in excess of that covered by its primary insurer, Kemper Insurance Group.⁵² Aetna denied coverage, asserting that since Rives had reached an agreement with Kemper to insure the shipment, Aetna's policy did not apply.⁵³ The court, noting that Aetna's agent had issued the policy pursuant to Rives' request for excess coverage for precisely the type of shipments involved, ruled that Rives' policy with Aetna covered the uninsured losses.⁵⁴ The exemption clause, the court held, applied only to agreements with third parties such as purchasers of Rives' goods, and not to agreements with other insurance companies.⁵⁵

North Carolina courts, however, have by no means accepted assertions of ambiguity in insurance policies whenever parties have so pleaded. Accordingly, in *Western World Insurance Co. v. Carrington*,⁵⁶ the court rejected arguments by a waterproofing contractor that his general liability policy covered damages in a suit for the costs of remediating faulty workmanship.⁵⁷ The court held that the policy's exclusion for "property damage to work performed by . . . the named insured arising out of the work"⁵⁸ effectively excused the insurer, since the claim consisted "solely for bringing the quality of the insured's work up to the standard bargained for."⁵⁹ Similarly, the court of appeals in *Wilkins v. American Motorists Insurance Co.*⁶⁰ held that claims of failure to warn or properly instruct persons leasing an aircraft from the insured fell within a homeowner's policy exclusion for injuries "arising out of the ownership, maintenance [or] use . . . of . . . an aircraft."⁶¹

As these cases suggest, the ability to determine whether ambiguity exists has afforded courts considerable flexibility in determining the precise extent to

48. *Id.* at 163, 155 S.E.2d at 505-06.

49. 92 N.C. App. 313, 374 S.E.2d 430 (1988), *disc. rev. denied*, 324 N.C. 342, 378 S.E.2d 809 (1989).

50. *Id.* at 315, 374 S.E.2d at 432.

51. *Id.* at 316, 374 S.E.2d at 432.

52. *Id.* at 315, 374 S.E.2d at 432.

53. *Id.* at 315-16, 374 S.E.2d at 432.

54. *Id.* at 316-17, 374 S.E.2d at 433.

55. *Id.* at 317, 374 S.E.2d at 433.

56. 90 N.C. App. 520, 369 S.E.2d 128 (1988).

57. *Id.* at 521, 524, 369 S.E.2d at 128-29.

58. *Id.* at 522, 369 S.E.2d at 129.

59. *Id.* at 525, 369 S.E.2d at 131.

60. 97 N.C. App. 266, 388 S.E.2d 191 (1990). The opinion was handed down after *Duke University*.

61. *Id.* at 268, 272, 388 S.E.2d at 193, 195.

which policies of insurance protect policyholders. This flexibility, however, is limited by the principle that even ambiguous terms must be construed in harmony with the express provisions of the policy and the intent of the parties.⁶² Most frequently, this maxim is expressed as a refusal, "under the guise of interpreting an ambiguous provision, [to] remake the contract and impose liability upon the company which it did not assume and for which the policyholder did not pay."⁶³ Occasionally, however, North Carolina courts have articulated this principle in terms of the intent of the parties as discernable through the contract. As one court wrote: "[T]he objective of construction of terms in an insurance policy is to arrive at the insurance coverage intended by the parties when the policy was issued."⁶⁴

The best illustration of this intent-oriented reasoning is the North Carolina Supreme Court's holding in *Waste Management, Inc. v. Peerless Insurance Co.*⁶⁵ In that case, a private trash collection company accused of depositing refuse containing hazardous materials in a local landfill sought to compel its insurers to defend the action.⁶⁶ The insurers argued that a policy exclusion for pollution damage exempted the underlying suit,⁶⁷ which involved deposits over a six-year

62. *Wachovia Bank & Trust Co. v. Westchester Fire Ins. Co.*, 276 N.C. 348, 354-55, 172 S.E.2d 518, 522 (1970).

63. *E.g.*, *Waste Management, Inc. v. Peerless Ins. Co.*, 315 N.C. 688, 694, 340 S.E.2d 374, 379 (1986) (quoting *Wachovia*, 276 N.C. at 354, 172 S.E.2d at 522).

64. *Wachovia*, 276 N.C. at 354, 172 S.E.2d at 522. The *Wachovia* court went on to enumerate a series of rules of construction to be used in gleanings the intent of the parties from the language of the insurance contract:

When the policy contains a definition of a term used in it, this is the meaning which must be given to that term wherever it appears in the policy, unless the context clearly requires otherwise. In the absence of such definition, nontechnical words are to be given a meaning consistent with the sense in which they are used in ordinary speech, unless the context clearly requires otherwise. . . .

Where the immediate context in which words are used is not clearly indicative of the meaning intended, resort may be had to other portions of the policy and all clauses of it are to be construed, if possible, so as to bring them into harmony. Each word is deemed to have been put into the policy for a purpose and will be given effect, if that can be done by any reasonable construction in accordance with the foregoing principles.

Id. at 354-55, 172 S.E.2d at 522 (citations omitted).

Interestingly, the clearest expression of the importance of intent is found in the opinions of courts that originally struggled with the theoretical justification of construing a policy against the insurer. In *Bray v. Virginia Fire & Marine Ins. Co.*, 139 N.C. 390, 51 S.E. 922 (1905), for example, the court wrote:

If the clause in question is ambiguously worded, so that there is any uncertainty as to its right interpretation, or if for any reason there is doubt in our minds concerning its true meaning, we should construe it rather against the [insurer] . . . giving, of course, legal effect to the intention, if it can be ascertained, although it may have been imperfectly or obscurely expressed.

Id. at 393, 51 S.E. at 923 (emphasis added).

65. 315 N.C. 688, 340 S.E.2d 374 (1986).

66. *Id.* at 689-90, 340 S.E.2d at 376-77.

67. *Id.* at 690, 340 S.E.2d at 376-77. The exact language of the policy was as follows:

This insurance does not apply . . . to bodily injury or property damage arising out of the discharge, dispersal, release or escape of . . . toxic chemicals . . . waste materials or other . . . contaminants or pollutants into or upon land, the atmosphere or any water course or body of water; but this exclusion does not apply if such discharge, dispersal, release or escape is sudden and accidental.

Id. at 693-94, 340 S.E.2d at 379.

period, unless such damage resulted from a "sudden and accidental" release of harmful materials.⁶⁸ In holding that the exclusion applied, the court looked not only to the language of the insurance contract, but also to the underlying policy reasons for the pollution exclusion.⁶⁹ In particular, the court noted that, from the insured's perspective, coverage for pollution damage would result in a temptation to "diminish . . . precautions and relax . . . vigilance."⁷⁰ Conversely, wrote the court, by "putting the financial responsibility for pollution that may occur over the course of time upon the insured [, the policy] places the responsibility to guard against such occurrences upon the party with the most control over the circumstances most likely to cause the pollution."⁷¹ The court also noted in a footnote that at the time the policyholder purchased the insurance in question, separate policies for "environmental impairment" were available to supplement general liability coverage.⁷² The existence of pollution coverage, the court opined, was "enlightening concerning the underwriters' understanding of the scope of coverage in the [general] liability policy."⁷³ These types of policy-based observations, though relatively rare in North Carolina insurance jurisprudence, indicate the court's willingness to look beyond the mechanical application of traditional rules of construction to more subtle indications regarding the intent of parties and the meaning of policy language.

Compared to the large amount of North Carolina case law addressing construction of insurance policies, authority regarding the nature of the term "professional services" is quite limited.⁷⁴ Indeed, the court in *Duke University* found

68. See *id.* at 690, 696-700, 340 S.E.2d at 376, 380-83. The insurers also argued that the routine dumping of waste materials alleged in the underlying suit was not an "occurrence" for the purposes of the policies involved. *Id.* at 695-96, 340 S.E.2d at 379-80. The court disagreed. *Id.* at 696, 340 S.E.2d at 380.

69. See *id.* at 697-98, 340 S.E.2d at 381.

70. *Id.* at 698, 340 S.E.2d at 381.

71. *Id.* The full language of the court's observation is as follows:

The policy reasons for the pollution exclusion are obvious: if an insured knows that liability incurred by all manner of negligence or careless spills and releases is covered by his liability policy, he is tempted to diminish his precautions and relax his vigilance. Relaxed vigilance is even more likely where the insured knows that the intentional deposit of toxic material in his dumpsters, so long as it is unexpected, affords him coverage. In this case, it pays the insured to keep his head in the sand.

From the insurer's perspective, the practical reasons for the pollution exclusion are likewise clear: the lessons of Love Canal and sites like it have revealed the yawning extent of potential liability arising from the gradual or repeated discharge of hazardous substances into the environment. In addition, putting the financial responsibility for pollution that may occur over the course of time upon the insured places the responsibility to guard against such occurrences upon the party with the most control over the circumstances most likely to cause the pollution.

Id. at 697-98, 340 S.E.2d at 381.

72. *Id.* at 698 n.5, 340 S.E.2d at 381 n.5.

73. *Id.*

74. Although a number of North Carolina statutes use the term "professional service," few have resulted in litigation specifically addressing the term. The few cases that have addressed the term provide little guidance for the purposes of this analysis. Two courts have refused to find the term unconstitutionally vague. In *State v. Covington*, 34 N.C. App. 457, 238 S.E.2d 794 (1977), *disc. rev. denied*, 294 N.C. 184, 241 S.E.2d 519 (1978), the court concluded that a statute defining the "practice of professional engineering" as "any professional service . . . requiring engineering education . . . and the application of special knowledge of the mathematical, physical and engineering sciences," N.C. GEN. STAT. § 89-2(6) (1965) (current version at N.C. GEN. STAT. § 89C-3(6))

its definitional authority for the term in *Smith v. Keator*,⁷⁵ a case wholly unrelated to insurance law. In *Smith* a group of masseurs operating in Fayetteville sought to avoid local licensing requirements by arguing that state law gave the North Carolina Commissioner of Revenue sole authority to license practitioners of any "professional art of healing."⁷⁶ The court held that masseurs could not be classified as professionals.⁷⁷ A professional service, it wrote, "is one arising out of a vocation, calling, occupation, or employment involving specialized knowledge, labor, or skill, and the labor of [sic] skill involved is predominantly mental or intellectual, rather than physical or manual."⁷⁸ This definition is substantially identical to that reported by Appleman,⁷⁹ who notes further that "[m]ere employment, even though on a compensated basis, does not of itself render the person employed a 'professional' in that particular field."⁸⁰

North Carolina law has been the basis for two holdings regarding the nature of professional services for the purposes of malpractice insurance. In *Mastrom, Inc. v. Continental Casualty Co.*,⁸¹ an accounting firm that had recommended investment in one of its financially troubled clients sought coverage for its liability in fraud.⁸² The court held that investment advising could not be brought within the firm's coverage for "damages . . . arising out of the per-

(1989)), was not unconstitutionally vague as applied to an engineer accused of practicing engineering without registering as required by statute. *Covington*, 34 N.C. App. at 460-61, 238 S.E.2d at 797. In *Roberts v. Durham County Hosp. Corp.*, 56 N.C. App. 533, 289 S.E.2d 875 (1982), *aff'd per curiam*, 307 N.C. 465, 298 S.E.2d 384 (1983), a statute establishing a period of limitations for "malpractice arising out of the performance of or failure to perform professional services," N.C. GEN. STAT. § 1-15(c) (1983), was held not to be unconstitutionally vague as applied to a doctor who failed to remove an intravenous catheter after rendering care. *Roberts*, 56 N.C. App. at 538-41, 289 S.E.2d at 878-80.

Two cases involving North Carolina's unfair and deceptive trade practice statute, N.C. GEN. STAT. § 75-1.1 (1988), have interpreted a professional services exemption in that law rather broadly. In *Cameron v. New Hanover Memorial Hosp., Inc.*, 58 N.C. App. 414, 293 S.E.2d 901, *disc. rev. denied*, 307 N.C. 127, 297 S.E.2d 399 (1982), the court held that the statute's exemption for "professional services rendered by a member of a learned profession," N.C. GEN. STAT. § 75-1.1(b), exempted doctors who, as members of a hospital administrative board, voted to deny staff privileges to a podiatrist. *Cameron*, 58 N.C. App. at 446-47, 293 S.E.2d at 920-21. The court concluded that "the nature of this consideration of whom to grant hospital staff privileges is a necessary assurance of good health care; certainly, this is the rendering of 'professional services.'" *Id.* at 447, 293 S.E.2d at 921. Similarly, the court held in *Abram v. Charter Medical Corp.*, 100 N.C. App. 718, 398 S.E.2d 331 (1990), that a drug rehabilitation services provider who requested a review of a competitor's certificate-of-need application was acting as a responsible provider of professional services and thus outside the scope of the statute. *Id.* at 722, 398 S.E.2d at 334. *But cf.* *Winston Realty Co. v. G.H.G., Inc.*, 70 N.C. App. 374, 375-76, 382, 320 S.E.2d 286, 287-88, 291 (1984) (employment agent who recommended convicted embezzler for accounting position was not rendering professional services for purposes of professional services exemption in unfair and deceptive trade practices statute), *aff'd*, 314 N.C. 90, 331 S.E.2d 677 (1985).

75. 21 N.C. App. 102, 203 S.E.2d 411, *aff'd*, 285 N.C. 530, 206 S.E.2d 203, *appeal dismissed*, 419 U.S. 1043 (1974).

76. *Id.* at 103, 105, 203 S.E.2d at 413, 415; *see* N.C. GEN. STAT. § 105-41(a) (1989) (amended 1990).

77. *Smith*, 21 N.C. App. at 105-06, 203 S.E.2d at 415.

78. *Id.* (quoting *Marx v. Hartford Accident & Indem. Co.*, 183 Neb. 12, 14, 157 N.W.2d 870, 872 (1968)).

79. 7A J. APPLEMAN, INSURANCE LAW AND PRACTICE § 4504.01, at 309-10 (W. Berdal ed. 1979).

80. *Id.* § 4501.12 at 282-83.

81. 78 N.C. App. 483, 337 S.E.2d 162 (1985).

82. *Id.* at 483-84, 337 S.E.2d at 163.

formance of professional services . . . as an accountant.”⁸³ Similarly, in *Smith v. Travelers Indemnity Co.*,⁸⁴ a federal district court found that an attorney’s professional malpractice policy did not cover his failure to return \$15,000 given to him for investment.⁸⁵ Noting that no attorney-client relationship had been established at the time the lawyer received the money, the court concluded that the plaintiff had not employed the attorney “for the purposes of obtaining any legal assistance.”⁸⁶ The transaction, therefore, was outside the coverage of the malpractice policy.⁸⁷

The second element of the *Duke University* court’s definition of professional services—the rule that the nature of the act rather than the title or character of the actor should determine the presence of such services—is new to North Carolina law. While at least three other jurisdictions have applied this rule,⁸⁸ each has done so in a largely mechanical fashion, with no discussion of underlying rationale or purpose.⁸⁹

Fundamental to an understanding of the holding in *Duke University* is a careful analysis of the types of activities that the term “professional services” may encompass. Relevant activities can be grouped into two categories. First are services that even the narrowest definition of the term will embrace. These are acts informed by special knowledge.⁹⁰ Examples include a surgeon’s selection of an incision site, a lawyer’s drafting of a will, and an accountant’s classification of expenses for bookkeeping purposes. The second, much broader, category of services consists of those actions that, though not arising directly from some specially informed judgment, are associated so closely with the per-

83. *Id.* at 484-85, 337 S.E.2d at 163. The court concluded: “Nowhere do we find any definition of ‘accountant’ broad enough to include the sale of securities, nor any definition of ‘accountant’ as one who offers a general range of financial services.” *Id.* at 485, 337 S.E.2d at 164.

84. 343 F. Supp. 605 (M.D.N.C. 1972).

85. *Id.* at 605-06, 610.

86. *Id.* at 609.

87. *Id.* at 609-10. The terms of the professional malpractice at issue involved provided for insurance for claims “arising out of the performance of professional services for others in the insured’s capacity as a lawyer.” *Id.* at 608.

88. See *Mason v. Liberty Mut. Ins. Co.*, 370 F.2d 925, 926 (5th Cir. 1967) (per curiam) (court applying Louisiana law held that claim by infirm patient for injuries received from hypodermic injection administered by student nurse was within professional services exclusion of general liability policy); *Danks v. Maher*, 177 So. 2d 412, 417-18 (La. App. 1965) (claim by patient injured when hospital nurse failed to count accurately laparotomy sponges removed from patient undergoing abdominal surgery was not within professional services exclusion of hospital’s general liability policy); *D’Antoni v. Sara Mayo Hosp.*, 144 So. 2d 643, 646-47 (La. App. 1962) (claim by patient injured in fall from bed after hospital attendants failed to raise bed rails was not within professional services exclusion of hospital’s general liability policy); *Marx v. Hartford Accident & Indem. Co.*, 183 Neb. 12, 14, 157 N.W.2d 870, 872 (1968) (when fire resulted from laboratory technician’s refilling steam sterilizer with benzene, court concluded that damages were not excluded by a general liability policy exclusion for the rendering of professional services); *Multnomah County v. Oregon Auto. Ins. Co.*, 256 Or. 24, 28-29, 470 P.2d 147, 150 (1970) (inmate’s claim of failure to provide necessary medical treatment against county jailer within professional services exclusion of general liability policy).

89. See *supra* note 88. Appleman reports the rule without comment. 7A J. APPLEMAN, *supra* note 79, § 4504.01, at 310 (“[I]n determining whether a particular act is a ‘professional service’ the court must look not to the title or character of the party performing the act, but to the act itself.”).

90. See *Duke University*, 96 N.C. App. at 639, 386 S.E.2d at 765. Although the definition of professional services used in the principal case is more complex and precise, see *id.*, this Note uses the phrase, “informed by special knowledge” or phrases similar thereto, as shorthand for the concept.

formance of informed services that they are arguably professional in character. Activities within this second category can be ordered according to the strength of their association with acts in the first category. For example, when a surgeon fails to remove an instrument from a patient before closure, the mechanical act of inventorying instruments or recognizing that one has been left in the patient requires no special skill or knowledge, yet it seems likely that even the most conservative of courts would classify such an action as a failure to render a professional service. On the opposite end of the spectrum are actions that are only loosely associated with the furnishing of a service requiring special skill or knowledge. Maintenance of a doctor's office premises, for example, may well constitute an action so remote from the application of specialized knowledge as to render meaningless any assertion that it is professional in character.

In interpreting professional service exclusion clauses, there is a temptation to assert that only those activities classified within the first category above—actions informed by some special skill or knowledge—should be deemed to fall within the exclusion language. Indeed, this is one possible interpretation of the holding in *Duke University*. Specifically, the court observed that stabilizing the dialysis chair and assisting the patient did “not require any special skills or training.”⁹¹ In its strictest sense, this statement would exclude almost any act not informed by some special knowledge or skill.

Most likely, however, the *Duke University* court concluded only that the acts and omissions at issue in the case could not be classified without ambiguity. This determination, in turn, required a decision in favor of the insured, since the applicable rules of construction demand that any ambiguity be resolved against the insurer. This interpretation of the holding is supported by the court's observation that “the claim in this case could come within [a professional liability policy] and yet not fall within a professional services exclusion.”⁹² Indeed, this notion that a single claim may be susceptible to coverage under both general and professional liability policies suggests that the court tacitly recognized that certain professional services may be classified along a spectrum according to the strength of their association with actions which are truly informed by special skill and knowledge.

While both the above modes of analysis achieve results commensurate with North Carolina insurance jurisprudence, there is little to recommend either. If professional services are restricted exclusively to actions informed by special knowledge, general liability insurance will cover a whole range of activities closely associated with informed actions, but not themselves requiring special knowledge, regardless of the applicable policy's exclusionary language. While the opinion in *Duke University* does not require this result, its largely mechanical analysis does little to dissuade future courts from this interpretation.

The second interpretation of the court's holding also presents important difficulties. By its express terms, the opinion institutionalizes an overlap between general liability policies, the exclusionary clauses of which must be con-

91. *Id.* at 641, 386 S.E.2d at 766.

92. *Id.* at 640, 386 S.E.2d at 765.

strued narrowly, and professional services liability policies, which must be construed broadly.⁹³ This overlap is likely to encourage litigation between competing insurers as they attempt to apportion liability with regard to actions that are covered under two policies. Policyholders, in contrast, will be encouraged to use any overlap in experience-rated policies to file claims under less expensive general liability policies, thus shifting the cost of insuring the particular activity away from populations most able to limit the risk involved. Finally, the overlap will hamper efforts by insurance providers to classify activities according to the nature of the risk involved and to set premiums accordingly.

The court in *Duke University* could have reduced the potential overlap between general liability and professional services policies by broadening its analysis to consider the intent of the parties and the underlying theory of the policy provisions at issue as a means of reducing ambiguity in the term "professional services." In particular, the court failed to note that most insurance companies attempt to maximize profitability by grouping policyholders into categories that share common risks.⁹⁴ Classification allows insurers to measure accurately the costs of insuring risks and to price policies, copayments, and deductibles in a manner that maximizes both risk avoidance and risk protection.⁹⁵ Furthermore, accurate classification serves important societal interests by placing the burden of insuring a risk on those persons most able to protect against it.⁹⁶

As suggested earlier, judicial consideration of these principles of risk allocation and insurance economics is not without precedent. In *Waste Management, Inc. v. Peerless Insurance Co.*⁹⁷ the North Carolina Supreme Court, looking to what it termed "policy reasons,"⁹⁸ gave considerable weight to the fact that the pollution exclusion at issue encouraged policyholders to "guard against" the pollution risk.⁹⁹ The *Waste Management* court noted that the exclusion was entirely consistent with the insurer's perception of the "yawning extent" of the potential liability involved¹⁰⁰ and the availability of special forms of insurance to

93. See *id.*

94. K. ABRAHAM, *DISTRIBUTING RISK: INSURANCE, LEGAL THEORY, AND PUBLIC POLICY* 15 (1986).

95. *Id.*

96. *Id.* at 138. Of course, insurance may serve a distributional function as well. The whole point of insurance is to enable a single individual to avoid the entire risk for a particular activity. See *id.* at 18. In virtually all cases, insurers' inability to predict risk for particular individuals perfectly, combined with the cost of information gathering, prevents classification of insureds into groups so small that the distributive benefit of insurance is lost. *Id.* at 15.

Some commentators have suggested that an appropriate function of insurance law is to assure that risk classification by insurers does not inappropriately limit the socially desirable goal of risk distribution. See *id.* at 18-31. A discussion of the appropriateness of judicial interpretation of private contracts of insurance for risk distribution purposes is beyond the scope of this Note. It should be noted, however, that the court in *Duke University* made no express or implied reference to this objective. Further, while the ever increasing cost of health care insurance may provide an argument that broader risk distribution is appropriate for professional services in the field of health care, the policy adopted by the court in *Duke University* would appear to apply to *all* forms of professional service, inside and outside the health care arena. See *infra* notes 103-06 and accompanying text.

97. 315 N.C. 688, 340 S.E.2d 374 (1986).

98. *Id.* at 697, 340 S.E.2d at 381.

99. *Id.* at 697-98, 340 S.E.2d at 381.

100. *Id.* at 698, 340 S.E.2d at 381.

cover the specific risk.¹⁰¹

Similar arguments can be made regarding professional services coverage. Specifically, when a particular action is related closely to some action informed by special skill or knowledge, the risks associated with the action are likely to be peculiarly within the control of the actor in his capacity as a professional. Under the theory that the costs of insuring risks should be borne by those most able to control the risk, such actions should be insured under professional services insurance policies rather than under contracts of general liability insurance. Alternatively, when a particular action is only loosely associated with the rendering of some specially informed service or skill, the risks associated with the action are more likely to be within the control of the actor as a member of society in general. In other words, the risks associated with the action are *not* disproportionately within the actor's control in his capacity as a professional. Such acts are, therefore, most appropriately insured under general liability coverage.

Notably, this mode of analysis is consistent with the rule that the nature of an act and not the title or position of the actor is determinative with respect to the presence of professional services. Looking to the title of an actor gives no indication of whether the particular action at issue is peculiarly within her control as a professional. To build on an example already suggested, a doctor's liability for failure to maintain safe premises is not closely related to the medical profession, but rises from a duty common to owners of business premises as a whole. The risk and liability are not closely related to the doctor's skill or learning, nor disproportionately within her control as a professional.

Applying the suggested analysis to the facts in *Duke University*, it is clear that the court, though failing to recognize the intent of the parties as a factor in construing the policy at issue, ultimately arrived at a result consistent with such an analysis. As the court intuitively concluded, the failure to fix the casters on the patient's dialysis chair was not an omission uniquely or disproportionately associated with the providing of skilled health care. The court suggested in its closing analysis:

Although the dialysis chair was a specialized piece of equipment, the injury was not related to any special function of the chair but merely resulted from the presence of casters on the chair which enable it to be easily moved. The injury may have been avoided by simply locking the casters or holding the chair.¹⁰²

As this statement suggests, much the same risk is found in the failure of any business operator to safely use and maintain equipment on his premises. In this light, the failure to lock the casters was not an action informed by special knowledge, nor so closely related to the employment of specially informed judgment as to be disproportionately associated with the provision of health care.

Finally, although *Duke University* presents the problem of analyzing professional liability exclusions in terms of the health care profession, the question is

101. *Id.* at 698 n.5, 340 S.E.2d at 381 n.5.

102. *Duke University*, 96 N.C. App. at 641, 386 S.E.2d at 766.

by no means unique to that field. Cases in other jurisdictions have addressed the problem in the context of engineering firms,¹⁰³ barbers and beauticians,¹⁰⁴ and hotel operations.¹⁰⁵ One innovative insurer has even suggested that a professional services exclusion apply to babysitting in the home.¹⁰⁶ Indeed, the problem is likely to arise in any professional area in which insurance companies offer or limit protection for professional liability.

Although the mode of analysis suggested by this Note may be applied to any of these fields, the classification scheme it proposes certainly is not without ambiguity. Close cases will continue to pose difficult choices. But consideration of the economic and risk allocation objectives of policy exclusions, and in particular the need to place the cost of insurance on those most able to prevent the injury insured, would provide a useful guidepost for jurists. Further, such considerations would serve as an appropriate balance to the well-established goal of protecting the insured against unfair and ambiguous policy language.

North Carolina has joined other jurisdictions in holding that courts should resolve ambiguities in contracts of insurance in favor of the insured.¹⁰⁷ This canon of construction, which opinions of early courts grounded solidly in the need to protect policyholders from the "slippery" wording of complex and technical contracts, is now a fixture of modern insurance jurisprudence. Yet, in many respects, contemporary application of the rule has become largely mechanical. Such was the case in *Duke University*, where the North Carolina Court of Appeals narrowly construed a general liability policy exclusion for professional services. While the *Duke University* court's result arguably is correct, the opinion fails to articulate a rationale that will guide future courts in resolving similar disputes over the status of professional services exclusion claims. Further, by openly acknowledging that current rules of construction create an area of overlap between general liability and professional services coverages, the

103. See, e.g., *First Ins. Co. v. Continental Casualty Co.*, 466 F.2d 807 (9th Cir. 1972) (engineering drawings requiring earthen fill which damaged oil pipeline constituted failure of professional services for purposes of contractor's general liability policy exclusion); *Womack v. Travelers Ins. Co.*, 251 So. 2d 463 (La. App. 1971) (engineering firm's failure to confirm drawings showing location of gas pipeline was professional malpractice within general liability policy exclusion); *Atlantic Mut. Ins. Co. v. Continental Nat'l Am. Ins. Co.*, 123 N.J. Super. 241, 302 A.2d 177 (1973) (failure of engineering firm to properly oversee trench fortification was failure to render professional services excluded by general liability policy).

104. See, e.g., *Ocean Accident & Guarantee Corp. v. Herzberg's Inc.*, 100 F.2d 171 (8th Cir.) (beautician's negligent use of electrical depilator was failure of professional services for purposes of exclusion in department store's general liability policy), *cert. denied*, 306 U.S. 645 (1939); *Ruotolo v. Aetna Casualty & Sur. Co.*, 56 Misc. 2d 45, 287 N.Y.S.2d 622 (1968) (barber who hit customer in eye with whisk broom not rendering tonsorial services for purposes of exclusion in general liability policy); *Lady Beautiful, Inc. v. Zurich Ins. Co.*, 16 Ohio Misc. 169, 240 N.E.2d 894 (1968) (beauty parlor patron's injuries from faulty hair dryer not a result of acts or omissions excluded by professional services clause of general liability policy).

105. See *Maryland Casualty Co. v. Crazy Water Co.*, 160 S.W.2d 102 (Tex. Civ. App. 1942) (hotel patron's injuries in bath prescribed by hotel physician and attended by hotel "tubber" (one who draws the bathwater, checks the water temperature, and otherwise assists a bather) not excluded by professional services clause of general liability policy).

106. See *Gulf Ins. Co. v. Tilley*, 280 F. Supp. 60 (N.D. Ind. 1967), *aff'd per curiam*, 393 F.2d 119 (7th Cir. 1968).

107. See *Grabbs v. Farmers' Mut. Fire Ins. Ass'n*, 125 N.C. 389, 34 S.E. 503 (1899); *supra* text accompanying note 1.

opinion encourages future litigation and impedes insurers' abilities to classify and price insurance policies covering risks related to professional services accurately. These difficulties will continue until North Carolina courts begin to consider systematically the underlying nature of insurance contracts and the importance of accurate risk allocation in interpreting and coordinating the insurance of professional services.

SAMUEL M. TAYLOR

***Brown v. Lumbermens Mutual Casualty Co.*: When Does Exhaustion of Policy Limits Terminate an Insurer's Duty to Defend?**

A liability insurance policy imposes two duties on an insurer with respect to the insured: the duty to indemnify and the duty to defend.¹ In the past, when an insurer believed liability would exceed policy limits, it argued that the latter duty is dependent on the former, and that therefore exhaustion of policy limits terminated the duty to defend. Although there is general consensus today that the duty to defend is independent of and broader than the duty to indemnify,² most courts have held that some forms of exhaustion can extinguish the duty to defend.³ Courts following this rule, however, have not agreed on the forms of exhaustion that suffice.⁴

In *Brown v. Lumbermens Mutual Casualty Co.*,⁵ the North Carolina Supreme Court addressed this problem, one of first impression in the state. The court held that when a policy reads, "Our duty to settle or defend ends when our limit of liability for this coverage has been exhausted,"⁶ the insurer's duty to defend continues until exhaustion of coverage occurs either through settlement of a claim against the insured or through a judgment against the insured.⁷

This Note examines the language changes insurers have made in their policies in an attempt to terminate the duty to defend on exhaustion of liability coverage. It discusses the split in opinion between courts that have held exhaustion ends the duty to defend, and those that have held it does not. The Note then addresses the disagreement among courts in the former group over the form of exhaustion that suffices to end the duty. It analyzes the *Brown* court's reasoning and looks at the policy implications of the decision. The Note concludes that the decision is correct as a matter of policy, but that the court should have rested its holding on less vulnerable grounds.

In *Brown* defendant Lumbermens Mutual Casualty Company (LMCC) sold plaintiffs Doyle and Coleen Brown a \$25,000 automobile insurance policy that contained the following duty-to-defend provision:

We will pay damages for bodily injury or property damage for

1. There is no North Carolina statutory requirement that an insurer provide its insured with a defense. See, e.g., N.C. GEN. STAT. § 20-279.21 (1989). A company, however, may provide this by contract. *Brown v. Lumbermens Mut. Casualty Co.*, 326 N.C. 387, 391, 390 S.E.2d 150, 152 (1990).

2. E.g., *Waste Management of Carolinas, Inc. v. Peerless Ins. Co.*, 315 N.C. 688, 691, 340 S.E.2d 374, 377, *reh'g denied*, 316 N.C. 386, 346 S.E.2d 134 (1986); see A. WINDT, *INSURANCE CLAIMS AND DISPUTES* 201 n.262 (2d ed. 1988 & Supp. 1990) (listing cases holding that the duty to defend is independent of the duty to pay); Zulkey & Pollard, *The Duty to Defend After Exhaustion of Policy Limits*, *FOR THE DEF.*, June 1985, at 21, 21 ("It is a well-recognized legal principle that an insurer's duty to defend is broader than the insurer's duty to indemnify the insured.").

3. See *infra* note 64 and accompanying text. For examples of forms of exhaustion that may extinguish the duty to defend, see *infra* text accompanying note 94.

4. See *infra* notes 65-67 and accompanying text.

5. 326 N.C. 387, 390 S.E.2d 150 (1990).

6. *Id.* at 389, 390 S.E.2d at 151 (quoting policy).

7. *Id.* at 394, 390 S.E.2d at 154.

which any covered person becomes legally responsible because of an auto accident. We will settle or defend, as we consider appropriate, any claim or suit asking for these damages. In addition to our limit of liability, we will pay all defense costs we incur. Our duty to settle or defend ends when our limit of liability for this coverage has been exhausted.⁸

When Joan Hinson brought a tort action against the Browns as a result of a car accident, LMCC hired a law firm to represent the Browns.⁹ In response to Hinson's complaint, the firm filed an answer denying any negligence by the Browns.¹⁰ In an affidavit, however, the law firm gave its opinion that the court probably would find the Browns liable, and it predicted a verdict between \$50,000 and \$75,000.¹¹ Against the Browns' wishes but at the direction of LMCC, the firm offered Hinson a \$25,000 settlement.¹² When Hinson refused the offer, LMCC paid her this amount in exchange for a release of its obligation; Hinson did not release the Browns.¹³ LMCC terminated the Browns' defense and discharged the law firm.¹⁴ The court later entered a \$45,000 judgment against the Browns, crediting to them the \$25,000 payment by LMCC.¹⁵

The Browns sued LMCC, claiming that it was obligated to defend them even after paying the policy limit. LMCC argued that by paying Hinson the full coverage, it discharged its duty to defend the Browns. The trial court entered summary judgment for LMCC.¹⁶

The North Carolina Court of Appeals reversed, holding that the duty-to-defend provision was ambiguous.¹⁷ The North Carolina Supreme Court affirmed in a four-three decision.¹⁸ It agreed with the court of appeals that the duty-to-defend provision was ambiguous because it did not specify what form of exhaustion terminated the duty to defend.¹⁹ Proof of ambiguity lay in the parties' conflicting interpretations: the Browns argued that only a settlement or judgment could exhaust the policy limits, but LMCC maintained that any man-

8. *Id.* at 389, 390 S.E.2d at 151.

9. *Id.* at 390, 390 S.E.2d at 151.

10. *Brown v. Lumbermens Mut. Casualty Co.*, 90 N.C. App. 464, 466, 369 S.E.2d 367, 368 (1988), *aff'd in part*, 326 N.C. 387, 390 S.E.2d 150 (1990).

11. *Brown*, 326 N.C. at 390, 390 S.E.2d at 151.

12. *Id.* at 390, 390 S.E.2d at 151-52.

13. *Id.* at 390, 390 S.E.2d at 152. LMCC paid Hinson pursuant to North Carolina General Statutes § 1-540.3, which allows advance payments to an injured person by a party or the party's insurer. *Id.* (citing N.C. GEN. STAT. § 1-540.3(a) (1983)).

14. *Id.*

15. *Id.*

16. *Id.* at 391, 390 S.E.2d at 152.

17. *Brown v. Lumbermens Mut. Casualty Co.*, 90 N.C. App. 464, 477, 369 S.E.2d 367, 374 (1988), *aff'd in part*, 326 N.C. 387, 390 S.E.2d 150 (1990). The supreme court did not review the decision of the court of appeals that an insured cannot hold her insurer vicariously liable for the negligence of an attorney it has hired, because the attorney is an independent contractor. For an analysis of this decision of the court of appeals, see Note, *Brown v. Lumbermens Mutual Casualty Co.: The Rock, the Hard Place, and the Insurance Defense Attorney*, 67 N.C.L. REV. 1424 (1989).

18. *Brown*, 326 N.C. at 397, 390 S.E.2d at 156. Chief Justice Exum wrote the majority opinion, which was joined by Justices Frye, Martin, and Mitchell. Justice Whichard wrote the dissent, which was joined by Justices Meyer and Webb.

19. *Id.* at 393-94, 390 S.E.2d at 153-54.

ner of payment would suffice.²⁰ The supreme court affirmed the court of appeals decision to interpret the provision against LMCC, and concluded that LMCC's duty to defend the Browns did not end until there was a settlement or judgment.²¹ In his opinion for the majority, Chief Justice Exum relied in part on supporting opinions from courts in three states and one federal district that have interpreted identical language.²²

Justice Whichard dissented, arguing that there was no ambiguity.²³ According to the dissent's examination of language changes in duty-to-defend provisions over the past twenty-five years, the majority's interpretation amended the disputed provision by including words that LMCC patently omitted.²⁴ Justice Whichard also claimed that the majority's interpretation of the provision "render[ed] it [the provision] meaningless surplusage," because the duty to defend, if it does not end earlier, always terminates on settlement or judgment.²⁵ The dissent went on to distinguish the cases the majority claimed to follow,²⁶ and then argued that the majority's interpretation ran counter to the public policy of the state because it discouraged payments to claimants before judgment or full settlement.²⁷

Because *Brown* decided an issue of first impression in North Carolina, it is helpful to examine other courts' analyses of the problem. It is important, however, to distinguish cases interpreting different policy language, for the insurance industry has changed the wording of the standard liability policy several times. Before 1955 the defense coverage clause was separate from the liability coverage clause, and the policy limited the insurer's duty to defend only by prefacing the defense provision. The preface stated that the insurer would defend only suits "[a]s respects the insurance afforded by the other terms of this policy under coverages A [bodily injury liability] and B [property damage liability]."²⁸ In 1955 the insurance industry changed the preliminary language slightly, to read: "'With respect to such insurance as is afforded by this policy for bodily injury liability and for property damage liability.'"²⁹ The defense coverage clause and the liability coverage clause remained separate.³⁰ In 1956 the drafters merged the defense coverage clause into the liability coverage clause and changed the

20. *Id.* at 394, 390 S.E.2d at 154.

21. *Id.* Courts resolve contractual ambiguity against the drafter, in this case the insurer. *Woods v. Nationwide Mut. Ins. Co.*, 295 N.C. 500, 506, 246 S.E.2d 773, 777 (1978).

22. *Brown*, 326 N.C. at 394-96, 390 S.E.2d at 154-55 (citing *Stanley v. Cobb*, 624 F. Supp. 536, 537 (E.D. Tenn. 1986); *Samplly v. Integrity Ins. Co.*, 476 So. 2d 79, 83 (Ala. 1985); *Anderson v. United States Fidelity & Guar. Co.*, 177 Ga. App. 520, 521, 339 S.E.2d 660, 661 (1986); *Pareti v. Sentry Indem. Co.*, 536 So. 2d 417, 420-21 (La. 1988)).

23. *Id.* at 397-98, 390 S.E.2d at 156 (Whichard, J., dissenting).

24. *Id.* at 399, 390 S.E.2d at 157 (Whichard, J., dissenting).

25. *Id.* at 400, 390 S.E.2d at 157 (Whichard, J., dissenting).

26. *Id.* at 400-01, 390 S.E.2d at 158 (Whichard, J., dissenting).

27. *Id.* at 402-03, 390 S.E.2d at 158-59 (Whichard, J., dissenting).

28. Van Vugt, *Termination of the Insurer's Duty to Defend by Exhaustion of Policy Limits*, 44 INS. COUNS. J. 254, 256 (1977) (quoting N. RISJARD & J. AUSTIN, *AUTOMOBILE LIABILITY INSURANCE CASES, STANDARD PROVISIONS AND APPENDIX 2* (1964)).

29. *Id.* (quoting N. RISJARD & J. AUSTIN, *AUTOMOBILE LIABILITY INSURANCE CASES, STANDARD PROVISIONS AND APPENDIX 16* (1964)).

30. *Id.*

language limiting the duty to defend. The revised policy provided that the insurer would defend only suits " 'seeking damages which are payable under the terms of this policy.' " ³¹ In 1966 the drafters again changed the language limiting the duty to defend: " '[T]he company shall not be obligated to pay any claim or judgment or to defend any suit after the applicable limit of the company's liability has been exhausted by payment of judgments or settlements.' " ³² The insurance industry effected these changes to make clear that the insurer's duty to defend extends only as far as its duty to indemnify, and that therefore an insurer terminates its duty to defend by paying the policy limits in any manner. ³³ Nonetheless, courts have been far from unanimous in rendering such an interpretation.

The decisions construing the pre-1955 policy language are split. Some cases focus on the restrictiveness of the preliminary language and hold that the duty to defend is dependent on the duty to indemnify; therefore, payment of the policy limit terminates the duty to defend. ³⁴ At least one court also based its opinion on the anomaly of requiring an insurer with no financial interest in the outcome of the case to defend an insured. ³⁵ Cases holding that the duty to defend is independent of the duty to pay, and that therefore exhaustion of coverage limits does not terminate the duty to defend, focus on the policy language providing that "the company shall defend any suit" and the separateness of the defense coverage clause and the liability coverage clause. ³⁶

There is also a split in opinions interpreting the 1955 policy language. Cases holding that exhaustion of policy limits ends the duty to defend focus on the changed language, construing "such insurance as is afforded by this policy" to refer to the amount of liability coverage afforded. ³⁷ They also point out that the limiting phrase immediately precedes the defense agreement. ³⁸ Courts that have interpreted the 1955 policy against the insurer have held that the duty to

31. *Id.* at 257 (quoting N. RISJARD & J. AUSTIN, *AUTOMOBILE LIABILITY INSURANCE CASES, STANDARD PROVISIONS AND APPENDIX 34* (1964)).

32. *Id.* (quoting N. RISJARD & J. AUSTIN, *AUTOMOBILE LIABILITY INSURANCE CASES, STANDARD PROVISIONS AND APPENDIX 265* app. (1964)).

33. *Id.* at 256.

34. *E.g.*, *General Casualty Co. v. Whipple*, 328 F.2d 353, 357 (7th Cir. 1964) (applying Illinois law); *Denham v. La Salle-Madison Hotel Co.*, 168 F.2d 576, 584 (7th Cir.) (same), *cert. denied*, 335 U.S. 871 (1948); *see supra* text accompanying note 28.

35. *Lumbermen's Mut. Casualty Co. v. McCarthy*, 90 N.H. 320, 323, 8 A.2d 750, 752 (1939).

36. *E.g.*, *American Casualty Co. v. Howard*, 187 F.2d 322, 326 (4th Cir. 1951) (applying South Carolina law); *Anchor Casualty Co. v. McCaleb*, 178 F.2d 322, 325 (5th Cir. 1949) (applying Texas law); *St. Paul Fire & Marine Ins. Co. v. Thompson*, 150 Mont. 182, 187-88, 433 P.2d 795, 798 (1967); *American Employers Ins. Co. v. Goble Aircraft Specialties, Inc.*, 205 Misc. 1066, 1073-74, 131 N.Y.S.2d 393, 400 (N.Y. Sup. Ct. 1954), *appeal withdrawn*, 1 A.D.2d 1008, 154 N.Y.S.2d 835 (N.Y. App. Div. 1956); *see supra* text accompanying note 28.

37. *E.g.*, *Allstate Ins. Co. v. Montgomery Trucking Co.*, 328 F. Supp. 415, 416-17 (N.D. Ga. 1971); *Commercial Union Ins. Co. v. Adams*, 231 F. Supp. 860, 866 (S.D. Ind. 1964); *Liberty Mut. Ins. Co. v. Mead Corp.*, 219 Ga. 6, 9, 131 S.E.2d 534, 535-36 (1963); *Oda v. Highway Ins. Co.*, 44 Ill. App. 2d 235, 255-56, 194 N.E.2d 489, 500 (1963); *Sutton Mut. Ins. Co. v. Rolph*, 109 N.H. 142, 144, 244 A.2d 186, 188 (1968); *see supra* text accompanying note 29.

38. *E.g.*, *Keene Corp. v. Insurance Co.*, 597 F. Supp. 946, 950-51 (D.D.C. 1984), *vacated on other grounds*, 631 F. Supp. 34 (D.D.C. 1985) (issues rendered moot by settlement pending appeal).

defend is independent of and broader than the duty to indemnify.³⁹ At least one case holds that the duty to defend continues after judgment or settlement;⁴⁰ the court supported its opinion by declaring the policy language ambiguous and by turning to the state's public policy of protecting insureds.⁴¹

Only one case interprets the 1956 policy. In *Simmons v. Jeffords*⁴² the court construed the language as obligating the insurer to defend the insured until final settlement or judgment.⁴³ The court rejected the insurer's attempt to tender the policy limits into court before settlement or judgment.⁴⁴

The only court interpreting the post-1966 policy reached the same conclusion. In *Conway v. Country Casualty Insurance Co.*,⁴⁵ the Illinois Supreme Court held that the duty to defend does not end when the insurer pays the policy limits to the claimant and that the payment does not result in a release of the insured.⁴⁶ The court underscored the unambiguous language of the policy, which referred to "judgments or settlements;" since neither had occurred, the insurer could not claim relief of its duty to defend.⁴⁷ In a case decided five years later by the same court, the court relieved the insurer of its duty to defend when it paid the policy limits through judgments and settlements.⁴⁸

The language variations described above are different forms of the standard liability policy. Other cases have interpreted policies that, like the *Brown* policy, do not fall into one of these categories. The policy in *Gross v. Lloyds of London Insurance Co.*⁴⁹ provided that exhaustion would occur by payment of judgments or settlements "or after such limit of the Company's liability has been tendered for settlements."⁵⁰ The Wisconsin Supreme Court interpreted the added language as contemplating payment before judgment or settlement; the insurer therefore could terminate its duty to defend by tendering the insured's policy limits into court.⁵¹ In *National Union Insurance Co. v. Phoenix Assurance Co.*,⁵² the court released the insurer from its duty to defend after it deposited the

39. *E.g.*, *Aetna Casualty & Sur. Co. v. Certain Underwriters at Lloyds of London*, 56 Cal. App. 3d 791, 804, 129 Cal. Rptr. 47, 55 (1976) (primary insurer's duty to defend continued, but court ordered excess carriers to reimburse primary insurer for costs beyond the policy limits); *Travelers Indem. Co. v. East*, 240 So. 2d 277, 279 (Miss. 1970) ("The defense clause . . . is . . . a contractual right of the insured, for which he is paid a premium. . . ."); *Prince v. Universal Underwriters Ins. Co.*, 143 N.W.2d 708, 717 (N.D. 1966) ("The defendant . . . had a duty to defend all actions against the plaintiff regardless of the amount involved.").

40. *Commercial Union Ins. Co. v. Pittsburgh Corning Corp.*, 553 F. Supp. 425, 431 (E.D. Pa. 1981).

41. *Id.* at 429, 431.

42. 260 F. Supp. 641 (E.D. Pa. 1966).

43. *Id.* at 642.

44. *Id.* at 641; see *supra* text accompanying note 31.

45. 92 Ill. 2d 388, 442 N.E.2d 245 (1982).

46. *Id.* at 395, 442 N.E.2d at 247.

47. *Id.* at 396, 442 N.E.2d at 248; see *supra* text accompanying note 32.

48. *Zurich Ins. Co. v. Raymark Indus.*, 118 Ill. 2d 23, 56, 514 N.E.2d 150, 165 (1987).

49. 121 Wis. 2d 78, 358 N.W.2d 266 (1984).

50. *Id.* at 83, 358 N.W.2d at 269.

51. *Id.* at 89, 358 N.W.2d at 272. The insurer could not terminate its duty to defend on this occasion, however, because it had not highlighted the added wording in the policy or otherwise given clear notice to the insured of the substantial change in the language. *Id.* at 88, 358 N.W.2d at 271.

52. 301 A.2d 222 (D.C. 1973).

amount of the policy limit with the court because the policy stated that the duty ended when the insurer "paid or tendered or deposited in court such part of such judgment as does not exceed the limit of the company's liability."⁵³ The *Delaney v. Vardine Paratransit, Inc.*⁵⁴ court held that use of the language "[o]ur payment of liability insurance limit ends our duty to defend or settle" did not release the insurer of its duty to defend when it paid the policy limits directly to the claimant who had sued the insured.⁵⁵ In *Batdorf v. Transamerica Title Insurance Co.*,⁵⁶ the court released the insurer from its duty to defend after it paid its insureds the policy limits. The *Batdorf* policy contained perhaps the most explicit language interpreted by a court so far:

The Company shall have the right to . . . defend the insured . . . reserving however, the option at any time of settling the claim or paying the amount of this policy in full.

. . . .

The Company may at any time pay this policy in full, whereupon all liability of the Company shall terminate. . . . The liability of the Company shall in no case exceed the actual loss of the insured and costs which the Company is obligated to pay.⁵⁷

All courts interpreting the *Brown* wording, which is the post-1966 language without the last six words, "by payments of judgments or settlements," have reached the conclusion the *Brown* court did: only exhaustion of policy limits by way of judgment or settlement extinguishes the insurer's duty to defend. The *Stanley v. Cobb*⁵⁸ court read the "[w]e will settle or defend" language as providing the only two options for exhaustion.⁵⁹ The courts in *Stanley*, *Samplly v. Integrity Insurance Co.*⁶⁰ and *Anderson v. United States Fidelity & Guaranty Co.*⁶¹ rejected the claim that an insurer's duty to defend ends when it pays the policy limits into court. The court in *Pareti v. Sentry Indemnity Co.*⁶² released the insurer from its duty to defend because a good-faith settlement had exhausted the policy limits.⁶³

This overview of cases shows that the issue in *Brown* concerns not one but two questions. First, does exhaustion of liability limits terminate the duty to defend? If not, the insurer is obligated to defend the insured until the very end—that is, until resolution of all claimants' actions and conclusion of all appeals. Very few courts have reached this conclusion; those that have were inter-

53. *Id.* at 224-25. Another court reached the same conclusion when an insurer using almost identical policy language tendered its policy limits into court in an interpleader action. See *Commercial Union Ins. Co. v. Adams*, 231 F. Supp. 860, 866 (S.D. Ind. 1964).

54. 132 Misc. 2d 397, 504 N.Y.S.2d 70 (N.Y. Sup. Ct. 1986).

55. *Id.* at 398, 504 N.Y.S.2d at 71.

56. 41 Wash. App. 254, 702 P.2d 1211 (1985).

57. *Id.* at 256-57, 702 P.2d at 1212-13.

58. 624 F. Supp. 536 (E.D. Tenn. 1986).

59. *Id.* at 537.

60. 476 So. 2d 79, 83-84 (Ala. 1985).

61. 177 Ga. App. 520, 522, 339 S.E.2d 660, 661 (1986).

62. 536 So. 2d 417 (La. 1988).

63. *Id.* at 424.

preting pre-1966 policy language.⁶⁴ Second, if exhaustion does terminate the duty to defend, what form of exhaustion suffices? The *Brown* court, following the majority trend, held that only exhaustion through settlement or judgment suffices.⁶⁵ Other courts have approved exhaustion by way of the insurer's tendering the policy limits into court or to the insured.⁶⁶ Notably, no court has held that LMCC's method of exhaustion, tendering the coverage limits to the claimant, terminates the duty to defend.⁶⁷

Against this backdrop it is easier to understand the ambiguity dispute in *Brown*. The majority and dissent agreed on the answer to the first question: the *Brown* policy language explicitly declares that exhaustion of liability limits terminates the duty to defend. All the justices conceded that the insurer is not obligated to pay for an appeal.⁶⁸ They disagreed, however, on the second ques-

64. *E.g.*, *Commercial Union Ins. Co. v. Pittsburgh Corning Corp.*, 553 F. Supp. 425, 430-31 (E.D. Pa. 1981); *Aetna Casualty & Sur. Co. v. Certain Underwriters at Lloyds of London*, 56 Cal. App. 3d 791, 804, 129 Cal. Rptr. 47, 55 (1976); *Travelers Indem. Co. v. East*, 240 So. 2d 277, 280 (Miss. 1970); *American Employers Ins. Co. v. Goble Aircraft Specialties, Inc.*, 205 Misc. 1066, 1073, 131 N.Y.S.2d 393, 400 (N.Y. Sup. Ct. 1954), *appeal withdrawn*, 1 A.D.2d 1008, 154 N.Y.S.2d 835 (N.Y. App. Div. 1956).

65. Although LMCC did not make the argument, an insurer might label payment to a plaintiff in partial satisfaction of his claim against the insured a "partial settlement" and contend that it has met the *Brown* requirement that exhaustion occur through judgment or settlement. A commentator from the insurance industry rejects this proposition:

The better reasoning would seem to support the view that an agreement whereby the insurer tenders its policy limits to a single claimant in exchange for a partial release of his claim against the insured is not in fact a "settlement" but merely establishes a credit against the insured's ultimate liability, and therefore a payment of policy limits pursuant to such an agreement would not terminate the duty to defend.

Van Vugt, supra note 28, at 264 (the author bases his position on a reading of the post-1966 standard liability policy, not the *Brown* policy).

66. *E.g.*, *Commercial Union Ins. Co. v. Adams*, 231 F. Supp. 860, 866 (S.D. Ind. 1964) (insurer's duty to defend ended when it tendered policy limits unconditionally into court in an interpleader action); *National Union Ins. Co. v. Phoenix Assurance Co.*, 301 A.2d 222, 225 (D.C. 1973) (duty to defend ended when insurer paid amount of policy limits into court); *Batdorf v. Transamerica Title Ins. Co.*, 41 Wash. App. 254, 258, 702 P.2d 1211, 1213 (1985) (court relieved insurer of duty to defend when it paid insureds the amount of liability coverage).

67. At least two courts have rejected this form of exhaustion. In *Conway v. Country Casualty Insurance Co.*, the insurer paid the claimant with the insured's consent, but she released neither the insurer nor the insured. 92 Ill. 2d 388, 395, 442 N.E.2d 245, 247 (1982). In rejecting the insurer's claim that payment of the policy limit terminated its duty to defend the insured, the Illinois Supreme Court relied on state law holding that the duty to defend is not dependent on the duty to indemnify, but instead arises from the insurance contract. *Id.* at 394, 442 N.E.2d at 247. Because the policy expressly provided that the insurer could terminate its duty to defend by paying the policy limits in "any judgments or settlements," and there was neither a judgment nor a settlement, the court held that the insurer did not terminate its duty to defend. *Id.* at 396, 442 N.E.2d at 248; *see supra* notes 45-47 and accompanying text.

In *Delaney v. Vardine Paratransit, Inc.*, the insurer "unilaterally, paid its policy limit of \$5,000 to plaintiffs without regard to settlement of the litigation or any further exposure of its insured." 132 Misc. 2d 397, 398, 504 N.Y.S.2d 70, 71 (N.Y. Sup. Ct. 1986). The New York Supreme Court, holding that the insurer had not terminated its duty to defend, relied on a state regulation that mandates certain provisions for automobile liability insurance policies. The regulation states: "The amounts so incurred under this subdivision, except settlement of claims and suits, shall be payable by the company in addition to the applicable policy limits." N.Y. COMP. CODES R. & REGS. tit. 11, § 60.1(b) (1985). The court held that the regulation "clearly anticipates that the insurer may not unilaterally avoid payment of defense costs and related expenses by the ex parte payment of the limits of its policy." *Delaney*, 132 Misc. 2d at 398, 504 N.Y.S.2d at 71; *see supra* note 54 and accompanying text.

68. Because the issue was not before the court, it is not clear how many justices would find that

tion. The majority found the *Brown* language ambiguous because it is unclear what form of exhaustion terminates the duty to defend.⁶⁹ The dissent claimed that there is no ambiguity: "[T]he only reasonable interpretation is that by paying its full policy limits to the party injured by its insured, defendant 'exhausted' its limit of liability and ended its duty to settle or defend."⁷⁰

The North Carolina Supreme Court is the first court to base its interpretation of the *Brown* policy language on a finding of ambiguity. Four other courts interpreting the same language have reached holdings consistent with *Brown*, but none found ambiguity. The *Stanley*⁷¹ and *Anderson*⁷² courts held that the language is not ambiguous: it clearly requires exhaustion through settlement or judgment. The *Samplly* court did not mention the ambiguity issue.⁷³ The *Pareti* court held that an insurer may terminate its defense obligations under the express terms of the policy by entering into a good-faith settlement for the policy limits.⁷⁴ The *Pareti* court found no ambiguity because there was a settlement, but had there been a tender of policy limits the same language arguably would have been ambiguous.⁷⁵

The vulnerability of the ambiguity argument is indicated to some extent by the justices' votes: although there was only one dissenter among all the judges deciding *Samplly*, *Stanley*, *Pareti*, and *Anderson*, *Brown* is a four-three decision. Ambiguity is a vulnerable basis because it is largely a question of fact, not of law. For example, the majority cited several cases holding that a difference in the parties' interpretations is a factor going toward a finding of ambiguity, although such a conflict is not dispositive.⁷⁶ "A difference of judicial opinion regarding proper construction of policy language" is also some evidence that the provision is ambiguous.⁷⁷ Arguably the potential for judges changing their minds is higher with such a fact-based inquiry than it is with a more law-based analysis.

the duty to defend ends when there is a judgment or settlement with only one of several claimants. Nor is it possible to know how many justices would find that the insurer has a continuing duty to defend the insured in a codefendant's suit for contribution or indemnification. At least one commentator declares it "well settled" that the duty to defend does not continue in these two situations. Van Vugt, *supra* note 28, at 263.

69. *Brown*, 326 N.C. at 394, 390 S.E.2d at 154.

70. *Id.* at 398, 390 S.E.2d at 156 (Whichard, J., dissenting).

71. *Stanley v. Cobb*, 624 F. Supp. 536, 537 (E.D. Tenn. 1986).

72. *Anderson v. United States Fidelity & Guar. Co.*, 177 Ga. App. 520, 521, 339 S.E.2d 660, 661 (1986).

73. *Samplly v. Integrity Ins. Co.*, 476 So. 2d 79, 81-83 (Ala. 1985).

74. *Pareti v. Sentry Indem. Co.*, 536 So. 2d 417, 423 (La. 1988). The court went on to state that the insurer must "make every effort to avoid prejudicing the insured by the timing of its withdrawal from litigation," and "make allowances for the time that the insured will need to retain new counsel, and should continue to represent the insured after the settlement, if necessary, until new counsel can be retained." *Id.*

75. *Id.* at 421 n.3.

76. *Brown*, 326 N.C. at 392, 390 S.E.2d at 153 (citing *St. Paul Fire & Marine Ins. Co. v. Freeman-White Assoc.*, 322 N.C. 77, 83, 366 S.E.2d 480, 484 (1988); *Mazza v. Medical Mut. Ins. Co.*, 311 N.C. 621, 630, 319 S.E.2d 217, 223 (1984); *Trust Co. v. Insurance Co.*, 276 N.C. 348, 354, 172 S.E.2d 518, 522 (1970)).

77. *Id.* (citing *Maddox v. Insurance Co.*, 303 N.C. 648, 654, 280 S.E.2d 907, 910 (1981); *Electric Co. v. Insurance Co.*, 229 N.C. 518, 521, 50 S.E.2d 295, 297 (1948); Annotation, *Insurance—Ambiguity—Split Court Opinions*, 4 A.L.R.4TH 1253, 1255 (1981)).

The decision would have had the same effect, yet stood on firmer ground, had the court based it on the reasonable interpretation that "[o]ur duty to settle or defend" unambiguously defines the only two methods of exhaustion.⁷⁸ This also would have given the North Carolina insurance industry clearer guidelines for changing the policy language. Such a holding would have told insurers that they need to erase "to settle or defend" from their policies. As the decision stands, it tells insurers only that the *Brown* policy language is ambiguous; it does not tell them how to remove the ambiguity to attain the effects they seek.

Although the dissent criticized the majority's claim of ambiguity, its attack is weak in that it relies on *Pareti*, *Stanley*, and *Anderson*, cases whose holdings are consistent with *Brown*. The dissent flip-flopped by turning to these cases to support the argument that the policy language is unambiguous,⁷⁹ and then distinguishing their holdings by claiming a significant factual difference.⁸⁰

The dissenters argued that an insurer necessarily exhausts policy limits by tendering the amount to a claimant, as in *Brown*, because it will not get the money back.⁸¹ In contrast, by tendering the amount into court, as in *Anderson*, *Samely*, and *Stanley*, the insurer will recover the sum if it makes the tender conditional on the insured being found liable and the insured is then found not liable.⁸² The weakness of this argument is that it considers only the claimant and ignores the insured, to whom, after all, the defense is of value. Professor Appleman asserts:

[T]he insurer's duty is both to defend actions *and* to pay judgment against the insured. Otherwise, where the damages exceed the policy coverage, the insurer could walk into court, toss the amount of the policy on the table, and blithely inform the insured that the rest was up to him. This would obviously constitute a breach of the insurer's contract to defend actions against the insured, for which premiums had been paid, and should not be tolerated by the courts.⁸³

Although it was an insurer's attempt to tender policy limits into court that prompted the *Simmons* court to declare that "a most significant protection afforded by the policy—that of defense—is rendered a near nullity,"⁸⁴ the same concern applies when the insurer pays the amount directly to the claimant.

To distinguish further the tendering of policy limits into court and to the claimant, the dissent argued that *Brown* runs counter to the North Carolina

78. This was the basis for the decisions in *Stanley* and *Anderson*. *Stanley v. Cobb*, 624 F. Supp. 536, 537 (E.D. Tenn. 1986); *Anderson v. United States Fidelity & Guar. Co.*, 177 Ga. App. 520, 521, 339 S.E.2d 660, 661 (1986). Nonetheless, the majority's finding of ambiguity is defensible. For a recitation of the court's chief argument that the method of exhaustion is ambiguous, see *infra* text accompanying note 94 (listing various methods LMCC may have intended).

79. *Brown*, 326 N.C. at 397-98, 390 S.E.2d at 156 (Whichard, J., dissenting).

80. *Id.* at 400-01, 390 S.E.2d at 158 (Whichard, J., dissenting). Opinions from other courts seem to indicate that the factual distinction does not support the dissent's argument, but cuts against it. See *supra* note 67 and accompanying text.

81. *Brown*, 326 N.C. at 401, 390 S.E.2d at 158 (Whichard, J., dissenting).

82. *Id.* (Whichard, J., dissenting).

83. 8 J. APPLEMAN, *INSURANCE LAW AND PRACTICE* § 4685, at 21-22 (1942).

84. *Simmons v. Jeffords*, 260 F. Supp. 641, 642 (E.D. Pa. 1966).

public policy that encourages insurers to make advance partial payments to claimants before final settlements.⁸⁵ Although *Brown* does not further encourage partial payments to claimants, neither does it discourage such payments, as the dissent contends. Section 1-540.3(a) still encourages insurers to make partial payments by declaring that the payments do not constitute an admission of liability by either the insured or the insurer;⁸⁶ *Brown's* only effect is that the payments will not relieve the insurer of its duty to defend the insured. Moreover, the dissent's policy argument ignores the state's countervailing public policy, which is to give effect to all parts of a contract,⁸⁷ including the "significant protection . . . of defense."⁸⁸

Another argument of the dissent is that the majority's interpretation amends the disputed provision to include words that LMCC patently omitted.⁸⁹ The *Brown* policy lacks language that the post-1966 policy contained: "[T]he company shall not be obligated to pay any claim or judgment or to defend any suit after the applicable limit of the company's liability has been exhausted by payment of judgments or settlements."⁹⁰ Because "[t]he language specifying means of exhaustion of policy limits is patently absent in the contract at issue here," Justice Whichard argued that the court should have given effect to the drafters' clear intent to leave limitless the forms of exhaustion that end the insurer's duty to defend.⁹¹ Although the argument is attractive in theory, it does not defeat the majority's finding of ambiguity. That is, whether or not the omission of "by payments of judgments or settlements" is purposeful, the court must resolve the ambiguous result against the drafter.

Even without a finding of ambiguity, the dissent's argument is unpersuasive because the language leaves the insured unwarned.⁹² It is unlikely that there are many insureds who, on reading the *Brown* proviso, contemplate the multiple avenues of escape open to the insurer:⁹³ paying the policy limits into court and

85. See N.C. GEN. STAT. § 1-540.3(a) (1983) (allowing for partial payments before settlement).

86. *Id.*

87. *Brown*, 326 N.C. at 393, 390 S.E.2d at 153 (citing *Bolton Corp. v. T.A. Loving Co.*, 317 N.C. 623, 628, 347 S.E.2d 369, 372 (1986)).

88. *Id.* at 396-97, 390 S.E.2d at 155 (quoting *Simmons v. Jeffords*, 260 F. Supp. 641, 642 (E.D. Pa. 1966)).

89. *Id.* at 399, 390 S.E.2d at 157 (Whichard, J., dissenting).

90. Van Vugt, *supra* note 28, at 257 (emphasis added) (quoting N. RISJARD & J. AUSTIN, *AUTOMOBILE LIABILITY INSURANCE CASES, STANDARD PROVISIONS AND APPENDIX* 265 app. (1964)).

91. *Brown*, 326 N.C. at 399, 390 S.E.2d at 157 (Whichard, J., dissenting). The dissent also argued that the majority's interpretation of the disputed language renders it "meaningless surplusage," because the duty to defend, if it does not end earlier, always terminates on settlement or judgment. *Id.* at 400, 390 S.E.2d at 157 (Whichard, J., dissenting). The argument fails, however, for the provision tells the insured that the insurer will not pay to appeal the case once there is a judgment. Although the *Brown* court did not decide the issue, the provision may also inform the insured that the duty to defend stops on judgment or settlement in other situations in which the insured continues to need a defense. See *supra* note 68.

92. See *Kosce v. Liberty Mut. Ins. Co.*, 159 N.J. Super. 340, 345-46, 387 A.2d 1259, 1262 (Super. Ct. Law Div. 1978) ("When construing language covering an obligation such as the duty to defend the insured, the court must look to the reasonable expectations of the insured.").

93. See *id.* at 346, 387 A.2d at 1262 ("We are dealing with language in a long, detailed insurance policy which an insured would find difficult to understand even after painstaking study.").

interpleading conflicting claimants, paying the limits to one of several claimants in exchange for a settlement of that one claim against the insured, paying the limits in full or partial satisfaction of a judgment against the insured, advancing the amount to the insured without investigating available defenses, or paying the limits to the claimant in return for a release of the insurer but not the insured.⁹⁴ When courts have approved termination of the duty to defend on unilateral tender of policy limits, they have had before them much more explicit policy language. Unlike the *Brown* policy, the policies in *Gross*,⁹⁵ *National Union*,⁹⁶ *Batdorf*,⁹⁷ and *Commercial Union*⁹⁸ did not leave the term "exhausted" undefined. Although by specifying the means of exhaustion an insurer may be faced with a form of exhaustion it did not contemplate, this route is the only fair one for insureds. To leave the forms of exhaustion limitless is to take advantage of the insured, who is not alerted to the issue without some affirmative language in the contract defining "exhaustion." Moreover, the insurer is the party in the best position to raise and clarify the issue.

Even assuming that an insured reading the policy language used in *Brown* understands that the insurer readily can escape the duty to defend, such a contract may be void as against public policy. Even those in the insurance industry concede that "an insurer cannot unilaterally pay money and walk away."⁹⁹ One industry article states:

[E]ven in jurisdictions that have . . . held that an insurer can terminate its defense obligation by paying the maximum amount of coverage specified by the policy, it has been widely accepted that the insurer cannot avoid its obligation to defend against an insured's contingent liability by early payment of the policy limits without effectuating a settlement or without obtaining the permission of the insured.¹⁰⁰

Professor Long agrees. He writes that the duty to defend should not end on exhaustion of the policy limit when the result would prejudice the insured.¹⁰¹ He argues that an insurer should not be allowed to pay the policy limit to its insured and thereby leave the insured to fend for himself, because this is an unjustified attempt by the insurer to free itself from its contractual duty to defend.¹⁰²

The duty-to-defend provision is a way to prevent an insurer from abandoning its insured by paying its policy limits into court or to the claimant, leav-

94. *Brown*, 326 N.C. at 394, 390 S.E.2d at 154. Of course, these options are not available to the insurer if a court declares them against public policy or if the insurer discharges them in bad faith.

95. *Gross v. Lloyds of London Ins. Co.*, 121 Wis. 2d 78, 83, 358 N.W.2d 266, 269 (1984).

96. *National Union Ins. Co. v. Phoenix Assurance Co.*, 301 A.2d 222, 224-25 (D.C. 1973).

97. *Batdorf v. Transamerica Title Ins. Co.*, 41 Wash. App. 254, 256-57, 702 P.2d 1211, 1212-13 (1985).

98. *Commercial Union Ins. Co. v. Adams*, 231 F. Supp. 860, 865 (S.D. Ind. 1964); see *supra* note 53.

99. Zulkey, *When Can and How Does an Insurer Properly Exhaust Its Limits?*, 56 DEF. COUNS. J. 202, 205 (1989).

100. Zulkey & Pollard, *supra* note 2, at 23.

101. R. LONG, *THE LAW OF LIABILITY INSURANCE* § 5-25, at 5-179 (1990).

102. *Id.* § 5.26, at 5-184.

ing the insured with litigation and possible appeals.¹⁰³ If an insurer can do this by specifically declaring in the policy its right to this option, the insurer is merely putting the insured on notice that he is obtaining a valueless contractual right.¹⁰⁴

If an insurer wishes to escape its duty to defend when it believes liability will exceed the policy limit, it should make the contract language unequivocal. The following suggestion puts the insured on notice by spelling out the forms of exhaustion without limiting them:

Our duty to settle or defend ends when our liability for this coverage has been exhausted. Coverage may be exhausted in any manner, including but not limited to the following methods of payment by the Company: to the insured if she consents; to a claimant or claimants in full or partial satisfaction of a settlement of the claimant's or claimants' actions against the insured; to a claimant or claimants in full or partial satisfaction of a judgment against the insured; to a claimant or claimants in return for a release of the Company, whether or not the insured is also released; to the court when there are conflicting claimants and the Company wishes to interplead.

Unlike the *Brown* policy, this language is not ambiguous. Moreover, it notifies the insured that the duty to defend can terminate quickly. Although this Note argues that such a provision is void as against public policy, only the courts can make such a decision.¹⁰⁵ Certainly an insurer can defend its contract more forcefully when it has used explicit language and can argue that it has acted in good faith¹⁰⁶ by making every effort to avoid prejudicing the rights of its insured. Because the LMCC policy language was not explicit, leaving the Browns without warning that LMCC's duty to defend them could end so quickly, the *Brown* court correctly decided to construe the ambiguous language against the defendant. The only mistake the court made was resting its opinion on so vulnerable a basis as ambiguity.

CHRISTINE J. WICHERS

103. 7C J. APPLEMAN, *supra* note 83, § 4682, at 36 (1979).

104. It was not the intent of the drafters of the 1966 and earlier policies to give the insured a valueless right. According to several representatives of the insurance industry, the purpose of the 1966 addition of "by payments of judgments or settlements" was merely to clarify the defense coverage clause, not to change its substance. Van Vugt, *supra* note 28, at 263. Therefore, "it could be persuasively argued that under all forms of the standard liability policy exhaustion of policy limits can occur only by payment of judgments or settlements." *Id.* If this is indeed the intent of the drafters, then they always have meant for the insured to obtain something of value. Of course, the argument on the other side is what the dissent in *Brown* asserted: that the drafters of the *Brown* policy, which omits the language added in 1966, did not intend to limit exhaustion to settlement or judgment. *Brown*, 326 N.C. at 399, 390 S.E.2d at 157 (Whichard, J., dissenting).

105. Moreover, there is significant support for the position that when the contract makes the option clear, payment of coverage limits into court is not against public policy. Several courts have upheld this form of exhaustion. *E.g.*, Commercial Union Ins. Co. v. Adams, 231 F. Supp. 860, 866 (S.D. Ind. 1964); National Union Ins. Co. v. Phoenix Assurance Co., 301 A.2d 222, 225 (D.C. 1973); Batdorf v. Transamerica Title Ins. Co., 41 Wash. App. 254, 702 P.2d 1211, 1213 (1985).

106. For an argument that insurers can exhaust policy limits in all cases with multiple claimants by interpleading, so long as they do so in good faith, see Zulkey, *supra* note 99, at 204-05.

Divorcing Public Policy from Economic Reality: The Fourth Circuit's Copyright Misuse Doctrine in *Lasercomb America, Inc. v. Reynolds*

For much of this century the United States Supreme Court has considered the grant of a patent¹ to be a public benefit that its owner may not misuse. Doing so deprives the patentee of the right to enforce an infringement claim.² A debate exists among lower federal courts and commentators, however, regarding the type of conduct sufficiently abusive to constitute misuse. Must the patentee use the patent to violate the antitrust laws?³ Or must the patentee merely extend the scope of the patent beyond the statute's permissible bounds, an act the courts consider contrary to public policy?⁴ In like fashion, lower courts and commen-

1. A patent is a statutory grant that offers the patentee the exclusive right to make, use, or sell his invention for 17 years. 35 U.S.C. § 154 (1988). Under the Patent Act, the law considers the patent to be personal property. *Id.* § 261. The grant of a patent primarily benefits society: by offering the patentee a limited monopoly in exchange for an early disclosure of his invention, society gains the benefit of knowledge that might otherwise be shrouded from view as a trade secret. See *SCM Corp. v. Xerox Corp.*, 645 F.2d 1195, 1203 (2d Cir. 1981); 8 E. LIPSCOMB, LIPSCOMB'S WALKER ON PATENTS § 28:2 (3d ed. 1989) [hereinafter WALKER ON PATENTS]. For further consideration of the patent law's procompetitive policy, see *infra* note 149.

Inventors seeking to patent their inventions must meet several statutory requirements. The threshold requirement is that the invention be useful. 35 U.S.C. § 101 (1988). The patent laws also require "novelty"—the inventor must be the first to invent or discover the device. *Id.* § 102(g). The law charges the inventor with constructive knowledge of all prior references to relevant technological developments. *Id.* § 102(a). If another discovery "anticipated" the invention, no patent will issue. Thus, to be patentable the invention must be sufficiently new and different from what came before it; or in the argot of the Patent Act, the invention must be "nonobvious." *Id.* § 103. The Patent Office requires an applicant to demonstrate that her invention meets all requirements and carefully scrutinizes each invention before issuing a patent. The Patent Act also establishes procedures for the mediation of competing claims. See *id.* §§ 131-35, 141-46. For the statutory requirements for a valid copyright see *infra* note 6.

2. *Morton Salt Co. v. G.S. Suppiger Co.*, 314 U.S. 488, 494 (1942); *Carbice Corp. v. American Patents Dev. Corp.*, 283 U.S. 27, 33 (1931); *Motion Picture Patents Co. v. Universal Film Mfg. Co.*, 243 U.S. 502, 518-19 (1917).

3. See *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 511 (7th Cir. 1982) ("Since the antitrust laws as currently interpreted reach every practice that could impair competition substantially, it is not easy to define a separate role for a doctrine also designed to prevent an anticompetitive practice—the abuse of a patent monopoly."); *G.S. Suppiger Co. v. Morton Salt Co.*, 117 F.2d 968, 970 (7th Cir. 1941) ("[A] patentee is not handicapped in his commercial transactions because he owns a patent. What he could lawfully do before he acquired the patent, he may lawfully do after acquirement."), *rev'd*, 314 U.S. 488 (1942); see also 4 D. CHISUM, PATENTS § 19.04[2] (1990) ("[T]he misuse doctrine presents greater analytical problems than antitrust analysis . . . [which] involves a balancing of patent interests and the impact or likely impact of a practice on competition. The misuse doctrine compounds the difficulty of balancing by substituting for competitive injury the vague notion of 'extension.'").

The Patent Act recognizes patent misuse, but only a narrow field of patentee conduct rises to the level of an antitrust violation to constitute misuse. See 35 U.S.C. § 271(d)(5) (1988), quoted *infra* at note 76. In the converse situation, where a patentee is charged with violating the antitrust laws, the Patent Act states that nothing in the patent laws "shall be deemed to convey to any person immunity from civil or criminal liability, or to create any defenses to actions, under any antitrust law." *Id.* § 211.

4. *Morton Salt*, 314 U.S. at 494; *Panther Pumps & Equip. Co. v. Hydrocraft, Inc.*, 468 F.2d 225, 231 (7th Cir. 1972) ("The policy underlying the misuse doctrine is designed to prevent a patentee from projecting the economic effect of his admittedly valid grant beyond the limits of his legal monopoly."); see Calkins, *Patent Law: The Impact of the 1988 Patent Misuse Reform Act and Noerr-Pennington Doctrine on Misuse Defenses and Antitrust Counterclaims*, 38 DRAKE L. REV. 175, 178-

tators⁵ have differed over whether these same principles should apply to another limited monopoly, the copyright.⁶ The Supreme Court has implied that copyright misuse might be a valid defense,⁷ but it has not suggested what types of misbehavior will bar an infringement claim.

In *Lasercomb America, Inc. v. Reynolds*⁸ the United States Court of Appeals for the Fourth Circuit became the first appellate court expressly to sanction copyright misuse as an infringement defense and to hold that such misuse barred the plaintiff's action.⁹ The Fourth Circuit refused to import antitrust requirements into the misuse doctrine; instead, the copyright holder's anticompetitive behavior towards a third party was deemed sufficiently contrary to pub-

92 (1988). One North Carolina jurist conjured up this colorful simile to describe patent misuse: "Like Aesop's dog, he seeks to take in too much, and must lose what he already had." *Dubuit v. Harwell Enters.*, 336 F. Supp. 1184, 1187 (W.D.N.C. 1971) (McMillan, J.).

5. See, e.g., *Fine, Misuse and Antitrust Defenses to Copyright Infringement Actions*, 17 *HASTINGS L.J.* 315, 330-36 (1965); Gibbs, *Copyright Misuse: Thirty Years Waiting for the Other Shoe*, 23 *COPYRIGHT L. SYMP. (ASCAP)* 31, 37-57 (1977); Note, *Redefining Copyright Misuse*, 81 *COLUM. L. REV.* 1291, 1310-14 (1981).

6. Like patents, copyrights offer rewards to individuals as an inducement to generate an increased number of unique expressions that benefit the general public. *Sony Corp. v. Universal City Studios*, 464 U.S. 417, 431-32 (1984). Despite the theoretical similarities, however, copyrights differ in practical ways from patents.

A copyright is a limited monopoly granted to a work satisfying two criteria: The work must be "original" in that it is the independent effort of the author seeking the copyright, and the work must be fixed in a tangible medium. 17 U.S.C. § 102(a) (1988). For works created on or after January 1, 1978 the copyright endures for the life of the author and fifty years after the author's death. *Id.* § 302(a). The copyright grants certain exclusive rights in the protected work, including the right to reproduce the work, prepare derivative works, distribute copies to the public (by sale, rental, lease, or lending), and perform or display the work publicly. *Id.* § 106. Protection from unauthorized copying extends only to the particular expressions originating with the author; it does not protect the ideas that are part of the copyrighted work. *Mazer v. Stein*, 347 U.S. 201, 217 (1954); 17 U.S.C. § 102(b).

There are two significant exceptions to a copyright grantee's claim of exclusivity. Unlike patents, the grantee need not be the first to create the unique expression. The alleged infringer's independent creation of an identical work is a defense to an infringement action. *Mazer*, 347 U.S. at 218. In addition, the Copyright Act provides that the unauthorized copying done pursuant to a "fair use," typically for nonprofit educational purposes, is not an infringement. 17 U.S.C. § 107 (1988). No analogous provision exists in patent law.

Finally, the standards for securing a copyright are far less severe than are patent standards, and copyrights are not subject to rigorous evaluations. For a useful comparison of the levels of protection afforded by patents and copyrights, see *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 100-04 (2d Cir. 1951). See *supra* note 1 for a discussion of the practical and theoretical underpinnings of a patent.

7. *Broadcast Music, Inc. v. Columbia Broadcasting Sys.*, 441 U.S. 1, 6 & n.9 (1979). For a discussion of *CBS* see *infra* notes 109-16 and accompanying text.

8. 911 F.2d 970 (4th Cir. 1990). *Lasercomb* chose not to petition for certiorari because it felt the added cost of litigation was not justified because "Interact," *Lasercomb's* copyrighted software program, had become obsolete and because *Lasercomb* "purged" its misuse by discontinuing the use of the noncompete clause in its standard licensing agreement. Telephone interview with Lee Bromberg, counsel for *Lasercomb America, Inc.* (Jan. 21, 1991). For a discussion of how a grantee may purge its misuse see *infra* notes 62-64 and accompanying text.

9. *Lasercomb*, 911 F.2d at 979. In 1948 one district court did deny the plaintiff-copyright holder's action to enjoin infringement on the grounds that the plaintiff used its copyright to gain monopolistic domination over musical compositions and therefore exceeded the privileges protected under the copyright laws. *M. Witmark & Sons v. Jensen*, 80 F. Supp. 843, 849-50 (D. Minn. 1948), *appeal dismissed sub nom. M. Witmark & Sons v. Berger Amusement Co.*, 177 F.2d 515 (8th Cir. 1949). For a discussion of *Jensen* see *infra* notes 84-93 & 166-69 and accompanying texts.

lic policy to deny relief.¹⁰

This Note traces the development of equitable defenses in the field of limited monopolies, from the origin of the "unclean hands" doctrine through the parallel evolutions of patent and copyright misuse. It explores the underlying policy justifications for granting limited monopolies and compares the varying amounts of protection afforded to grantees. The Note examines the role that antitrust principles have played in the misuse doctrine and concludes that the Fourth Circuit incorrectly rejected those principles as a foundation for copyright misuse. The Note maintains that the Fourth Circuit's embrace of the "public policy" rationale creates a standard too ambiguous to guide those who seek to exploit their property rights without losing the protection of the law.

Lasercomb America, Inc. created a computer software program called "Interact," which companies use to design and manufacture steel rule dies for the paper box and carton industries.¹¹ Lasercomb also manufactures steel rule dies, as did its competitor, Holiday Steel Rule Die Corporation.¹² In 1983, before it was ready to put its software on the market, Lasercomb licensed four prerelease copies of Interact to Holiday Steel under an oral agreement.¹³ Each Interact copy contained a "chronoguard," an antitheft device that allowed Interact to work only on those machines to which it was attached.¹⁴ Larry Holliday, Holiday Steel's president and sole shareholder, ordered his computer programmer, Job Reynolds, to remove the chronoguards and to make three copies of Interact.¹⁵ Holiday Steel then began marketing a CAD/CAM program called PDS-1000, which essentially was Interact under a new name.¹⁶

Once Lasercomb discovered what Holiday Steel had done, it registered the copyright¹⁷ for Interact and sued for copyright infringement.¹⁸ Defendants

10. *Lasercomb*, 911 F.2d at 979. For a discussion of violations of the antitrust laws, see *infra* note 113.

11. *Lasercomb*, 911 F.2d at 971. Companies use Interact for CAD/CAM (computer assisted design and computer assisted manufacture). *Id.* at n.2. Interact permits a box designer to create a "box blank"—a flat paperboard sheet that is cut so that it may be folded into a box. An Interact-controlled computer stores information about the prototype box blank, and a die maker can use that information to control the manufacture of a steel rule die. The die then stamps out quantities of box blanks. Opening Brief for Appellant at 4-5, *Lasercomb* (No. 89-3245).

12. *Lasercomb*, 911 F.2d at 971.

13. Although Lasercomb had asked Holiday Steel to sign its standard licensing agreement, Holiday Steel never signed, and Lasercomb never demanded compliance. *Id.* at 973. Later, however, Holiday admitted in a letter that an oral agreement with Lasercomb bound Holiday Steel. *Id.* at 973 n.7. Lasercomb charged Holiday Steel \$35,000 for the first copy, \$17,500 for the next two, and \$2,000 for the fourth copy. *Id.* at 971.

14. Brief of Plaintiff-Appellee at 14, *Lasercomb* (No. 89-3245).

15. *Lasercomb*, 911 F.2d at 971.

16. *Id.*

17. Registering the work with the Copyright Office is a prerequisite for filing an infringement action. 17 U.S.C. § 411 (1988). For works published before March 1, 1989 the author has five years from the date of publication to register the work to preserve the copyright. *Id.* § 405(a). The Copyright Act protects computer software as a "literary work," defined as "a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result." *Id.* § 101. The Act limits the grantee's exclusive rights by permitting certain noninfringing uses. For example, the owner of a copy of a computer program may make another copy freely, if the new copy "is created as an essential step in the utilization of the computer program in conjunction with a machine" or "is for archival purposes only." *Id.* § 117.

Holiday Steel, Holliday, and Reynolds claimed that Lasercomb misused its copyright because of anticompetitive clauses in Lasercomb's standard licensing contract, clauses that restricted a licensee from creating its own computer-assisted die-making software.¹⁹ The district court rejected the copyright misuse defense for three reasons. First, although another Interact licensee was bound by the anticompetitive clauses, the defendants had not explicitly agreed to them, and therefore had not suffered from the misuse.²⁰ Second, the clause was "reasonable in light of the delicate and sensitive area of computer software."²¹ Finally, the court doubted whether such a defense existed as a matter of law.²² The trial court found the three defendants jointly and severally liable for actual damages on both the copyright infringement and fraud claims, awarded punitive damages against Holliday and Reynolds on the fraud claim, and permanently enjoined all defendants from making or selling PDS-1000.²³

The United States Court of Appeals for the Fourth Circuit reversed the injunction and the damage award for copyright infringement,²⁴ affirmed the trial court's finding of fraud,²⁵ and remanded the case to the district court for recalculation of the damages.²⁶ In doing so, the Fourth Circuit acknowledged that

18. *Lasercomb*, 911 F.2d at 972. Lasercomb also claimed breach of contract and fraud. *Id.* The trial court dismissed Lasercomb's claims for misappropriation of trade secret, false designation of origin, and unfair competition because the Copyright Act preempted them. *Id.*

19. *Id.* The clauses at issue in the Lasercomb licensing agreement read:

- D. Licensee agrees during the term of this Agreement that it will not permit or suffer its directors, officers and employees, directly or indirectly, to write, develop, produce or sell computer assisted die making software.
- E. Licensee agrees during the term of this Agreement and for one (1) year after the termination of this Agreement, that it will not write, develop, produce or sell or assist others in the writing, developing, producing or selling computer assisted die making software, directly or indirectly without Lasercomb's prior written consent. Any such activity undertaken without Lasercomb's written consent shall nullify any warranties or agreements of Lasercomb set forth herein.

Id. at 973. The term of the licensing agreement was for 99 years. *Id.*

20. *Id.* Holiday Steel never signed the standard licensing agreement but was bound under an oral contract. *See supra* note 13.

21. *Lasercomb*, 911 F.2d at 973 (quoting the trial court's findings).

22. *Id.*

23. *Id.* at 972. The trial court found that "the evidence at trial clearly disclosed not only defendants' liability as copyright infringers, it also indicated their culpability . . . [as] malicious infringers who acted with full knowledge of and gross disregard to plaintiff's right." Brief of Plaintiff-Appellee at 5-6 (quoting trial court's findings) (emphasis in Brief). Following the judgment against it, Holiday Steel went bankrupt and the bankrupt estate stipulated it would not appeal. *Id.* at 3 n.2.

24. *Lasercomb*, 911 F.2d at 979.

25. *Id.* at 980.

26. *Id.* The trial court found that Lasercomb had suffered damages of \$105,000 from the infringement and another \$105,000 from the fraud. The damages arose, however, from lost sales of the same copies of Interact; Lasercomb, therefore, could not recover twice. *Id.* at 972 n.3. The court based the \$105,000 damages amount on Interact's sale price of \$35,000 to first-time buyers, multiplied by three—the number of unauthorized copies defendants made of Interact. *Id.* at 980. The Fourth Circuit, in contrast, noted that it could only measure damages by lost sales to Holiday Steel, which was receiving a discount for additional purchases of Interact. *Id.* at 981. It was up to the district court on remand to decide whether the defendants' fraudulent activities induced Lasercomb to discount its price. *Id.*

the copyright misuse defense was far from being a settled area of the law,²⁷ but considered the defense to be "inherent in the law of copyright."²⁸

The touchstone of the court's analysis was the body of patent law that produced the patent misuse defense and the parallel public policies behind patent and copyright law. After discussing the seventeenth century English Crown's practice of granting limited monopolies,²⁹ the court found that the Federal Constitution's grouping of patent and copyright into the same clause reflected "a unitary purpose—to promote progress."³⁰ The court then examined a seminal patent misuse case, *Morton Salt Co. v. G.S. Suppiger Co.*,³¹ and found that its "phraseology adapts easily to the copyright context."³² The result of that easy adaptation is that public policy excludes from protection all that is not embraced by the copyright and forbids the use of a copyright to secure rights not granted by the Copyright Office.³³

The *Lasercomb* panel next considered whether copyright misuse depends on antitrust principles. The trial court deemed *Lasercomb*'s noncompete clause to be "reasonable," a notion that the Fourth Circuit assumed was derived from the antitrust concept of the "rule of reason."³⁴ While stating that an antitrust

27. For a discussion of the Fourth Circuit's discomfort with the apparent lack of certainty surrounding the copyright misuse doctrine, see *infra* note 144.

28. *Lasercomb*, 911 F.2d at 973.

29. *Id.* at 974.

30. *Id.* at 975. The Constitution provides that "[t]he Congress shall have Power . . . [t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. CONST. art. I, § 8, cl. 8.

The Supreme Court has noted the "historic kinship between patent law and copyright law," but also has warned that "[t]he two areas of the law, naturally, are not identical twins, and we exercise the caution which we have expressed in the past in applying doctrine formulated in one area to the other." *Sony Corp. v. Universal City Studios*, 464 U.S. 417, 439 & n.19. (1984). Despite that warning, however, the Court in *Sony Corp.* readily applied the doctrine of contributory infringement to copyrights. The Court noted that the Patent Act expressly defines this concept but the Copyright Act does not. *Id.* at 434-35. Such a doctrine, the Court concluded, was necessary to provide "effective . . . protection of the statutory monopoly" when a defendant sells copying equipment that is capable only of substantial infringing uses. *Id.* at 442. The Court found that, although the general public was using video recorders sold by Sony to tape the plaintiffs' copyrighted motion pictures, Sony was not liable for contributory infringement because its video recorders were also capable of substantial noninfringing uses. *Id.* at 456.

31. 314 U.S. 488 (1942). The Fourth Circuit noted that although *Morton Salt* was not the first Supreme Court case to recognize the patent misuse defense, it is generally considered the "foundational" case. *Lasercomb*, 911 F.2d at 976. For a discussion of *Morton Salt*, see *infra* notes 49-68 and accompanying text.

32. *Lasercomb*, 911 F.2d at 977. The Fourth Circuit quoted a passage from *Morton Salt* and replaced patent terms and phrases with those applicable to copyright. The Fourth Circuit thus found:

"The grant to the [author] of the special privilege of a [copyright] carries out a public policy adopted by the Constitution and laws of the United States, 'to promote the Progress of Science and useful Arts, by securing for limited Times to [Authors] . . . the exclusive Right . . . to their ["original"] works'. United States Constitution, Art. I, § 8, cl. 8, [17 U.S.C.A. § 102]. But the public policy which includes [original works] within the granted monopoly excludes from it all that is not embraced in the [original expression]. It equally forbids the use of the [copyright] to secure an exclusive right or limited monopoly not granted by the [Copyright] Office and which it is contrary to public policy to grant."

Id. (quoting *Morton Salt*, 314 U.S. at 492) (bracketed material added by the *Lasercomb* court).

33. *Id.* See *supra* note 6 for a discussion of the rights protected under the copyright laws.

34. *Lasercomb*, 911 F.2d at 977. Although the Sherman Act forbids "every contract . . . in restraint of trade," 15 U.S.C. § 1 (1988), courts interpret this language to mean that only those

violation would trigger the copyright misuse defense, the Fourth Circuit held that an antitrust violation is not a prerequisite to barring an infringement action. Instead, any attempt to use a copyright "to control competition in an area outside the copyright, *i.e.* the idea of computer-assisted die manufacture," would give rise to misuse "regardless of whether such conduct amounts to an antitrust violation."³⁵ The Fourth Circuit reached its conclusion because *Morton Salt* did not require an antitrust violation to apply the patent misuse defense.³⁶ In rejecting the copyright misuse defense, the district court had emphasized that the defendants had not explicitly agreed to the noncompete clause, and therefore could not claim injury from the alleged misuse in order to bar the infringement action against them.³⁷ The Fourth Circuit, however, relied on the venerable *Morton Salt* decision to conclude that the copyright misuse defense is applicable even though the infringing parties did not themselves suffer injuries from the misuse.³⁸

Copyright misuse cases such as *Lasercomb* are the culmination of complex, intertwined conceptions of unclean hands, rights under the patent and copyright laws, and antitrust principles. Some cases rest the copyright misuse defense solely on antitrust principles,³⁹ while others deny that an antitrust violation can provide an affirmative defense to infringement.⁴⁰ Still others, like *Lasercomb*, find that an antitrust violation, while not a prerequisite, may trigger a misuse

contracts that unreasonably restrain trade are illegal. W. HOLMES, INTELLECTUAL PROPERTY AND ANTITRUST LAW § 5.01[2] (1990). See *infra* note 113 for a discussion of the antitrust rule of reason. For a discussion of the noncompete clause's purported reasonableness, see *infra* note 175.

35. *Lasercomb*, 911 F.2d at 979. The *Lasercomb* panel described the noncompete clause as containing language that was "extremely broad," and found it to be "at least as egregious" as similar language in a patent licensing agreement that the Fourth Circuit had declared to be a patent misuse. *Id.* at 978-79 (citing *Compton v. Metal Prod., Inc.*, 453 F.2d 38, 45 (4th Cir. 1971), *cert. denied*, 406 U.S. 968 (1972)). For a discussion of *Compton* and other cases involving noncompete clauses, see *infra* note 175.

Although the Fourth Circuit found that the noncompete clause constituted copyright misuse, it rejected the defendants' claim that *Lasercomb* created an illegal "tying" arrangement by discounting Interact's price for those licensees who also purchased steel rule dies from *Lasercomb*. *Lasercomb*, 911 F.2d at 972 n.6. Because *Lasercomb* did not require a buyer to purchase Interact or steel rule dies when buying the other product, or require its customers to agree not to purchase goods from other sellers, the court found no tie-in. *Id.* A tying arrangement is one in which a seller requires that a purchaser buy one item (the tied product) to buy another (the tying product). See *Northern Pac. Ry. v. United States*, 356 U.S. 1, 5-6 & n.4 (1958).

36. *Lasercomb*, 911 F.2d at 978. For a discussion of the unified economic theory behind patent and antitrust law, see *infra* note 149.

37. See *Lasercomb*, 911 F.2d at 979.

38. *Id.*

39. *Saturday Evening Post Co. v. Rumbleseat Press, Inc.*, 816 F.2d 1191, 1200 (7th Cir. 1987); see also *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 106 (2d Cir. 1951) (balancing the equities between harm caused by copyright owner's alleged antitrust violation and infringer's culpability).

40. See, e.g., *BellSouth Advertising & Publishing Corp. v. Donnelley Information Publishing, Inc.*, 719 F. Supp. 1551, 1563 (S.D. Fla. 1988) (denying antitrust defense to copyright infringement but recognizing misuse defense when there is an attempt to extend the limited monopoly beyond the protected work); *Rural Tel. Serv. Co. v. Feist Publications, Inc.*, 663 F. Supp. 214, 220 (D. Kan. 1987); *Ortho-O-Vision, Inc. v. Home Box Office*, 474 F. Supp. 672, 686 (S.D.N.Y. 1979); *Broadcast Music, Inc. v. Grant's Cabin*, 204 U.S.P.Q. (BNA) 633, 634-45 (E.D. Mo. 1979); *M. Witmark & Sons v. Pastime Amusement Co.*, 298 F. 470, 480 (E.D.S.C.), *aff'd mem.*, 2 F.2d 1020 (4th Cir. 1924).

defense.⁴¹

The animating principle behind denying relief from infringement is equitable in nature; a court will not render assistance to a party coming into court with unclean hands.⁴² The unclean hands doctrine bars legal or equitable relief to a plaintiff who has harmed the defendant in a way that bears directly on the matter in controversy.⁴³ In *Keystone Driller Co. v. General Excavator Co.*⁴⁴ the Supreme Court denied relief in a patent infringement case because the patentee, in an earlier case, had suppressed evidence that would have shown his patent to be invalid.⁴⁵ The Court stated that for the doctrine to apply there must be an immediate and necessary relationship between the plaintiff's unconscionable act and the relief sought.⁴⁶ Courts have also applied the unclean hands doctrine in copyright infringement actions. One early court, for example, applied the doctrine when the copyright holder's advertisements were false and misleading, thereby causing a public injury egregious enough to bar relief.⁴⁷ Subsequent copyright cases disavowed this broad application and confined the equitable defense to "misuses that frustrate the particular purposes of the copyright . . . statute."⁴⁸

In *Morton Salt Co. v. G.S. Suppiger Co.* the Supreme Court modified the unclean hands doctrine, permitting a patent infringer to assert the patent misuse defense even though the patentee's misuse did not harm the infringer.⁴⁹ The Court previously had recognized the patent misuse defense only when the infringing party suffered injury by the patentee's attempt to enlarge the scope of

41. See *Lasercomb*, 911 F.2d at 978; *Broadcast Music, Inc. v. Moor-Law, Inc.*, 203 U.S.P.Q. (BNA) 487, 488 (D. Del. 1978); *M. Witmark & Sons v. Jensen*, 80 F. Supp. 843, 850 (D. Minn. 1948) ("One who unlawfully exceeds his copyright monopoly and violates anti-trust laws is not outside the pale of the law, but where the Court's aid is requested . . . it should be denied."), *appeal dismissed sub nom. M. Witmark & Sons v. Berger Amusement Co.*, 177 F.2d 515 (8th Cir. 1949).

42. *Precision Instrument Mfg. Co. v. Automotive Maintenance Mach. Co.*, 324 U.S. 806, 814 (1945); *Mitchell Bros. Film Group v. Cinema Adult Theatre*, 604 F.2d 852, 863 (5th Cir. 1979); *Jensen*, 80 F. Supp. at 850.

43. *Mitchell Bros.*, 604 F.2d at 863; see also *Tempo Music, Inc. v. Myers*, 407 F.2d 503, 507 n.8 (4th Cir. 1969) (equitable estoppel barring plaintiff the right to plead copyright infringement applicable both in law and equity); *F.E.L. Publications, Ltd. v. Catholic Bishop*, 506 F. Supp. 1127, 1137 (N.D. Ill. 1981) (unclean hands applies in equity suits as well as in actions for damages), *rev'd on other grounds*, 214 U.S.P.Q. (BNA) 409 (7th Cir. 1982).

44. 290 U.S. 240 (1933).

45. *Id.* at 246-47.

46. *Id.* at 245.

47. *Stone & McCarrick, Inc. v. Dugan Piano Co.*, 220 F. 837, 841-43 (5th Cir. 1915).

48. See, e.g., *Mitchell Bros. Film Group v. Cinema Adult Theater*, 604 F.2d 852, 864 (5th Cir. 1979). The Fifth Circuit, largely repudiating its earlier decision in *Stone & McCarrick*, refused to permit the defendant to assert a "public injury" defense based on the claim that the infringed work was obscene. *Id.* at 865. In limiting *Stone & McCarrick*, the Fifth Circuit contrasted that case with *Morton Salt v. G.S. Suppiger Co.*, 314 U.S. 488 (1942), finding that *Morton Salt* properly required clean hands when the patentee subverted the public policy behind the patent statute. *Id.* at 864.

It is interesting to note, however, that *Morton Salt* itself relied on *Stone & McCarrick* to deny the patentee's infringement claim. *Morton Salt*, 314 U.S. at 494. Seemingly unconcerned about the prior requirement that there be an "immediate and necessary" relationship, the Court in *Morton Salt* justified barring the patentee's claim, even though the infringer had not suffered from the patent misuse, because of "the adverse effect upon the public interest." *Id.* (emphasis added). For a discussion of the Supreme Court's departure from traditional equitable doctrine when confronted with enforcing patent claims, see *infra* note 56 and accompanying text.

49. *Morton Salt*, 314 U.S. at 494.

the patent privileges.⁵⁰ In *Morton Salt* the plaintiff owned the patent for a canning industry device that inserted measured amounts of salt in tablet form into canned goods.⁵¹ The plaintiff leased its machines but required that licensees purchase salt tablets exclusively from the plaintiff's subsidiary.⁵² The Court denied injunctive relief from the defendant's patent infringement on the grounds that the patentee was using its patent to restrain competition in the market for salt tablets, an unpatented item, and that continued restraint would permit the patentee to create a limited monopoly not within the scope of the patent grant.⁵³

Although the infringer also competed with the patentee in the salt tablet market,⁵⁴ the Court considered injury to the public to be of primary importance.⁵⁵ The Court recognized that the unclean hands doctrine typically balances the relative misconduct of the parties to the suit, but found that "additional considerations must be taken into account where maintenance of the suit concerns the public interest as well as the private interests of suitors."⁵⁶ Because the patentee derives privileges from a statutory grant enacted to further a public policy, the Court would not protect a patent being used to subvert that policy.⁵⁷ The policy has two components: protecting inventions included in the patent while excluding from the patent all that is not embraced by the invention,⁵⁸ and forbidding a patentee from using its patent to secure an exclusive privilege that the Patent Office did not grant.⁵⁹ To the Court, a successful infringement claim would adversely affect the public interest.⁶⁰ That adverse impact, in combination with the misuse, disqualified the patentee from protection regardless of whether the infringer suffered from the patent misuse.⁶¹

However harsh the patent misuse doctrine may appear, the Court noted that the misuse could be "purged" by abandoning the improper conduct and dissipating the harmful consequences of the misuse.⁶² The patent misuse doc-

50. See *Carbice Corp. v. American Patents Dev. Corp.*, 283 U.S. 27, 31-35 (1931) (plaintiff-patentee denied relief against supplier who sold dry ice to plaintiff's licensees with the knowledge that the license required dry ice to be purchased only from patentee); *Motion Picture Patents Co. v. Universal Film Mfg. Co.*, 243 U.S. 502, 518-19 (1917) (owner of a patented film projector barred from enforcing license restriction against film distributors when patent license required that only films rented from the patentee's licensee could be used with the patented projector).

51. *Morton Salt*, 314 U.S. at 489.

52. *Id.* at 490-91.

53. *Id.* at 491.

54. *Id.*

55. *Id.* at 493.

56. *Id.* The Supreme Court later rationalized this change from the traditional unclean hands doctrine through the fiction of joining the public as a party to the suit. *Precision Instrument Mfg. Co. v. Automotive Maintenance Mach. Co.*, 324 U.S. 806, 816 (1945). Because of the "far-reaching social and economic consequences of a patent," the public has a "paramount interest in seeing that patent monopolies spring from backgrounds free from fraud or other inequitable conduct and that such monopolies are kept within their legitimate scope." *Id.*

57. *Morton Salt*, 314 U.S. at 494.

58. *Id.* at 492.

59. *Id.*

60. *Id.* at 494.

61. *Id.*

62. *Id.* at 493. See generally 4 D. CHISUM, *supra* note 3, § 19.04[4] (discussion of purging and dissipating patent misuse).

trine therefore does not invalidate a patent; the patentee may bring an action for an infringement occurring after the misuse is satisfactorily purged.⁶³ In fact, a patentee may purge the misuse even after commencing the infringement claim.⁶⁴

In reversing the decision of the Court of Appeals for the Seventh Circuit, the *Morton Salt* Court found it unnecessary to decide whether the patentee's tying arrangement⁶⁵ violated antitrust laws under the Clayton Act.⁶⁶ In the Seventh Circuit's view, acquiring a patent does not impose a higher standard of commercial behavior on patentees because the antitrust laws were not designed to "discriminate between patentees and others when it comes to use or sale of unpatented articles."⁶⁷ The Supreme Court, however, found that an inquiry into whether the patentee substantially decreased competition or tended to create a monopoly missed the point. Although the patent misuse doctrine is concerned with a patentee's anticompetitive conduct, the proper analysis must be grounded in the patent laws and the public policy behind them.⁶⁸

Morton Salt did not predicate patent misuse on antitrust principles; instead, the Court intimated that antitrust violations may serve as the basis for patent misuse.⁶⁹ Subsequently, however, the Court and commentators⁷⁰ have muddled the distinction by linking the two concepts. In *Mercoird Corp. v. Minneapolis-Honeywell Regulator Co.*⁷¹ the Court opined that "[t]he legality of any attempt

63. See *Hartford-Empire Co. v. United States*, 323 U.S. 386, 419-20 (1945); *Westinghouse Elec. Corp. v. Bulldog Elec. Prod. Co.*, 179 F.2d 139, 145 (4th Cir. 1950); see also WALKER ON PATENTS, *supra* note 1, § 28:35 ("[I]mproper use does not invest an infringer with a license for the life of the patent, for if a patent owner purges himself of the improper business practices, he may then enforce his patent and recover damages for infringement occurring thereafter.").

64. See *Westinghouse Elec.*, 179 F.2d at 145 ("Since plaintiff here has abandoned any illegal conduct, it is entitled to the protection of the courts and there is no reason why relief should not be afforded in the presently pending litigation.").

65. See *supra* note 35 for a definition of a tying arrangement. For a discussion of congressional reform of patent misuse based on tying clauses, see *infra* note 76.

66. *Morton Salt*, 314 U.S. at 494. Section 3 of the Clayton Act provides that:

[I]t shall be unlawful for any person engaged in commerce . . . to lease or make a sale or contract for sale of goods . . . whether patented or unpatented . . . or fix a price charged therefor, or discount from, or rebate upon, such price, on the condition, agreement, or understanding that the lessee or purchaser thereof shall not use or deal in the goods . . . or other commodities of a competitor or competitors of the lessor or seller, where the effect . . . may be to substantially lessen competition or tend to create a monopoly in any line of commerce.

15 U.S.C. § 14 (1988) (emphasis added).

67. *G.S. Suppiger Co. v. Morton Salt Co.*, 117 F.2d 968, 970 (7th Cir. 1941), *rev'd*, 314 U.S. 488 (1942).

68. *Morton Salt*, 314 U.S. at 490.

69. See *Hartford-Empire Co. v. United States*, 323 U.S. 386, 415 (1945) ("[S]o long as the patent owner is using his patent in violation of the antitrust laws, he cannot restrain infringement of it by others."). Indeed, Judge Posner has repeatedly argued that an antitrust violation must be the basis for patent or copyright misuse. See *Saturday Evening Post Co. v. Rumbleseat Press, Inc.*, 816 F.2d 1191, 1200 (7th Cir. 1987) (copyright misuse); *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 511-12 (7th Cir. 1982) (patent misuse). See *infra* notes 77-83 and accompanying text for a discussion of antitrust and nonantitrust violations constituting patent misuse.

70. See, e.g., 3 M. NIMMER & D. NIMMER, NIMMER ON COPYRIGHT § 13.09[A] (1989) (describing patent cases like *Morton Salt* and *Mercoird Corp. v. Minneapolis-Honeywell Regulator Co.*, 320 U.S. 680 (1944), as holding that "a patentee who uses his patent privilege contrary to the public interest by violating the antitrust laws will be denied the relief of a court of equity in a patent infringement action.").

71. 320 U.S. 680 (1944).

to bring unpatented goods within the protection of the patent is measured by the anti-trust laws not by the patent law. . . . [T]he effort here made to control competition in this unpatented device plainly violates the anti-trust laws."⁷² This holding may be interpreted as requiring an antitrust violation as a predicate for patent misuse, but it may be viewed also as a call to permit a patent misuse to be a per se violation of the antitrust laws, without requiring an inquiry into the patentee's market power.⁷³ Although the Supreme Court implicitly has rejected this concept⁷⁴ and reaffirmed its allegiance to *Morton Salt's* patent misuse doctrine,⁷⁵ Congress has recently circumscribed the limits of the patent misuse defense by explicitly requiring courts to employ antitrust principles when the claimed misconduct involves tying arrangements and package licensing.⁷⁶ The interplay and tension between limited monopolies and antitrust doctrines is re-

72. *Id.* at 684. In a companion case, the Court again noted that "the anti-trust acts or other laws not the patent statutes define the public policy." *Mercoird Corp.*, 320 U.S. at 666.

73. Calkins, *supra* note 4, at 184-85. Market power is a critical component of an antitrust inquiry. If the purported anticompetitive conduct involves a tying arrangement, the court must evaluate whether the defendant-seller

has the power, within the market for the tying product, to raise prices or to require purchasers to accept burdensome terms that could not be exacted in a completely competitive market. In short, the question is whether the seller has some advantage not shared by his competitors in the market for the tying product.

United States Steel Corp. v. Fortner Enters., Inc., 429 U.S. 610, 620 (1977) (citation omitted).

74. *Zenith Radio Corp. v. Hazeltine Research, Inc.*, 395 U.S. 100, 140 (1969) ("[I]t does not necessarily follow that the [patent] misuse embodies the ingredients of a violation of either § 1 or § 2 of the Sherman Act, or that a violation threatened Zenith so as to entitle it to an injunction under § 16 of the Clayton Act.").

75. *Dawson Chem. Co. v. Rohm & Haas Co.*, 448 U.S. 176, 221 (1980) (Without importing antitrust principles into patent doctrine, the Court noted that "[t]he policy of free competition runs deep in our law. It underlies both the doctrine of patent misuse and the general principle that the boundary of a patent monopoly is to be limited by the literal scope of the patent claims.").

76. Package licensing may or may not constitute misuse. Two types of package licenses are permissible. One type is when patents "block" each other so that the patents may not be exploited without each infringing the other. A second type is when the patents are "complementary"—i.e., each patent protects a discreet unit within a larger device, making it inefficient, if not impossible, to produce the device without the complementary patents. *W. HOLMES, supra* note 34, at § 22.02. Package licensing may become an illegal tying arrangement when a patentee requires the licensee to take a license on one patent as a condition for the grant of a license on another. *See Automatic Radio Mfg. Co. v. Hazeltine Research, Inc.*, 339 U.S. 827, 830-31 (1952). The 1988 Patent Misuse Reform Act, Pub. L. No. 100-73, tit. II, § 201, 102 Stat. 4676 (1988), amended 35 U.S.C. § 271(d) to read:

(d) No patent owner otherwise entitled to relief for infringement or contributory infringement of a patent shall be denied relief or deemed guilty of misuse or illegal extension of the patent right by reason of having done one or more of the following: . . . (5) conditioned the license of any rights to the patent or the sale of the patented product on the acquisition of a license to rights in another patent or purchase of a separate product, unless, in view of the circumstances, the patent owner has market power in the relevant market for the patent or patented product on which the license or sale is conditioned.

35 U.S.C. § 271(d) (1988) (emphasis added).

Although the *Lasercomb* panel remarked on this change in patent misuse doctrine, *Lasercomb*, 911 F.2d at 976 & n.15, the court failed to take notice of the strong congressional effort to predicate patent misuse doctrine entirely on antitrust principles in accordance with the view espoused by Judge Posner. *See Calkins, supra* note 4, at 192-200. Although Congress ultimately did not adopt such a radical departure from existing patent misuse doctrine, Senate leaders gave notice that Congress intends to make more extensive changes in the future. *Id.* at 197-200. Under existing federal trademark law, the trademark owner's present or past use of the trademark to violate the antitrust laws is a defense to an infringement action. 15 U.S.C. § 1115 (1988).

flected by the broad scope of what may constitute patent misuse,⁷⁷ including agreements not to deal in competing goods,⁷⁸ royalties based on total sales,⁷⁹ price fixing,⁸⁰ tying arrangements,⁸¹ package licensing,⁸² and field-of-use restrictions.⁸³

The admixture of antitrust standards and policy-based misuse doctrine frequently spills over into the area of copyright law. The Supreme Court has never expressly declared the copyright misuse defense to be legitimate. This failure may explain why the lower courts, with one exception prior to *Lasercomb*, have not rushed to transplant *Morton Salt*'s misuse defense into copyright doctrine.

In *M. Witmark & Sons v. Jensen*⁸⁴ a Federal district court, relying on *Morton Salt*, denied injunctive relief because of the copyright owners' anticompetitive conduct.⁸⁵ The *Jensen* plaintiffs owned copyrighted musical works and belonged to the American Society of Composers, Authors and Publishers (ASCAP).⁸⁶ Through ASCAP the plaintiffs licensed to motion picture producers "synchronization rights"—the right to record plaintiffs' musical compositions onto the soundtracks of certain motion pictures.⁸⁷ The plaintiffs' licensing agreement did not include the right to perform publicly the compositions. The plaintiffs claimed that defendant motion picture theater operators infringed by showing films containing the copyrighted musical works without first obtaining a license to perform the music.⁸⁸ Although theater operators may bargain with individual copyright owners for a performance license, for practical reasons they have no real choice but to take a "blanket license" through a performance soci-

77. See Calkins, *supra* note 4, at 187 n.38 (for a discussion of various patent misuses). See generally 4 D. CHISUM, *supra* note 3, at § 19.04[3][a]-[i] (comprehensive discussion of various patent misuses).

78. *Compton v. Metal Prod., Inc.*, 453 F.2d 38, 44-45 (4th Cir. 1971), *cert. denied*, 408 U.S. 968 (1972), cited in *Lasercomb*, 911 F.2d at 979; *Berlenbach v. Anderson & Thompson Ski Co.*, 329 F.2d 782, 784-85 (9th Cir. 1964); *National Lockwasher Co. v. George K. Garrett Co.*, 137 F.2d 255, 256 (3d Cir. 1943). Courts uniformly have found noncompete clauses in patent licenses to be a per se misuse of the patent. See 4 D. CHISUM, *supra* note 3, at § 19.04[3][b] (citing cases). For a discussion of noncompete clauses as the measure for misuse, see *infra* notes 170-75 and accompanying text.

79. Although coercing a licensee to pay royalties based on total sales constitutes misuse, *Zenith Radio Corp. v. Hazeltine Research, Inc.*, 395 U.S. 100, 138 (1969), the patentee and its licensee may agree to the arrangement if it is convenient for the parties to collect royalties in such a way. *Automatic Radio*, 339 U.S. at 834.

80. *United States v. New Wrinkle, Inc.*, 342 U.S. 371, 378 (1952); *United States v. Line Material Co.*, 333 U.S. 287, 306 (1948).

81. *Northern Pac. Ry. v. United States*, 356 U.S. 1, 5-6 (1958). Under the 1988 Patent Reform Act, tying arrangements must be analyzed under antitrust standards to constitute patent misuse. See *supra* note 76 for the language of the statute.

82. *Automatic Radio*, 339 U.S. at 830-31; see *supra* note 76 for a definition of package licensing.

83. A patentee may restrict a licensee to make or use the patent only in a particular field. *WALKER ON PATENTS*, *supra* note 1, § 28:17. It is patent misuse, however, to demand such a restriction after the first sale. *General Talking Picture Corp. v. Western Elec. Co.*, 304 U.S. 175, 181-82, *aff'd on rehearing*, 305 U.S. 124 (1938).

84. 80 F. Supp. 843 (D. Minn. 1948), *appeal dismissed sub nom. M. Witmark & Sons v. Berger Amusement Co.*, 177 F.2d 515 (8th Cir. 1949).

85. *Id.* at 850. Later, the Fourth Circuit cited *Jensen* as the only "case which has actually applied copyright misuse to bar an action for infringement." *Lasercomb*, 911 F.2d at 976.

86. *Jensen*, 80 F. Supp. at 844.

87. *Id.*

88. *Id.*

ety like ASCAP.⁸⁹

The *Jensen* court agreed with defendants that the consequence of ASCAP's licensing practices created the economic power to coerce theater owners into taking blanket performance licenses, giving the plaintiffs advantages far greater than those enjoyed by a single copyright owner.⁹⁰ Although the court found that the copyright owners violated the antitrust laws,⁹¹ it was more concerned with the practice of using a copyright to expand illegitimately the rights granted under the copyright laws to the detriment of the public interest.⁹² The court concluded that it was unnecessary to decide whether antitrust violations alone would constitute copyright misuse barring the plaintiffs' infringement claim.⁹³

The law of antitrust-based copyright misuse is much more developed, although highly fractured.⁹⁴ Due to mixed signals from the Supreme Court, two mutually exclusive doctrines have evolved. One doctrine permits an antitrust-based copyright misuse defense as a result of the Court's linking antitrust violations to the improper enlargement of the copyright monopoly.⁹⁵ The other forbids that type of copyright misuse defense because of the Court's prohibition against "collateral" antitrust defenses.⁹⁶

In *United States v. Paramount Pictures, Inc.*⁹⁷ the Supreme Court implicitly recognized the copyright misuse defense and hinted that antitrust analysis ought to play a role in considering whether a grantee has misused its copyright. The Court found that defendant motion picture distributors' practice of "block-booking"⁹⁸ violated the antitrust laws.⁹⁹ In citing *Morton Salt*, however, the court also remarked that the practice improperly enlarged what the copyright

89. Rather than deal with multitudes of copyright owners, the blanket licensee pays an annual fee to ASCAP and in exchange receives the right to perform all the musical works authorized to ASCAP, regardless of whether the licensee needs or desires all the works in the ASCAP catalogue. *Id.* at 845. When the performance licensee pays ASCAP, that association distributes the royalties to ASCAP members based upon a formula it has devised. *Id.* at 846.

90. *Id.* at 847. The court noted that the licensing scheme gave ASCAP members a monopoly of "80% of all of the music recorded in motion picture films." *Id.*

91. *Id.* at 850.

92. *Id.*

93. *Id.* A more recent case, however, cites *Jensen* as supporting a copyright misuse defense due to practices violating the antitrust laws. *Broadcast Music, Inc. v. Moor-Law, Inc.*, 203 U.S.P.Q. (BNA) 487, 488 (D. Del. 1978).

94. *See Broadcast Music, Inc. v. Hearst/ABC Viacom Entertainment Servs.*, 746 F. Supp. 320, 328 (S.D.N.Y. 1990) (describing copyright misuse defense as "ill-received" by lower courts, typically either rejected out of hand, or if recognized, finding misuse defense inapplicable on the facts).

95. *See infra* notes 97-123 and accompanying text.

96. An antitrust violation may be collateral when the contract in restraint of trade is not the subject matter of the suit. Defendants have asserted, without much success, that the courts should not permit a plaintiff in violation of the antitrust laws to enforce a claim for breach of contract, even though the disputed contract did not form the basis for the antitrust violation. *See, e.g., Kelly v. Kosuga*, 358 U.S. 516, 518-21 (1959). For a discussion of the judicial treatment given to collateral antitrust defenses, *see infra* notes 124-35 and accompanying text.

97. 334 U.S. 131 (1948).

98. Block-booking is a motion picture tying arrangement in which distributors license films "on condition that the exhibitor will also license another feature or group of features released by the distributors during a given period." *Id.* at 156.

99. *Id.* at 159.

granted.¹⁰⁰

In *United States v. Loew's, Inc.*¹⁰¹ the Supreme Court again addressed the practice of block-booking films, but this time to television stations rather than movie theaters. The Court found that the practice violated section 1 of the Sherman Act,¹⁰² and that for tying arrangements the law requires that the seller have "sufficient economic power with respect to the tying product to appreciably restrain free competition in the market for the tied product."¹⁰³ The Court rejected defendants' attempt to distinguish *Paramount Pictures* on the grounds that, because television stations rely on films only for a small fraction of their programming, the defendants lacked sufficient economic power.¹⁰⁴ The Court responded that when the tying product is patented or copyrighted the "requisite economic power is presumed."¹⁰⁵ Therefore, the court held that compelling a licensee to take unwanted films as a condition for the desired license is illegal, and the copyright monopoly must be strictly confined by the antitrust laws.¹⁰⁶ Because the case involved a government suit brought under the antitrust laws, rather than a claim for copyright infringement, the Court did not need to decide whether a copyright misuse defense existed. The Court's specific reference to the patent misuse defense, however, in conjunction with its uniform application of the antitrust laws to both patents and copyrights,¹⁰⁷ suggests that the *Loew's* Court tacitly approved of a copyright misuse defense.¹⁰⁸

In *Broadcast Music, Inc. v. Columbia Broadcasting System*¹⁰⁹ the Supreme Court gave the lower courts further reason to believe that antitrust violations can trigger the copyright misuse defense. Columbia Broadcasting System (CBS)

100. *Id.* at 157. By forcing a licensee to take an inferior copyrighted work as a condition to receiving a superior one, the licensor "strengthens its monopoly" beyond what the copyright intended and reaps disproportionate rewards. *Id.* at 158.

101. 371 U.S. 38 (1962).

102. *Id.* at 39. The Sherman Act provides in part: "Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal." 15 U.S.C. § 1 (1988). Although the statute prohibits "every contract," courts construe this language narrowly by forbidding only those contracts that unreasonably restrain trade. W. HOLMES, *supra* note 34, § 5.01[2]. For a discussion of the various judicial modes of antitrust analysis, see *infra* note 113.

103. *Loew's*, 371 U.S. at 45 (quoting *Northern Pac. Ry. v. United States*, 356 U.S. 1, 6 (1958)).

104. *Id.* at 47-48.

105. *Id.* at 48. Prior to the 1988 Patent Misuse Reform Act, see *supra* note 76, several appellate courts rejected, either expressly or implicitly, the presumptive economic power of patented or copyrighted works used in tying arrangements. The courts instead inquired into the grantee's actual market power in the patented or copyrighted product. See *Nobel Scientific Indus., Inc. v. Beckman Instruments, Inc.*, 670 F. Supp. 1313, 1328 (D. Md. 1986), *aff'd*, 831 F.2d 537 (4th Cir. 1987) (patents); *A.I. Root Co. v. Computer/Dynamics, Inc.*, 806 F.2d 673, 676 (6th Cir. 1987) (case involving copyrighted computer software in which Sixth Circuit observed that "[m]ore often than not . . . a patent or copyright provides little, if any, market power." (quoting Note, *The Presumption of Economic Power for Patented and Copyrighted Products in Tying Arrangements*, 85 COLUM. L. REV. 1140, 1156 (1985))).

106. *Loew's*, 371 U.S. at 49.

107. *Id.* at 46-49. The *Loew's* Court noted that "[a]ccommodation between the statutorily dispensed monopoly in the combination of contents in the patented or copyrighted product and the statutory principles of free competition demands that extension of the patent or copyright monopoly by the use of tying agreements be strictly confined." *Id.* at 49.

108. 3 M. NIMMER & D. NIMMER, *supra* note 70, § 13.09[A].

109. 441 U.S. 1 (1979).

sued Broadcast Music, Inc. (BMI) and ASCAP claiming that the performance societies' blanket licensing of copyrighted musical works amounted to price fixing and was per se illegal under the antitrust laws.¹¹⁰ CBS also asked for a declaratory judgment that BMI and ASCAP were misusing their copyright.¹¹¹

The Court reversed both the Second Circuit's finding that the blanket licensing was illegal per se under the antitrust laws and the copyright misuse judgment that the trial court had predicated on an antitrust violation.¹¹² The Court remanded to the lower court for an assessment of the claim under the antitrust rule of reason.¹¹³ Writing for the Court, Justice White stated:

Although the copyright laws confer no rights on copyright owners to fix prices among themselves or otherwise to violate the antitrust laws, we would not expect that any market arrangements reasonably necessary to effectuate the rights that are granted would be deemed a *per se* violation of the Sherman Act.¹¹⁴

The implication, therefore, is quite strong that the Court not only recognized that the copyright misuse defense exists, but that the antitrust laws are the proper measurement of copyright misuse.¹¹⁵ This implied approval, in both

110. *Id.* at 6.

111. *Columbia Broadcasting Sys. v. American Soc'y of Composers*, 400 F. Supp. 737, 771 (S.D.N.Y. 1975), *rev'd*, 562 F.2d 130 (2d Cir. 1977), *rev'd sub nom.* *Broadcast Music, Inc. v. Columbia Broadcasting Sys.*, 441 U.S. 1 (1979).

112. *Broadcast Music*, 441 U.S. at 24.

113. *Id.* at 24-25. Under the rule of reason an agreement unreasonably restrains trade if, under all the facts and circumstances, its anticompetitive effects outweigh its procompetitive effects. See *National Soc'y of Professional Eng'rs v. United States*, 435 U.S. 679, 692 (1978); J. VAN CISE, W. LIFLAND & L. SORKIN, *UNDERSTANDING THE ANTITRUST LAWS* 35 (9th ed. 1986). This factual inquiry has its limits in that "the Rule does not open the field of antitrust inquiry to any argument in favor of a challenged restraint that may fall within the realm of reason. Instead, it focuses directly on the challenged restraint's impact on competitive conditions." *Professional Eng'rs*, 435 U.S. at 688.

Agreements that are illegal per se are those "whose nature and necessary effect are so plainly anticompetitive that no elaborate study of the industry is needed to establish their illegality." *Id.* at 692. The law has deemed certain practices facially anticompetitive: price fixing, tying arrangements, agreements between competitors to divide markets, boycotts, pooling profits, and bans on competitive bidding. J. VAN CISE, W. LIFLAND & L. SORKIN, *supra*, at 116. Before courts will find that the challenged trade practice is presumptively illegal, the courts require proof that the practice creates adverse economic effects in light of judicial experience that "the practice is of a type that will 'always or almost always tend to restrict competition.'" W. HOLMES, *supra* note 34, § 5.01[2] (quoting *Broadcast Music*, 441 U.S. at 19-20).

A defendant may rebut the presumption of illegality by two methods: (1) Showing that the restraint is essential to allow the product to become available; if so, then the restraint is scrutinized under the rule of reason and will be upheld if the restraint's potential for anticompetitive effects is never realized and the purpose for the restraint is lawful; (2) If the restraint does not involve price, the restraining party does not have market power, and the restraint is "reasonably ancillary to efficiency in competition." J. VAN CISE, W. LIFLAND & L. SORKIN, *supra*, at 118-19.

114. *Broadcast Music*, 441 U.S. at 19 (emphasis in original).

115. Several commentators support this view. See, e.g., W. HOLMES, *supra* note 34, § 4.09; Comment, *Copyright Misuse and Cable Television: Ortho-O-Vision, Inc. v. Home Box Office*, 35 FED. COMM. L.J. 347, 360 (1980) ("[T]he Court agreed that the limits of the copyright grant are defined by the antitrust laws."); Note, *supra* note 5, at 1305; Note, *Broadcast Music, Inc. v. Columbia Broadcasting System, Inc.: The Copyright Misuse Doctrine*, 15 NEW. ENG. L.J. 683, 693 (1980) ("[T]he Court appears to be endorsing a copyright misuse doctrine which is analogous to the patent misuse doctrine and requires no analysis of copyright policy.").

This viewpoint is not universally shared. See, e.g., Phelan, *The Continuing Battle with the Performing Rights Societies: The Per Se Rule, the Rule of Reason Standard and Copyright Misuse*, 15

Broadcast Music and *Loew's*, has influenced several lower courts in their finding of copyright misuse predicated on antitrust illegality.¹¹⁶

The Seventh Circuit, and Judge Posner in particular, has accepted the Supreme Court's apparent invitation to permit a copyright misuse defense, but only if predicated exclusively on antitrust violations. In *Saturday Evening Post Co. v. Rumbleseat Press, Inc.*¹¹⁷ the plaintiff-licensor sued its licensee for breach of the copyright licensing agreement. The agreement contained a clause stating that the licensee would not dispute the validity of any copyrights the licensor may have obtained.¹¹⁸ The licensee argued that the no-contest clause violated the public policy of the Copyright Act and therefore was unenforceable.¹¹⁹ The Seventh Circuit responded that no-contest clauses in copyright licenses are valid unless they violate the antitrust laws.¹²⁰ The court cited a previous Seventh Circuit patent misuse opinion for its reasoning:

"If misuse claims are not tested by conventional antitrust principles, by what principles shall they be tested? Our law is not rich in alternative concepts of monopolistic abuse; and it is rather late in the day to try to develop one without in the process subjecting patent holders to debilitating uncertainty."¹²¹

The court stated that this reasoning was even more persuasive regarding

TEX. TECH. L. REV. 349, 377 (1984) ("[T]he Supreme Court did not expressly rule that a finding of copyright misuse is always dependent upon an antitrust violation."). Similarly the *Lasercomb* panel rejected the idea that *Broadcast Music* mandates an antitrust-based misuse defense, stating that viewed "[i]n context . . . it is apparent that misuse was linked to antitrust in that case simply as a matter of litigation strategy. Copyright misuse was not asserted as a defense to an infringement suit, and the primary claim was an antitrust claim." *Lasercomb*, 911 F.2d at 977 n.17.

116. See *United Tel. Co. v. Johnson Publishing Co.*, 855 F.2d 604, 612 (8th Cir. 1988) (recognizing the concept of antitrust-based copyright misuse, but finding no antitrust violation in this case); *Saturday Evening Post Co. v. Rumbleseat Press, Inc.*, 816 F.2d 1191, 1200 (7th Cir. 1987) (holding that a no-contest clause in a copyright licensing agreement is valid unless violative of antitrust law); *Broadcast Music, Inc. v. Hearst/ABC Viacom Entertainment Servs.*, 746 F. Supp. 320, 327-28 (S.D.N.Y. 1990) (recognizing that an affirmative defense of copyright misuse is cognizable following a finding that defendant pleaded a colorable monopolization counterclaim); *F.E.L. Publications, Ltd. v. Catholic Bishop*, 506 F. Supp. 1127, 1136 (N.D. Ill. 1981) (tying arrangement held to be a per se Sherman Act violation and therefore a misuse of plaintiff's copyright), *rev'd*, 214 U.S.P.Q. (BNA) 409, 413 n.9 (7th Cir. 1982) (reversing the district court's finding of a per se Sherman Act violation, but noting that copyright misuse is an equitable defense that requires a balancing of equities); *Broadcast Music, Inc. v. Moor-Law, Inc.*, 484 F. Supp. 357, 364-65 (D. Del. 1980) (denying plaintiff-copyright holder's motion for summary judgment against allegedly infringing defendant's copyright misuse defense and antitrust counterclaim on grounds that a factual question existed whether plaintiff's conduct violated antitrust rule of reason); *Edward B. Marks Music Corp. v. Colorado Magnetics, Inc.*, 357 F. Supp. 280, 287 (W.D. Okla. 1973) (plaintiff's antitrust violations constituted copyright misuse, rendering copyright unenforceable), *rev'd*, 497 F.2d 285, 290 (10th Cir. 1974) (assuming that antitrust violations are copyright misuse, no basis to support finding of antitrust violations), *cert. denied*, 419 U.S. 1120 (1975).

117. 816 F.2d 1191 (7th Cir. 1987). The plaintiff magazine licensed to the defendant the right to manufacture porcelain dolls based on Norman Rockwell illustrations. *Id.* at 1193. Although the license warranted that the *Saturday Evening Post* owned the copyright in the Rockwell illustrations, it had a copyright only in each of the magazines and not in the illustrations themselves. *Id.* After the license terminated the defendant continued making the dolls, precipitating the litigation. *Id.*

118. *Id.*

119. *Id.* at 1199.

120. *Id.* at 1200.

121. *Id.* (quoting *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 512 (7th Cir. 1982) (Posner, J.)).

copyright misuse because copyrights present a smaller threat of monopolization than do patents.¹²² The court observed that if a copyright holder seeks to extend its monopoly power beyond that granted by the statute, the proper avenue of attack is under the Sherman Act, rather than "a federal common law rule that would jostle uncomfortably with the Sherman Act."¹²³

In marked contrast to this strain of legal thought that demands an anti-trust-based misuse defense is a line of cases that rejects the idea that antitrust violations can preclude enforcement of a copyright infringement claim. This judicial hostility borrows from a long-held doctrine that an antitrust violation is not a defense to a breach of contract.¹²⁴ Beginning in the early part of this century, the Supreme Court distinguished between contracts that are inherently illegal under the antitrust laws and contracts that are merely collateral to the alleged illegality.¹²⁵ When the plaintiff sought to enforce a contract, the Court repeatedly ruled that the plaintiff's violation of the antitrust laws is not a defense if the contract itself is not inherently illegal under the antitrust laws.¹²⁶

For example, in *Bruce's Juices, Inc. v. American Can Co.*¹²⁷ the plaintiff sought to collect on promissory notes the defendant had signed following purchases of goods from the plaintiff.¹²⁸ The defendant asserted that the notes were uncollectible because the plaintiff's pricing policy for other buyers was illegally discriminatory under the Robinson-Patman Act.¹²⁹ The Court rejected this antitrust defense because the contract was not inherently illegal¹³⁰ and because voiding contracts was not among the Act's exclusive remedies.¹³¹

Copyright jurisprudence mirrors this doctrine, flatly denying that a collateral antitrust defense can bar an action to enforce an infringement claim.¹³² The

122. *Id.*

123. *Id.*

124. For a review of the role antitrust violations have played in actions to enforce contracts see Note, *The Defense of Antitrust Illegality in Contract Actions*, 27 U. CHI. L. REV. 758, 759-68 (1960); Note, *supra* note 5, at 1291-93.

125. Compare *Connolly v. Union Sewer Pipe Co.*, 184 U.S. 540, 547 (1902) (plaintiff's contract to sell pipe to the defendant unrelated to plaintiff's conspiracy to fix prices) with *Continental Wall Paper Co. v. Louis Voight & Sons Co.*, 212 U.S. 227, 262 (1909) (enforcement of requirements contract denied because it was part of a scheme to violate the antitrust laws).

126. See *Kelly v. Kosuga*, 358 U.S. 516, 521 (1959); *Bruce's Juices, Inc. v. American Can Co.*, 330 U.S. 743, 756-57 (1947).

127. 330 U.S. 743 (1947).

128. *Id.* at 744.

129. *Id.* at 744-45. The defendant claimed that the plaintiff's practice of giving discounts to volume buyers of cans exerted anticompetitive pressures against a small cannery like itself. *Id.* at 745. The Robinson-Patman Act outlaws price discrimination between purchasers "where the effect of such discrimination may be substantially to lessen competition or tend to create a monopoly . . . or prevent competition with any person who either grants or knowingly receives the benefit of such discrimination." 15 U.S.C. § 13(a) (1988).

130. *Bruce's Juices*, 330 U.S. at 755-56.

131. *Id.* at 750-51.

132. *BellSouth Advertising & Publishing Corp. v. Donnelley Information Publishing*, 719 F. Supp. 1551, 1563 (S.D. Fla. 1988) (denying antitrust defense to copyright infringement but recognizing misuse defense when there is an attempt to extend the limited monopoly beyond the protected work); *Rural Tel. Serv. Co. v. Feist Publishing, Inc.*, 663 F. Supp. 214, 220 (D. Kan. 1987); *Ortho-Vision, Inc. v. Home Box Office*, 474 F. Supp. 672, 686 (S.D.N.Y. 1979); *Broadcast Music, Inc. v. Grant's Cabin*, 204 U.S.P.Q. (BNA) 633, 634-45 (E.D. Mo. 1979); *United Artists Associated v. NWL Corp.*, 198 F. Supp. 953, 957 (S.D.N.Y. 1961) (rejecting antitrust defense but denying plain-

rationale behind this doctrine is twofold. First, the courts will not construe the Sherman Act to divest a violator of his property rights, including ownership of a valid copyright.¹³³ Second, the antitrust laws offer specific remedies that are exclusive.¹³⁴ Both of these reasons are ones the Fourth Circuit expressly has supported in the past,¹³⁵ yet the *Lasercomb* court failed to address these points. *Lasercomb* simply gives no indication whether the Fourth Circuit now questions the continuing vitality of this precedent.

In the copyright-antitrust dynamic, a less clear-cut third choice has emerged. In *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*¹³⁶ the Second Circuit embroidered upon the collateral antitrust defense by providing that courts must balance the competing policies of preventing copyright piracy and enforcing the antitrust laws.¹³⁷ The district court in *Alfred Bell* had rejected the antitrust defense as a matter of law,¹³⁸ recognizing that the Supreme Court's denial of the collateral antitrust defense in *Bruce's Juices* had limited the misuse principles outlined in *Morton Salt*.¹³⁹ Although the Second Circuit affirmed, it chose instead to take "into account the comparative innocence or guilt of the parties, the moral character of their respective acts, the extent of the harm to the public interest, [and] the penalty inflicted on the plaintiff if we deny it relief."¹⁴⁰ Under this balancing test, the Second Circuit found a clear copyright infringement, but only a marginal antitrust violation, and therefore, held that the defendants "did not establish the anti-trust 'unclean hands' defense."¹⁴¹ Subsequent courts, including the Fourth Circuit, have agreed that balancing the equities is the proper response to a copyright misuse defense.¹⁴²

Against the backdrop of the last half-century of patent and copyright law,

tiff's motion to strike copyright misuse defense); *Peter Pan Fabrics, Inc. v. Candy Frocks, Inc.*, 187 F. Supp. 334, 336 (S.D.N.Y. 1960); *Harms, Inc. v. Sansom House Enters.*, 162 F. Supp. 129, 135 (E.D. Pa. 1958), *aff'd sub nom. Leo Feist, Inc. v. Lew Tendler Tavern, Inc.*, 267 F.2d 494 (3d Cir. 1959); *M. Witmark & Sons v. Pastime Amusement Co.*, 298 F. 470, 480 (E.D.S.C.), *aff'd mem.*, 2 F.2d 1020 (4th Cir. 1924).

133. *Pastime Amusement*, 298 F. at 480 ("The Sherman Act does not make the party to an interstate monopoly an outlaw.").

134. *Id.*

135. The Fourth Circuit affirmed *Pastime Amusement* without comment. *M. Witmark & Sons v. Pastime Amusement Co.*, 2 F.2d 1020 (4th Cir. 1924).

136. 191 F.2d 99 (2d Cir. 1951).

137. *Id.* at 106.

138. *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 74 F. Supp. 973, 979 (S.D.N.Y. 1947), *aff'd*, 191 F.2d 99 (2d Cir. 1951).

139. *Id.* at 978.

140. *Alfred Bell*, 191 F.2d at 106.

141. *Id.* at 105.

142. See *F.E.L. Publications, Ltd. v. Catholic Bishop*, 214 U.S.P.Q. (BNA) 409, 413 n.9 (7th Cir. 1982) ("Dismissal of a copyright claim for misuse is an equitable defense which requires a balancing of equities."); *Inge v. Twentieth Century-Fox Film Corp.*, 143 F. Supp. 294, 298-99 (S.D.N.Y. 1956) (employing *Alfred Bell*'s balancing test; finding that the facts did not support an antitrust defense). In *Stein v. Mazer*, 204 F.2d 472, 480 (4th Cir. 1953), *aff'd*, 347 U.S. 201 (1954), the defendant asserted a copyright misuse defense to plaintiff's action for infringement on the grounds that it reproduced the copyrighted work of art for utilitarian purposes. The defendant presented no unfair competition claims. The Fourth Circuit, citing *Morton Salt*, stated that it was "not impressed" by the copyright misuse claim and observed that "the equities . . . lie with the plaintiffs, not with the defendants, who deliberately and meticulously copied the copyrighted statutes." *Id.*

the *Lasercomb* decision is significant because it marks the first time an appellate court has used the copyright misuse defense to deny an infringement action.¹⁴³ Yet *Lasercomb* is far from surprising.¹⁴⁴ Given the widespread acceptance of *Morton Salt*'s patent misuse doctrine, it seemed almost inevitable that an identical doctrine would appear in the law of copyright. Essentially the *Lasercomb* court's analysis boils down to a syllogism: Patent and copyright law have parallel constitutional goals of promoting progress. Under *Morton Salt* a patentee who misuses his patent by exceeding the scope of the limited monopoly frustrates that policy goal and is unworthy of legal protection. Therefore, copyright law must recognize copyright misuse as a defense to infringement. While this reasoning has some appeal, it fails to distinguish critically between patent and copyright theory and application, to comprehend the role copyrights play in modern economics, or to provide any guideposts for what constitutes copyright misuse.

By framing copyright policy as one designed to "promote progress," the Fourth Circuit sidestepped two important points. First, although patents and copyrights both offer limited monopolies to individuals primarily for the benefit of the public, there are significant theoretical and practical differences between them;¹⁴⁵ these differences show that patents are more difficult than copyrights to obtain, but offer more protection from infringement.¹⁴⁶ For example, because patent infringers cannot interpose the defenses of independent creation or fair use as can copyright infringers, a patent affords substantially greater economic power than does a copyright.¹⁴⁷ Thus, the danger to the public interest is

143. At the district court level, only two cases have used the misuse defense to bar the copyright holder's claim: *F.E.L. Publications, Ltd. v. Catholic Bishop*, 506 F. Supp. 1127, 1136-37 (N.D. Ill. 1981) (granting defendant's motion for summary judgment on grounds plaintiff misused its copyright), *rev'd*, 214 U.S.P.Q. (BNA) 409, 413 n.6 (7th Cir. 1982) (recognizing copyright misuse defense exists, but finding copyright holder did not violate the antitrust laws); *M. Witmark & Sons v. Jensen*, 80 F. Supp. 843, 850 (D. Minn. 1948), *appeal dismissed sub nom. M. Witmark & Sons v. Berger Amusement Co.*, 177 F.2d 515 (8th Cir. 1949).

144. A substantial number of cases have considered the nature of copyright misuse; two of these cases were from the Fourth Circuit. See *supra* notes 133-35 & 142 and accompanying texts. Interestingly, the *Lasercomb* court cited only a single case, *Jensen*, that directly addressed the copyright misuse defense. The opinion, however, did note repeatedly that this is an unsettled area of the law and used such phrases as: "much uncertainty engulfs the 'misuse of copyright defense,'" "there is little case law," "the paucity of precedent," and "uncertainty persists." *Lasercomb*, 911 F.2d at 973-76. Although the *Lasercomb* panel expressed these concerns partly due to the lack of a definitive ruling from the Supreme Court, it is unclear whether these statements are the product of disingenuousness or insufficient research. Also, instead of analyzing prior relevant cases, the court expended its energy scrutinizing the monopoly practices of the British Crown during the sixteenth and seventeenth centuries. *Id.* at 974-75. Rather than reaching so far afield, perhaps if the court had analyzed the case law that has developed over the past few decades, the court's opinion would have demonstrated a firmer grasp of the nuances associated with the misuse doctrine.

145. For a comparison of patents and copyrights, see *supra* notes 1 & 6; *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 100-04 (1951). *Contra* *Lee v. Runge*, 404 U.S. 887, 890 (1971) (order denying certiorari) ("No distinction is made in the constitutional language between copyrights and patents and I would not create one by judicial gloss.") (Douglas, J., dissenting); Gibbs, *supra* note 5, at 74-78 ("It should be concluded that there is no case for applying a distinction in the availability of the misuse defense between patent infringement and copyright infringement cases.").

146. *Stein v. Mazer*, 204 F.2d 472, 474 (4th Cir. 1953), *aff'd*, 347 U.S. 201 (1954).

147. See *Saturday Evening Post Co. v. Rumbleseat Press, Inc.*, 816 F.2d 1191, 1200 (7th Cir. 1987). Although a patentee's economic power may be greater, that power is offset by the shorter life

greater when a patentee seeks to enlarge his monopoly than when a copyright holder acts similarly.

Second, the "promotion of progress" desired by the authorities is inextricably bound to an incentive-based scheme that rewards the grantee according to the amount of copyrighted items sold or licensed.¹⁴⁸ As a result of this royalty scheme, owners of copyrighted works compete with one another to provide the highest quality and most desirable works at a price the market is willing to pay. The public policy in need of protection, therefore, is one that ensures that anticompetitive practices do not impede or restrain the public's demand for copyrighted works.¹⁴⁹

The *Lasercomb* decision presents three significant problems. Beyond condemning noncompete clauses in copyright licenses, it does not delineate the boundaries of copyright misuse, because the Fourth Circuit opted to ground misuse on an ill-defined public policy rather than on the antitrust laws. Furthermore, enforcement of a policy-based misuse doctrine announced in *Lasercomb* necessarily depends on a morally culpable party. Finally, the opinion, which states that antitrust-based misuse is acceptable, fails to grapple with the problem of collateral antitrust defenses.

The *Lasercomb* panel never clearly articulated the public policy that informs the copyright misuse defense. The court noted that society benefits from "the efforts of authors to introduce new ideas and knowledge into the public domain"¹⁵⁰ and that copyright law seeks "to increase the store of human knowl-

span of a patent compared to a copyright. See *supra* notes 1 & 6 for a comparison of patents and copyrights.

148. The Supreme Court has observed that

[t]he economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors in "Science and useful Arts." Sacrificial days devoted to such creative activities deserve rewards commensurate with the services rendered.

Mazer v. Stein, 347 U.S. 201, 219 (1954).

149. See Note, *supra* note 5, at 1307-08. The Supreme Court has acknowledged repeatedly that the copyright law's rewards to a grantee act as a mechanism to ensure that the public gains access to new expressions: "[P]rivate motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts." *Sony Corp. v. Universal City Studios*, 464 U.S. 417, 432 (1984) (quoting *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975)) (emphasis added). In a similar vein, the patent grant also encourages competition:

When a new product appears on the market, it stimulates all concerned in the trade to improve on it or to develop a competing product. After the Patent Office grants a patent for the invention, the publication to the world of the specification adds to the fund of public knowledge. The process continues *ad infinitum*. In this connection, it appears that a fundamental object of the Patent Laws is in effect to encourage the replacement of inferior goods or processes by superior ones

WALKER ON PATENTS, *supra* note 1, § 28:2. At the heart of both patent and antitrust doctrine is the objective of maximizing consumer wealth by producing desirable goods at the lowest cost. There may be a diversity of approaches on how to reach this goal, but patent and antitrust law are wedded in a unitary purpose. W. BOWMAN, PATENT AND ANTITRUST LAW 1 (1973).

150. *Lasercomb*, 911 F.2d at 975. A copyright does not protect ideas, only expressions. Ideas are in the public domain and may not be copyrighted. See *Mazer*, 347 U.S. at 217. For example, in the literary world there are certain common ideas or themes that all are free to use, such as young love tragically destroyed by fierce antagonisms between the families of the two lovers. This is an idea as old as Shakespeare and is not copyrightable; what is copyrightable are the expressions that convey

edge and arts.”¹⁵¹ If that is true, then the goal of copyright policy should be to encourage authors to meet the demand for such expressions and to remove obstacles that interfere with that demand. Such obstacles then may appropriately constitute misuse. Rather than explore this aspect of copyright policy, the court repeated *Morton Salt*’s command that “[public policy] . . . forbids the use of a patent to secure an exclusive right or limited monopoly not granted by the Patent Office and which it is contrary to public policy to grant.”¹⁵² The circular reasoning within that statement is apparent: Public policy forbids what is contrary to public policy. The only concrete idea in that tautology is the stricture against expanding rights beyond what is conferred by the statute. Although *Lasercomb* does not analyze why this expansion is against public policy, the implication is that a copyright holder who extends his limited monopoly earns a reward greater than what the statute contemplated. The courts, therefore, will not protect one who receives an unearned property right because of the harm done to the public interest.

In light of the Fourth Circuit’s unwillingness to explore the contours of a policy-based misuse defense, examining the sources on which the court relied is instructive to better understand the policy the court invoked. One commentator the court cited argues that the improper extension of the limited monopoly requires a court to determine whether this conduct has anticompetitive effects, because “the copyright law necessarily incorporated within it the public policy of encouraging a competitive economy.”¹⁵³ Misuse analysis, therefore, is dependent on “an uncodified law of anticompetitive practices”¹⁵⁴ rather than on federal antitrust law. This same commentator, however, then argues that courts “should be less interested in actual anticompetitive effects and more interested in preserving the limits on monopoly grants”¹⁵⁵ to prevent the copyright holder from obtaining a reward greater than the copyright permits. This statement suggests that the copyright grant is a safe harbor beyond which the grantee dare not venture for fear of losing the court’s protection. Such reasoning offers a purist’s view of the copyright grant as pristine and inviolable. The result is a formalistic approach to the law. The copyright grant protects conduct it expressly authorizes, while other conduct may expose the grantee to infringements that cannot be remedied because the court assumes that such conduct is anticompetitive in a way harmful to the public interest.

Rather than make assumptions about the allegedly anticompetitive effects generated by conduct outside the copyright grant, courts should look to the established body of antitrust law for their misuse analysis.¹⁵⁶ Some commentators

the idea—the dialogue, unique plot twists, or clearly delineated characters. For an instructive essay on the distinction between idea and expression, see Judge Hand’s opinion in *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 120-23 (2d Cir. 1930).

151. *Lasercomb*, 911 F.2d at 976.

152. *Id.* (quoting *Morton Salt Co. v. G.S. Suppiger Co.*, 314 U.S. 488, 492 (1942)).

153. Gibbs, *supra* note 5, at 42.

154. *Id.*

155. *Id.* at 46-47.

156. See *supra* note 113 for a discussion of antitrust principles.

argue that because copyright law predates antitrust law, the latter is unnecessary to enforce the former.¹⁵⁷ Judge Posner has observed, however, that because the misuse doctrine evolved before the antitrust laws, its current viability in judging anticompetitive effects is questionable. Furthermore, he has argued that because the antitrust laws affect all practices that could impair competition substantially, establishing a separate doctrine with an identical purpose is difficult.¹⁵⁸ Judge Posner mused that courts would use such a doctrine to condemn "any . . . licensing practice that is even trivially anticompetitive, at least if it has no socially beneficial effects."¹⁵⁹

The *Lasercomb* opinion reflects the tendency for the misuse doctrine to assume harmful anticompetitive effects. In examining the noncompete clause in *Lasercomb's* licensing agreement,¹⁶⁰ the court asserted that the licensee is "required to forego utilization of the creative abilities of all its officers, directors and employees in the area of CAD/CAM die-making software. Of yet greater concern, these creative abilities are withdrawn from the public."¹⁶¹ The withdrawal of one company from the relevant market does not, however, guarantee that the noncompete clause has harmful consequences; that would depend upon, among other factors, the relative market strength the licensor already enjoyed.¹⁶² In *Lasercomb's* case, its Interact program did not dominate the marketplace, and was actually one of seven competing systems available on the market.¹⁶³ Market power, an essential element in antitrust analysis, is thus irrelevant to the purist's view of copyright law, which the *Lasercomb* court followed.¹⁶⁴ The inconsistency in this approach is striking. The Fourth Circuit seeks to condemn misuse on economic grounds without analyzing its economic effects.

It is interesting to contrast *Lasercomb's* analysis with the economic inquiry undertaken by *Lasercomb's* judicial ancestor, *Jensen*.¹⁶⁵ In *Jensen* the court found that the grantees' licensing methods allowed them to capture eighty per-

157. See, e.g., Gibbs, *supra* note 5, at 36 ("Analysis of allegedly anticompetitive conduct from the standpoint of copyright policy alone is absolutely proper, for federal copyright protection existed for one hundred years before the Sherman Act became law.").

158. *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 511 (7th Cir. 1982).

159. *Id.*

160. See *supra* note 19 for the language of the licensing agreement.

161. *Lasercomb*, 911 F.2d at 978.

162. In this respect it may be helpful to analogize the noncompete clause to a merger between competing companies. Both practices eliminate competition between firms that are, or could be, rivals. The economic result, however, is not entirely clear-cut. If the two rivals have substantial market shares, the result may well be monopoly and a restriction of output, with an attendant rise in prices. If, in contrast, the rivals have insignificant market shares, their combination may either have a neutral effect or result in economies of scale, improved management, or greater access to capital, which in turn can produce a more "efficient" system, *i.e.*, one that produces goods that consumers desire. See R. BORK, *THE ANTITRUST PARADOX* 105, 219 (1978).

Analogizing between noncompete clauses and mergers, however, is fraught with imprecision. In the merger situation, the market shares are a quantifiable amount; with the licensee who agrees not to compete, it is pure speculation what the licensee's market share could have been had it not agreed to the noncompete clause. For the procompetitive effects that may flow from noncompete or exclusive dealing arrangements, see *infra* notes 176-78 and accompanying text.

163. Brief of Plaintiff-Appellee at 13.

164. See *supra* note 73 for a definition of market power.

165. *M. Witmark & Sons v. Jensen*, 80 F. Supp. 843 (D. Minn. 1948), *appeal dismissed sub nom. M. Witmark & Sons v. Berger Amusement Co.*, 177 F.2d 515 (8th Cir. 1949).

cent of the market for music integrated into motion picture soundtracks.¹⁶⁶ Although the *Jensen* court chose not to predicate copyright misuse on an anti-trust violation, the court obviously was influenced by the scope of the grantees' anticompetitive behavior. The *Jensen* court observed palpable effects flowing from this conduct, noting that it effectually restrained competition¹⁶⁷ giving the grantees the power to "sound the death knell of every motion picture theatre in America."¹⁶⁸ The court concluded that such power clearly extended beyond what was granted under the copyright laws; the court therefore would not condone using one monopoly to obtain another.¹⁶⁹

In contrast, in its *Lasercomb* opinion the Fourth Circuit could not detail any tangible effects resulting from Lasercomb's licensing agreement. By the time of trial Lasercomb had sold approximately forty licenses for Interact, but the defendants could prove that only one Interact licensee was bound by the standard licensing agreement that included the noncompete clause.¹⁷⁰ The court was unimpressed with Lasercomb's assertion that the noncompete clause was negotiable and had been modified on previous occasions.¹⁷¹ Instead, the court expressed concern that the noncompete clause was effective for ninety-nine years, which could last longer than the life of the copyright.¹⁷² But most important to the *Lasercomb* panel was the similarity between the language of this noncompete clause and the language found in a patent license that the Fourth Circuit earlier had condemned as patent misuse. Misuse resulted from such a condition because it "'unreasonably lessen[ed] the competition which the public has a right to expect.'" ¹⁷³ The Fourth Circuit, however, did not expound on what would be a reasonable lessening of competition that would not constitute misuse. Nor did the court document any actual anticompetitive effects. Instead, the Fourth Circuit retreated to its purist's view of copyright law: for the *Lasercomb* panel, a quantum of harm sufficient to injure the public interest could be found in "Lasercomb's attempt to use its copyright in . . . the Interact software, to control competition in an area outside the copyright, *i.e.*, the idea of computer-assisted die manufacture."¹⁷⁴ While the Fourth Circuit asserted that

166. *Id.* at 847.

167. *Id.* at 849.

168. *Id.* at 847.

169. *Id.* at 850.

170. *Lasercomb*, 911 F.2d at 973 & n.8. For the language of the disputed noncompete clause, see *supra* note 19.

171. *Lasercomb*, 911 F.2d at 973 n.8.

172. *Id.* at 978 & n.21. For works created on or after January 1, 1978, the copyright endures "for the life of the author and fifty years after the author's death." 17 U.S.C. § 302(a) (1988).

173. *Lasercomb*, 911 F.2d at 979 (quoting *Compton v. Metal Prod., Inc.* 453 F.2d 38, 45 (4th Cir. 1971) (emphasis added)). For a discussion of the judicial hostility towards noncompete clauses in patent licenses, see *infra* note 175.

174. *Lasercomb*, 911 F.2d at 979 (emphasis added). Despite its professed allegiance to a pure, nonantitrust misuse doctrine, the Fourth Circuit seems to be importing the antitrust concept of attempted monopolization, which is forbidden under the Sherman Act. 15 U.S.C. § 2 (1988). Judicial analysis of attempted monopolization claims involves three parts: (1) Inquiry into the relevant market; (2) whether the defendant possessed a specific intent to achieve monopoly power, as opposed to engaging in mere vigorous competition; (3) whether there is a "dangerous probability" that the defendant, if left unchecked, would achieve an actual monopoly. W. HOLMES, *supra* note 34, § 6.05.

restraints in competition injure the public, in the same breath the court found a public injury even if a party only attempts to impair competition.¹⁷⁵

Lasercomb's assumption of public harm also fails to consider that some seemingly restrictive clauses actually have procompetitive consequences.¹⁷⁶ For example, by conditioning a license on the licensee's agreement not to create competing items, the licensor must compensate the licensee for the restriction.¹⁷⁷ Such compensation lowers the price for the license, creating a cost savings that the licensee can pass to the consumer. Although the licensee effectively takes itself out of the particular market, the agreement should be enforceable unless it demonstrably violates the Sherman Act's prohibition against contracts that restrain trade in a commercially unreasonable manner.¹⁷⁸ The overall effect of ignoring the well-established body of antitrust law and replacing it with a vague

Proof that the defendant already had a substantial market share may satisfy the "dangerous probability" inquiry. *Id.*

By holding a grantee liable for attempts to control competition, the Fourth Circuit thus is trying to have it both ways. It is holding grantees to a higher standard of commercial behavior than those who do not possess a copyright, but without engaging in any serious economic analysis.

175. Although the Fourth Circuit did not scrutinize rigorously the nature of noncompete clauses, given the precedents found in patent law, it likely assumed it did not have to. For nearly fifty years the federal courts have consistently considered noncompete clauses in patent licenses to be misuse per se. See *Compton v. Metal Prod., Inc.*, 453 F.2d 38, 45 (4th Cir. 1971); *Berlenbach v. Anderson & Thompson Ski Co.*, 329 F.2d 782, 784 (9th Cir. 1964); *McCullough v. Kammerer Corp.*, 166 F.2d 759, 761-62 (9th Cir. 1948); *National Lockwasher Co. v. George K. Garrett Co.*, 137 F.2d 255, 256 (3d Cir. 1943); *Park-In Theatres v. Paramount Richards Theatres*, 90 F. Supp. 730, 734 (D. Del.), *aff'd*, 185 F.2d 407 (3d Cir. 1950). One early case, relying in part on *Morton Salt*, observed that "[a] patentee's right does not extend to the use of the patent to purge the market of competing non-patented goods, except of course, through the process of fair competition." *National Lockwasher Co.*, 137 F.2d at 256.

Later courts adhered to this view, finding patent misuse when the patentee did not enforce the noncompete clause, *Berlenbach*, 329 F.2d at 785, or despite arguments that the licensee had an implicit duty of good faith, under an exclusive license, to exploit the patent and refrain from using competing nonpatented items. *Park-In Theatres*, 90 F. Supp. at 733-34. In *Park-In Theatres* the patentee argued that the licensee who enjoyed the exclusive right to exploit the patent could choose to use another nonpatented good, thereby depriving the patentee of any royalties, an act the patentee considered to be in bad faith. *Id.* at 733.

Lasercomb advanced essentially the same argument in its brief, asserting that a licensee owes a "duty of care not to undertake to create a new work to substitute for the licensed work, even though a stranger to the agreement is entirely free to undertake to create a substitute work." Brief of Plaintiff-Appellee at 21. Such a view is supported by some courts and commentators. See *Conan Properties, Inc. v. Mattel, Inc.*, 712 F. Supp. 353, 364-65 (S.D.N.Y. 1989); *SAS Inst., Inc. v. S & H Computer Systems*, 605 F. Supp. 816, 827-28 (M.D. Tenn. 1985); 3 M. NIMMER & D. NIMMER, *supra* note 70, § 10.11[B]. Because *Lasercomb* provided certain codes that permitted a licensee to convert *Interact* for the purpose of making the program compatible with the licensee's hardware, *Lasercomb* argued that the noncompete license was essential to protect against a licensee misappropriating *Interact*. Brief of Plaintiff-Appellee at 21. This vulnerability to misappropriation is probably what the district court referred to when it said that "'such a clause is reasonable in light of the delicate and sensitive area of computer software.'" *Lasercomb*, 911 F.2d at 973 (quoting the trial court's findings). The *Lasercomb* panel, however, did not address this issue, finding the noncompete clause to be misuse per se. *Id.* at 979.

176. See *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 510-11 (7th Cir. 1982) (Posner, J.).

177. *Id.* at 510.

178. See *supra* note 102 for the language of the statute. One commentator notes that per se condemnation of noncompete or exclusive dealing arrangements is unwarranted due to certain procompetitive effects that they may generate. W. HOLMES, *supra* note 34, § 21.03. Purchasers may be benefited, for example, by an assured supply, protection of the purchaser's capital investment, and ease in long-term planning. *Id.*

policy-based doctrine is to generate the debilitating uncertainty of which Judge Posner warned.¹⁷⁹

An antitrust-based misuse doctrine is not without problems. One commentator, in discussing patent misuse, notes that antitrust-based misuse creates two possible results: Either the limited monopoly is allowed to extend beyond the grant because there is no antitrust violation, or, if there is a violation, the grantee is subject to the stern penalties of the antitrust laws, including treble damages and attorney's fees.¹⁸⁰ A policy-based misuse doctrine, therefore, affords an escape valve from those penalties while it enforces the patent laws.¹⁸¹ In essence, because policy-based misuse is available, "[t]he full weight of the antitrust laws is not required" to enforce patent policy.¹⁸² This position, however, necessarily depends on a belief that the patent and antitrust laws have distinct, if not antagonistic, policy goals—a belief that fails to withstand analysis.¹⁸³

Another factor militating against antitrust-based misuse is the litigation expenses an infringer must incur when seeking to prove the grantee was in restraint of trade. The result of these exorbitant expenses is that infringers may not prosecute some antitrust actions against grantees in restraint of trade.¹⁸⁴ This scenario is especially true in the case of an infringer asserting a collateral antitrust defense. If the infringing defendant proves misuse because the grantee violated the antitrust laws, the defendant escapes infringement liability; because the defendant is not injured by reason of the antitrust violation, however, neither treble damages nor attorney's fees will be available.¹⁸⁵ Thus, there is less incentive to litigate the question,¹⁸⁶ although ways are available to avoid this dilemma.¹⁸⁷

One of the more troubling aspects of a policy-based misuse doctrine is that enforcement of the policy lies solely in the hands of a copyright infringer. If a public policy is so important that it warrants depriving a property owner of the court's protection against theft, why is a thief's conduct the only vehicle for

179. *USM Corp.*, 694 F.2d at 512. Not only is there less certainty due to the nature of a policy-based misuse, but copyright holders also will have to speculate whether each practice condemned under the patent misuse doctrine is equally vulnerable under copyright misuse. See *supra* notes 77-83 and accompanying text for the types of conduct constituting patent misuse. If the Fourth Circuit finds patent and copyright policy in virtual lockstep, may copyright holders now assume that courts will engage in a market power inquiry in copyright cases involving tying arrangements and package licensing, as is required only under the patent laws? See 35 U.S.C. § 271(d) (1988). Although precedent exists for importing statutory requirements under the patent laws into copyright doctrine, see *Sony Corp. v. Universal City Studios*, 464 U.S. 417, 434-42 (1984), *Lasercomb* leaves copyright owners with only a guess for whether this might be the case.

180. Calkins, *supra* note 4, at 187.

181. *Id.*

182. *Id.*

183. See *supra* notes 148-49 and accompanying text for a discussion of how the patent and antitrust laws reflect parallel procompetitive policies.

184. Calkins, *supra* note 4, at 187.

185. 15 U.S.C. § 15(a) (1988).

186. Note, *supra* note 5, at 1312 n.106.

187. One commentator suggests two possible solutions. One alternative is permitting the infringing defendant to join the antitrust victim as a codefendant, thereby covering the litigation costs should the antitrust counterclaim prevail. A second possibility is requiring that the government intervene to enforce the antitrust laws, which places the litigation expenses on the government. *Id.*

enforcing that policy? The pure copyright misuse theory holds that the public interest is harmed by improper enlargement of the copyright, but such conduct need not rise to the level of an antitrust violation. By varying the facts of *Lasercomb*, one can imagine a situation in which no party has standing to enforce this presumably important public policy.¹⁸⁸ A public policy that relies so heavily on happenstance and the willful wrongdoing of a private party to enforce that policy is not much of a policy at all.¹⁸⁹

Significantly, the Fourth Circuit did not attempt to address the issue whether collateral antitrust violations can trigger the misuse defense.¹⁹⁰ The court instead straddled the fence and found that misuse results from either antitrust or public policy violations.¹⁹¹ Because the Fourth Circuit's holding was grounded in a policy-based misuse, the court did not contemplate the validity of prior Fourth Circuit caselaw that bars collateral antitrust defenses to copyright infringement actions.¹⁹² Had it affirmed this principle, while at the same time deciding *Lasercomb* as it did, the court would have placed itself in an untenable position. In both cases the infringing party would be in a position collateral to the misuse, yet only the grantee who committed the more egregious offense—the antitrust violator—would still be allowed to enforce an infringement claim.

A better approach would have been to disavow this precedent and distinguish collateral antitrust defenses in cases involving contracts and copyrights. The former involve state-created rights while the latter involve rights created

188. Assuming the defendants had not broken the chronoguard and stolen the software for themselves, they would not have been liable for infringement and would never have asserted the copyright misuse defense; *Lasercomb* then would never have had the incentive to purge its misuse. If a court found that the noncompete clause withstood an antitrust attack, the clause would remain in force. At this point the only way to confine *Lasercomb* to its limited monopoly and discontinue the anticompetitive conduct would be to use copyright misuse as an affirmative attack on the copyright holder. The *Lasercomb* panel did not need to contemplate this notion, but one court recently rejected use of the copyright misuse doctrine as "a vehicle for affirmative relief." *Broadcast Music, Inc. v. Hearst/ABC Viacom Entertainment Servs.*, 746 F. Supp. 320, 328 (S.D.N.Y. 1990).

A party may seek a declaratory judgment on the question of copyright misuse. *See id.* at 327; *Columbia Broadcasting Sys. v. American Soc'y of Composers*, 400 F. Supp. 737, 771 (S.D.N.Y. 1975), *rev'd*, 562 F.2d 130 (2d Cir. 1977), *rev'd sub nom.* *Broadcast Music, Inc. v. Columbia Broadcasting Sys.*, 441 U.S. 1 (1979). Yet an affirmative judgment of misuse would permit a party to infringe with impunity only until the copyright holder purged the misuse. The result here is the same: enforcement of copyright policy is entirely dependent on one who willfully violates the copyright laws.

Judge Posner has also cast doubt on the importance of the policy that is meant to be protected. In seeking to explain a patent misuse case arising out of the Fourth Circuit. Judge Posner suggested that misuse doctrine, "unlike antitrust law, condemns any patent licensing practice that is even trivially anticompetitive, at least if it has no socially beneficial effects." *USM Corp. v. SPS Technologies, Inc.*, 694 F.2d 505, 511 (7th Cir. 1982). If so, one may wonder whether "*De minimis non curat lex*"—"the law does not care for trifling matters"—holds true in the Fourth Circuit.

189. In denying an antitrust defense to a breach of contract action, Justice Brennan has observed that "[i]f the defense of illegality is to be allowed as a collateral method of enforcement of the antitrust laws, . . . it must be said that [this] creates a very strange class of private attorneys general." *Kelly v. Kosuga*, 358 U.S. 516, 520 (1959). Given the holding in *Lasercomb*, the Fourth Circuit is in effect expanding this "very strange class" for the purpose of enforcing an ill-defined public policy.

190. *See supra* notes 132-35 and accompanying text.

191. *Lasercomb*, 911 F.2d at 978.

192. *M. Witmark & Sons v. Pastime Amusement Co.*, 298 F. 470, 480 (E.D.S.C.), *aff'd mem.*, 2 F.2d 1020 (4th Cir. 1924). For a discussion of *Pastime Amusement's* holding, *see supra* notes 133-35 and accompanying text.

under federal law. Any judicial repugnance towards injecting federal policy into disputes governed by state law, therefore, is not warranted in actions already concerning federal law.¹⁹³ Furthermore, the rationale that courts should not use the Sherman Act to divest a copyright holder of her property rights is unavailing in light of the temporary effect that misuse doctrine has upon a copyright. Once the grantee purges her misuse she regains all rights to her property. Moreover, while the remedies offered by the antitrust laws are said to be exclusive, no rational basis exists for preventing federal courts from employing a well-established body of legal analysis as a measuring device that is useful in enforcing equally important federal laws.¹⁹⁴ The result in allowing collateral antitrust violations to constitute misuse involves a trade-off. This choice requires infringers asserting a misuse defense to satisfy the requirements of the antitrust laws. It balances the burden of meeting this demand against permitting infringers who have not suffered antitrust injuries to raise the misuse defense. This in turn broadens the class of copyright holders who will be unable to pursue an infringement claim and should act as a deterrent to copyright abuse.

Although it is too early to gauge the impact of the Fourth Circuit's decision, it hardly can be welcome news for copyright holders.¹⁹⁵ It remains an open question, however, whether courts will construe *Lasercomb* narrowly or will give it widespread application in infringement actions involving computer software as well as in other fields.¹⁹⁶

193. *But see* Gibbs, *supra* note 5, at 70.

194. *See* Note, *supra* note 5, at 1314.

195. The *Lasercomb* decision already has attracted criticism outside the legal community. One commentator in the computer software field suggests that the result in *Lasercomb* is "ridiculous," and the product of a judiciary that fails to comprehend that computer software is unique; software is neither "like a book or a hammer," but an "entirely new beast." Gruenfeld, *Copyrights: You Abuse, You Lose*, INFORMATION WEEK, Mar. 18, 1991, at 52.

196. One recent software infringement action suggests that "shrink wrap" or "box top" licenses may make the licensor vulnerable to charges of copyright misuse. Shrink wrap licenses are often used in the marketing of consumer software. The license consists of a form attached to the exterior of the software packaging, which alerts consumers that opening the package binds the consumer to the terms of the license. 1 L. KUTTEN, COMPUTER SOFTWARE, § 8.03[1] (1991). The form then lists various provisions that seek to control the way the purchaser uses the software. *Id.*

In *Vault Corp. v. Quaid Software, Ltd.*, 847 F.2d 255 (5th Cir. 1988), the plaintiff copyright holder brought an infringement action against another software producer, who purportedly was bound by the terms of the plaintiff's shrink wrap license. In addition to the infringement action, the plaintiff asserted a claim that the defendant breached its license agreement by decompiling or disassembling the plaintiff's program in violation of a Louisiana statute. *Id.* at 258. That statute permitted licensors to restrict licensees from decompiling or disassembling computer programs, as well as enforcing a complete ban on copying the programs. *Id.* at 269. The Fifth Circuit found that such restrictions conflict with the Copyright Act, which permits archival copies and copies made as an essential step in the use of a computer program. *See* 17 U.S.C. § 117 (1988). The court then held that the Copyright Act preempted the Louisiana statute, rendering the license agreement unenforceable. *Vault*, 847 F.2d at 270.

The potential application of *Lasercomb*'s misuse doctrine is clear. When the copyright holder seeks to usurp rights granted under the Copyright Act, this action ostensibly enlarges the grantee's copyright beyond what was originally contemplated. If a court were to declare the grantee to have misused the copyright in a case like *Vault*, the irony is remarkable: a grantee who has relied on a state legislative expression of what the public policy should be concerning copyrighted programs now has no remedy against those who infringe.

In areas outside computer software, *Lasercomb*'s implications are less clear-cut. One could imagine a film or video producer licensing the reproduction and distribution rights to another, on the condition that the licensee agree not to enter into the business of film or video production. If the

Despite the uncertain reach of the Fourth Circuit's decision, the *Lasercomb* opinion is troubling in its flat rejection of the idea of defining copyright misuse according to antitrust doctrine. The Fourth Circuit chose not to inquire into the contested practice's actual effects. The result is that even though a licensing agreement may be only marginally anticompetitive, have a neutral effect, or in practice actually have procompetitive consequences,¹⁹⁷ it still will be harmful to the public interest if it does not fall squarely within the defined bounds of the statutory grant. This formalistic approach is out of step with modern economic analysis and even traditional equitable principles of balancing one party's culpability against the other's.¹⁹⁸ Moreover, opting for a vaguely defined public policy rationale rather than antitrust principles likely will generate "debilitating uncertainty" for copyright holders seeking to license their copyrights.¹⁹⁹ Although *Lasercomb* may be a landmark holding, its celebrity could well be short-lived. Given the recent congressional attempt to subsume the patent misuse defense under antitrust principles,²⁰⁰ it may not be long before the Fourth Circuit is legislatively preempted, at least far more quickly than the languid jurisprudential pace that transformed *Morton Salt* into *Lasercomb*.

SCOTT A. MISKIMON

licensee were to create an unauthorized derivative work (for example, a film sequel) in violation of the noncompete clause, *Lasercomb*'s misuse defense would bar any infringement action.

197. See *supra* notes 176-78 and accompanying text.

198. See *supra* notes 136-41 and accompanying text for the Second Circuit's balancing approach.

199. See *supra* note 179.

200. See *supra* note 76.

An Update on Contract Damages When the Landlord Breaches the Implied Warranty of Habitability: *Surratt v. Newton* and *Allen v. Simmons*

By enacting the Residential Rental Agreements Act in 1977,¹ the North Carolina General Assembly established an implied warranty of habitability in residential leases.² The Act, however, does not specify the damages that a tenant may recover if her landlord violates the implied warranty. Ten years after the enactment of the statute, in *Miller v. C.W. Myers Trading Post*,³ the North Carolina Court of Appeals adopted a "benefit of the bargain" formula for computing such damages.⁴ This measure of damages seeks to put the plaintiff in the position she would have attained if the defendant had performed his contract duty.⁵

In *Surratt v. Newton*⁶ and *Allen v. Simmons*⁷ the court of appeals added substance to the structure it had established in *Miller*. Most notably, the court placed a "rent paid" ceiling on the *Miller* formula for damages;⁸ defined the contract rights of tenants who ceased paying rent before vacating their rental units;⁹ and held that the actions of the breaching landlord may constitute unfair trade practices, for which the landlord may be liable for treble damages.¹⁰

This Note examines the court's opinions in *Surratt* and *Allen* in light of judicial precedent and the Residential Rental Agreements Act. The Note concludes that the holdings are reasonable, although not clearly expressed. The Note then offers three principles to guide future holdings, focusing primarily on reducing the uncertainty surrounding legal liability in this area.

Plaintiff Katherine Surratt rented a house in Winston-Salem, North Carolina from 1974 or 1975 through March 1987.¹¹ Defendant Jerry Newton was the rental agent and property manager responsible for the house during the last fourteen months of her tenancy.¹² Plaintiff alleged that defendant failed to respond to her many requests for repairs of defective conditions existing in the

1. Residential Rental Agreements Act, ch. 770, 1977 N.C. Sess. Laws 1006 (codified as amended at N.C. GEN. STAT. §§ 42-38 to -44 (1984)).

2. N.C. GEN. STAT. § 42-42(a) (1984). An implied warranty of habitability requires a landlord to deliver and maintain habitable premises. See R. SCHOSHINSKI, AMERICAN LAW OF LANDLORD AND TENANT § 3:16, at 127 (1980).

3. 85 N.C. App. 362, 355 S.E.2d 189 (1987).

4. *Id.* at 371, 355 S.E.2d at 194.

5. See E. FARNSWORTH, CONTRACTS § 12.1, at 840 (2d ed. 1990).

6. 99 N.C. App. 396, 393 S.E.2d 554 (1990).

7. 99 N.C. App. 636, 394 S.E.2d 478 (1990).

8. *Id.* at 642, 394 S.E.2d at 482; *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560.

9. *Allen*, 99 N.C. App. at 641-42, 394 S.E.2d at 482; *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560.

10. *Allen*, 99 N.C. App. at 645, 394 S.E.2d at 484.

11. *Surratt*, 99 N.C. App. at 399, 393 S.E.2d at 556.

12. *Id.* at 400, 393 S.E.2d at 556. For Ms. Surratt's entire tenancy before 1986, Paul Jeffrey Newton, d/b/a Newton Brothers, was the rental agent. He was also a defendant in Surratt's suit, but the trial court dismissed his appeal based on his failure to file a notice of appeal within ten days of the judgment. The court of appeals affirmed this dismissal. *Id.* at 400-03, 393 S.E.2d at 556-58.

house throughout her occupancy.¹³ These conditions included "electrical failures, flooding of sewage and water into the house, rodent infestation, and other deteriorating conditions throughout the house."¹⁴ Plaintiff ceased paying rent in November 1986, which precipitated a summary ejectment action by defendant.¹⁵ The court entered judgment against plaintiff, and she vacated the premises at the end of February or the beginning of March of 1987.¹⁶

In March 1987 Surratt appealed to the district court for a trial de novo and filed an answer asserting that she owed no rent because the defendants had failed to maintain the premises in a safe and habitable condition.¹⁷ In addition to moving to dismiss the summary ejectment claim, plaintiff counterclaimed for rent abatement and other consequential and actual damages.¹⁸ Defendant voluntarily dismissed the summary ejectment action and moved for summary judgment on plaintiff's counterclaims, but the court denied the motion and plaintiff prevailed on the merits of the counterclaims.¹⁹ Defendant then appealed to the North Carolina Court of Appeals.²⁰

Newton's appeal raised several issues concerning the nature and extent of landlord liability under the Residential Rental Agreements Act.²¹ Judge Eagles, writing for the court, first defined the extent of damages that a tenant can receive in a rent abatement action. The court disagreed with defendant's contention

Ms. Surratt reached a settlement with the owners of the rental house, who were also defendants in the original action. *Id.* at 399, 393 S.E.2d at 556.

13. *Id.* at 400, 393 S.E.2d at 556.

14. *Id.*

15. *Id.*

16. *Id.*

17. *Id.*

18. *Id.*

19. *Id.* at 400-01, 393 S.E.2d at 556.

20. *Id.* at 401, 393 S.E.2d at 557.

21. The Residential Rental Agreements Act provides in relevant part:

(a) The landlord shall:

(1) Comply with the current applicable building and housing codes, whether enacted before or after October 1, 1977, to the extent required by the operation of such codes; . . .

(2) Make all repairs and do whatever is necessary to put and keep the premises in a fit and habitable condition;

(3) Keep all common areas of the premises in safe condition; and

(4) Maintain in good and safe working order and promptly repair all electrical, plumbing, sanitary, heating, ventilating, air conditioning, and other facilities and appliances supplied or required to be supplied by him provided that notification of needed repairs is made to the landlord in writing by the tenant except in emergency situations.

N.C. GEN. STAT. § 42-42(a) (1984).

The court also rejected defendant Newton's argument that he was not a proper defendant. *Surratt*, 99 N.C. App. at 403-05, 393 S.E.2d at 558-59. The court held that Newton was a "landlord," defined for the purposes of the Act as "any owner and any rental management company, rental agency, or any other person having the actual or apparent authority of an agent to perform the duties imposed by this Article." N.C. GEN. STAT. § 42-40(3) (1984). The court distinguished the case from its earlier holding in *Collingwood v. General Elec. Real Estate Equities*, 89 N.C. App. 656, 659, 366 S.E.2d 901, 903 (1988), *rev'd in part on other grounds*, 324 N.C. 63, 376 S.E.2d 425 (1989), that a property manager is not liable for defects in design and construction. *Surratt*, 99 N.C. App. at 404-05, 393 S.E.2d at 558-59.

that he was liable for no more than the amount of his rental commission.²² Instead, the court based the damages recoverable on the damages formula it had adopted in *Miller*:

"[A] tenant may recover damages in the form of a rent abatement calculated as the difference between the fair rental value of the premises if as warranted (i.e. in full compliance with G.S. 42-42(a)) and the fair rental value of the premises in their unfit condition for any period of the tenant's occupancy during which the finder of fact determines the premises were uninhabitable, plus any special or consequential damages alleged and proved."²³

The court qualified this rent abatement formula, however, by declaring that "the amount of rent paid is a limit on recovery."²⁴ Thus, the *Surratt* court employed a literal reading of *Miller*, which had found a cause of action "for recovery of rent paid."²⁵ Thus, a more precise statement of the rent abatement formula is: the lesser of (1) fair rental value as warranted less fair rental value "as is," and (2) the rent paid during the period that the premises were uninhabitable.

In addition, the *Surratt* court addressed the nature of the tenant's responsibilities in the following three aspects of a rent abatement claim: The requirement of notice to the landlord concerning defective conditions; the evidence used to establish fair rental values; and the consequence of withholding rent. Defendant argued that a tenant cannot recover for defects that fall under section 42-42(a)(4), which includes "electrical, plumbing, sanitary, heating, ventilating, air conditioning, and other facilities and appliances,"²⁶ if she has failed to provide the landlord with the written notice required under that subsection.²⁷ The court rejected that argument, noting that the jury had found that the orally requested repairs were needed to put the house in "fit and habitable condition."²⁸ Because section 42-42(a)(2), which requires the landlord to "put and keep the premises in a fit and habitable condition,"²⁹ does not explicitly require written notice from the tenant, the court held that "where the conditions enumerated in G.S. 42-42(a)(4) are the same conditions that render the premises unfit and uninhabitable no written notice is required under the statute."³⁰

Defendant next argued that the damages awarded were contrary to the evidence of fair market value produced at trial.³¹ In response, the court noted that defendant himself had testified that the house had a fair rental value of \$600 per

22. *Surratt*, 99 N.C. App. at 406-07, 393 S.E.2d at 560 ("No lesser measure of damages is recoverable against a landlord (as defined by G.S. 42-40(3)) merely because he is not the owner but is an agent.").

23. *Id.* at 406, 393 S.E.2d at 560 (quoting *Miller v. C.W. Myers Trading Post*, 85 N.C. App. 362, 371, 355 S.E.2d 189, 194 (1987)).

24. *Id.*

25. *Id.* at 407, 393 S.E.2d at 560 (quoting *Miller*, 85 N.C. App. at 368, 355 S.E.2d at 193).

26. N.C. GEN. STAT. § 42-42(a)(4) (1984); see *supra* note 21.

27. *Surratt*, 99 N.C. App. at 405, 393 S.E.2d at 559.

28. *Id.*

29. N.C. GEN. STAT. § 42-42(a)(2).

30. *Surratt*, 99 N.C. App. at 405-06, 393 S.E.2d at 559.

31. *Id.* at 408, 393 S.E.2d at 561. Specifically, he contended that plaintiff had failed to prove that the fair rental value of the house was different from the amount of rent charged. *Id.*

month and that "it only rented for less because of the nature of the neighborhood."³² Noting also that the plaintiff testified to an "as is" rental value of between \$100 and \$150, the court held the testimony to be sufficient for the jury to determine damages.³³

Finally, the court agreed with defendant that the jury could not award rent abatement damages for the period that plaintiff had withheld rent payments, although it rejected his rationale for this conclusion.³⁴ Defendant argued that the Residential Rental Agreements Act's provision that "[t]he tenant may not unilaterally withhold rent prior to a judicial determination of a right to do so"³⁵ precluded rent abatement during the period of rent withholding.³⁶ Instead of accepting this argument, the court simply characterized rent abatement as a refund of all or part of rent actually paid and held, in effect, that one cannot refund an amount never paid.³⁷ The court noted, however, that there is "nothing in the Act" to preclude entirely a tenant from collecting damages because she withholds rent; she just may not collect damages for the period of the withholding.³⁸

Significantly, the *Surratt* court split on the fundamental question of exactly how rent abatement relates to an action for breach of the warranty of habitability. Judge Eagles distinguished an action for rent abatement from an action for breach of the warranty of habitability.³⁹ He stated, "Since the pleadings here pray for relief in rent abatement and do not seek damages for breach of the covenant of habitability, we expressly decline to address here the issue of whether damages for the breach of a covenant of habitability are limited to the amount of rent paid."⁴⁰ Judge Greene, who concurred in the result, took issue with the court's "suggestion . . . that an action in rent abatement somehow differs from an action for breach of warranty for habitability."⁴¹ He characterized rent abatement as only one of the remedies available for a breach of the warranty of habitability.⁴²

In *Allen* the court of appeals reviewed a fact pattern similar to that in *Surratt*. Defendant Warnell Simmons rented a house from Scott Realty, the agent for plaintiff Harvey Allen, after Scott Realty allegedly agreed to make specified

32. *Id.*

33. *Id.*

34. *Id.* at 407, 393 S.E.2d at 560.

35. N.C. GEN. STAT. § 42-44(c) (1984).

36. *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560.

37. *Id.*

38. *Id.* In an opinion concurring in the result, Judge Greene went further in explaining the position of a tenant who has withheld rent in response to the landlord's failure to provide fit premises. Judge Greene stated that such a tenant, as a defendant in a summary ejectment action, can seek an abatement of rent due. Under those circumstances, however, Judge Greene would modify the court's rent abatement formula by replacing the "as warranted" term with "agreed rent." *Id.* at 411, 393 S.E.2d at 562 (Greene, J., concurring in result).

39. *Id.* at 409, 393 S.E.2d at 561.

40. *Id.*

41. *Id.* at 411, 393 S.E.2d at 562 (Greene, J., concurring in result).

42. *Id.* (Greene, J., concurring in result). Judge Greene identified special and consequential damages as two other remedies for a breach of the warranty. *Id.* (Greene, J., concurring in result).

repairs prior to occupancy.⁴³ Defendant alleged that Scott Realty failed to make the repairs before she moved into the house in November 1985 and failed to correct defects brought to its attention throughout her tenancy.⁴⁴ Simmons occupied the house until July 1987, although she ceased paying rent after August 1986.⁴⁵ Defendant's initial rent withholding occurred concurrently with the Winston-Salem Housing Services Department's declaration, issued on September 5, 1986, that the house was unfit for human habitation.⁴⁶

Allen brought a summary ejectment action and prevailed in an April 1987 hearing in magistrate's court.⁴⁷ Defendant then appealed to the district court and, in addition, counterclaimed for rent abatement and damages for fraud, intentional infliction of emotional distress, and unfair and deceptive trade practices.⁴⁸ The district court entered directed verdicts on all of the counterclaims against the defendant, and she appealed.⁴⁹ The court of appeals affirmed the trial court's directed verdicts on the fraud⁵⁰ and intentional infliction of emotional distress⁵¹ counterclaims, but reversed the trial court on the other counts, finding sufficient evidence to reach the jury on the rent abatement⁵² and unfair trade practice claims.⁵³

In the majority opinion, Judge Eagles reiterated his *Surratt* position that, section 42-44(c) notwithstanding,⁵⁴ a tenant's withholding of rent does not bar him from recovering damages, except rent abatement for the period of withholding.⁵⁵ His discussion of rent abatement, including the formula for determining damages, mirrored his *Surratt* opinion.⁵⁶ Judge Eagles again distinguished between an action for rent abatement and an action for breach of the implied war-

43. *Allen*, 99 N.C. App. at 638, 394 S.E.2d at 480. "The defects included: holes in some of the walls; a damaged faucet on the kitchen sink; electrical problems, plumbing that leaked in the bathroom and in the basement; a damaged commode; a damaged hot water heater; fleas; broken glass; and no furnace." *Id.*

44. *Id.* These alleged ongoing problems included pipes bursting from lack of heat, rats entering through holes in the walls, and a fire caused by defective wires. The defendant testified that a furnace was not installed until the day before she vacated the house. *Id.* at 638-39, 394 S.E.2d at 480.

45. *Id.* at 638, 394 S.E.2d at 480.

46. *Id.* at 639, 394 S.E.2d at 480.

47. *Id.* at 638, 394 S.E.2d at 480.

48. *Id.*

49. *Id.* at 640, 394 S.E.2d at 481.

50. *Id.* at 643, 394 S.E.2d at 483 (finding "no evidence that at the time of his promise plaintiff intended not to make the repairs" he promised to make).

51. *Id.* at 646, 394 S.E.2d at 484 (finding no evidence of serious mental distress or other bodily harm).

52. *Id.* at 642, 394 S.E.2d at 482.

53. *Id.* at 645, 394 S.E.2d at 484. Unfair trade practices are addressed in N.C. GEN. STAT. § 75-1.1 (1988). Subsection (a) provides: "Unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are declared unlawful." *Id.* § 75-1.1(a). The statute defines commerce broadly to include "all business activities," *id.* § 75-1.1(b), although it expressly excludes services rendered by someone in a "learned profession" and certain activities of advertising media. *Id.* § 75-1.1(b), (c).

54. The Residential Rental Agreements Act provides that the "tenant may not unilaterally withhold rent." N.C. GEN. STAT. § 42-44(c) (1984); see *supra* notes 35-38 and accompanying text.

55. *Allen*, 99 N.C. App. at 642, 394 S.E.2d at 482.

56. See *id.* at 641-42, 394 S.E.2d at 482; *supra* notes 22-40 and accompanying text.

ranty of habitability.⁵⁷ "Tenants may bring an action seeking damages for breach of the implied warranty of habitability and may also seek rent abatement for their landlord's breach of the statute,"⁵⁸ he stated.

The court also found that defendant had presented evidence from which a jury could find that plaintiff had committed an unfair trade practice under section 75-1.1.⁵⁹ After determining that Simmons and Allen were proper parties to such an action,⁶⁰ the court noted that Simmons provided evidence that plaintiff failed to respond to numerous notices concerning the unfit and uninhabitable condition of the house.⁶¹ The court placed special significance on the fact that, "[d]espite the unfit conditions of the house, Scott Realty attempted to collect rent."⁶² It found that "[Allen's] behavior [could] be considered 'immoral, unethical, oppressive, unscrupulous, or substantially injurious to consumers.'"⁶³ Evidence that Simmons suffered additional expenses because of the unfit conditions was the final element cited by the court in finding that a jury should have determined the unfair trade practice issue.⁶⁴

Surratt and *Allen* are the latest chapters in North Carolina landlord and tenant law as it has evolved since the enactment of the Residential Rental Agreements Act in 1977. Prior to the Act, the common-law doctrine of caveat emptor applied to residential leases in North Carolina and absolved landlords of any duty to maintain and repair the leased premises.⁶⁵ Courts in the United States applied this common-law concept into the 1960s.⁶⁶ By the late 1970s, approximately one-third of the states had moved away from caveat emptor by judicial creation of landlord duties under an implied warranty of habitability.⁶⁷ Also during this period, adoption of the Uniform Residential Landlord and Tenant Act (URLTA),⁶⁸ which includes an implied warranty of habitability, established a statutory implied warranty in an equivalent number of states.⁶⁹ Courts now

57. *Allen*, 99 N.C. App. at 641, 394 S.E.2d at 482.

58. *Id.*

59. *Id.* at 645, 394 S.E.2d at 484.

60. *Id.* at 644, 394 S.E.2d at 483. Plaintiff was a physician who owned the house rented by Simmons and retained Scott Realty as rental agent. Plaintiff-Appellee's Brief at 9, *Allen* (No. 8921DC1155). Dr. Allen argued that he was not in the business of real estate. *Id.* Finding Scott Realty to be in the business of real estate, however, the court used the concept of agency to find Allen to be a proper defendant in the unfair trade practice action. See *Allen*, 99 N.C. App. at 644, 394 S.E.2d at 483.

61. *Allen*, 99 N.C. App. at 644, 394 S.E.2d at 484.

62. *Id.* at 644-45, 394 S.E.2d at 484. "[P]laintiff even went to defendant's house in February 1987 in an effort to collect past due rent for the unfit house." *Id.*

63. *Id.* at 645, 394 S.E.2d at 484 (quoting *Mosley & Mosley Builders, Inc. v. Landin Ltd.*, 97 N.C. App. 511, 517, 389 S.E.2d 576, 579, *disc. rev. denied*, 326 N.C. 801, 393 S.E.2d 898 (1990)).

64. *Id.*

65. Fillette, *North Carolina's Residential Rental Agreements Act: New Developments for Contract and Tort Liability in Landlord-Tenant Relations*, 56 N.C.L. REV. 785, 785 (1978).

66. See R. CUNNINGHAM, W. STOEBUCK & D. WHITMAN, *THE LAW OF PROPERTY* § 6.38, at 313 (1984).

67. See *id.* at 314.

68. UNIF. RESIDENTIAL LANDLORD AND TENANT ACT, 7B U.L.A. 427 (1985). The North Carolina Residential Rental Agreements Act is patterned after the Uniform Act. See Residential Rental Agreements Act, ch. 770, 1977 N.C. Sess. Laws 1006 (codified as amended at N.C. GEN. STAT. §§ 42-38 to -44 (1984)).

69. R. SCHOSHINSKI, *supra* note 2, § 3:31, at 152.

recognize an implied warranty of habitability in residential leases in all but a handful of jurisdictions.⁷⁰

Generally, when a landlord breaches the implied warranty of habitability, the tenant is entitled to vacate the premises and terminate the lease.⁷¹ It is not necessary for a rental unit to be "literally uninhabitable" for a breach to occur.⁷² Thus, a tenant can seek relief for the landlord's breach while continuing in occupancy of an unfit unit, or she may seek relief after vacating the premises for the period during which she occupied the premises.⁷³

Courts have developed at least three different formulas to calculate the damages suffered by a tenant occupying an unfit rental unit.⁷⁴ First, courts can measure the damages as the difference between the fair rental value of the premises as warranted and the fair rental value of the premises in their unfit condition, or "as is."⁷⁵ This approach seeks to compensate the tenant for the loss of her bargain.⁷⁶ Second, courts may compute the damages as the difference between the agreed rent and the "as is" fair rental value.⁷⁷ This view effectively requires that the tenant pay the fair rental value of the defective premises.⁷⁸ Third, the damages may be a percentage of the agreed rent representing the percentage reduction in the tenant's use and enjoyment of the premises resulting from the breach.⁷⁹ In a modified version of the third formula, the court uses the "as warranted" and "as is" fair market values in computing the percentage to be applied to the agreed rent in calculating damages.⁸⁰ When the court computes damages as a percentage of agreed rent, the resulting reduction in the rent obligation commonly is called a rent abatement.⁸¹

In addition to collecting damages for the landlord's breach, the tenant may assert a breach of the implied warranty of habitability as a defense to an action for rent.⁸² This defense is a consequence of the tenant's right to terminate the lease following the landlord's breach,⁸³ which precludes the landlord from collecting the full agreed rent during a period when unfit conditions constituting a

70. See *id.* § 3:16, at 67 n.30 (Supp. 1990) (implied warranty of habitability not recognized by courts in Alabama, Florida, Idaho, Kentucky, Oregon, and South Carolina).

71. R. CUNNINGHAM, W. STOEBUCK & D. WHITMAN, *supra* note 66, § 6.41, at 333. This entitlement derives from the mutuality of landlord and tenant obligations under a lease. *Id.* Termination of a lease in this manner is "very similar to a traditional 'constructive eviction' based on the landlord's failure to perform an express covenant to maintain the premises." *Id.* at 333 n.3.

72. R. SCHOSHINSKI, *supra* note 2, § 3:17, at 68 (Supp. 1990).

73. See *id.*

74. R. CUNNINGHAM, W. STOEBUCK & D. WHITMAN, *supra* note 66, § 6.42, at 336.

75. R. SCHOSHINSKI, *supra* note 2, § 3:25, at 141-42.

76. *Id.* at 141.

77. *Id.* at 142.

78. *Id.*

79. R. CUNNINGHAM, W. STOEBUCK & D. WHITMAN, *supra* note 66, § 6.42, at 337.

80. R. SCHOSHINSKI, *supra* note 2, § 3:25, at 142-43. It is not obvious that this modified formula produces a different measure of damages than the formula it modifies.

81. *Id.* Rent is abated when damages constitute an effective refund of prior rent payments or when a tenant who has withheld rent is relieved of the obligation to pay all or part of past due rent. BLACK'S LAW DICTIONARY 4 (6th ed. 1990).

82. R. SCHOSHINSKI, *supra* note 2, § 3:22, at 136.

83. See *supra* note 71 and accompanying text.

breach exist.⁸⁴ Indeed, in the landmark implied warranty case of *Javins v. First National Realty Corp.*⁸⁵ the United States Court of Appeals for the District of Columbia Circuit went even further and held that a breach-of-implied-warranty defense could be used not only in an action for nonpayment of rent, but also in an action for possession.⁸⁶

The first recognition of an implied warranty of habitability in North Carolina was judicial.⁸⁷ In 1974 the North Carolina Supreme Court held that such a warranty applies to the sale of newly constructed houses.⁸⁸ The following year, however, the court of appeals refused to recognize an implied warranty of habitability in residential leases.⁸⁹ The North Carolina General Assembly countered that decision by enacting the Residential Rental Agreements Act in 1977.⁹⁰ The Act provides that the tenant's obligations to pay rent and maintain clean premises and the landlord's obligation to provide fit premises are mutually dependent.⁹¹ The landlord's obligation includes specific responsibilities concerning building and housing code compliance, common-area maintenance, and maintenance of electrical and other facilities, and a general responsibility to "[m]ake all repairs and do whatever is necessary to put and keep the premises in a fit and habitable condition."⁹² The statute does not delineate specific remedies, but provides that "[a]ny right or obligation declared by this Chapter is enforceable by civil action, in addition to other remedies of law and in equity."⁹³

Because the Act lacks specificity concerning a tenant's remedies for landlord violations of the Act, the North Carolina Court of Appeals began filling the void. The initial cases before the court of appeals were tort actions involving personal injury and wrongful death.⁹⁴ In those cases, the court found landlords to be in violation of the Act in instances of dimly lit steps in a common area,⁹⁵ failure to repair steps of the rental unit,⁹⁶ and failure to repair a heating flue.⁹⁷

84. R. SCHOSHINSKI, *supra* note 2, § 3:21, at 136.

85. 428 F.2d 1071 (D.C. Cir.), *cert. denied*, 400 U.S. 925 (1970).

86. *Id.* at 1082.

87. *Hartley v. Ballou*, 286 N.C. 51, 62, 209 S.E.2d 776, 783 (1974).

88. *Id.* (implied warranty that house is free of major structural defects and satisfies workmanlike quality standard).

89. *Knuckles v. Spough*, 26 N.C. App. 340, 340, 215 S.E.2d 825, 826, *cert. denied*, 288 N.C. 241, 217 S.E.2d 665 (1975).

90. Residential Rental Agreements Act, ch. 770, 1977 N.C. Sess. Laws 1006 (codified as amended at N.C. GEN. STAT. §§ 42-38 to -44 (1984)).

91. N.C. GEN. STAT. § 42-41 (1984).

92. *Id.* § 42-42(a). For a further explanation of § 42-42(a), see *supra* note 21.

93. N.C. GEN. STAT. § 42-44(a) (1984). The other two provisions of this section, however, limit the right to recover for violations of the Act by providing that "[t]he tenant may not unilaterally withhold rent prior to a judicial determination of a right to do so" and that "[a] violation of this Article shall not constitute negligence per se." *Id.* § 42-44(c), (d).

94. See *Jackson v. Housing Auth.*, 73 N.C. App. 363, 364, 326 S.E.2d 295, 296 (1985), *aff'd*, 316 N.C. 259, 341 S.E.2d 523 (1986); *Brooks v. Francis*, 57 N.C. App. 556, 557, 291 S.E.2d 889, 890 (1982); *Allen v. Equity & Investors Mgt. Corp.*, 56 N.C. App. 706, 706, 289 S.E.2d 623, 624 (1982); *O'Neal v. Kellett*, 55 N.C. App. 225, 226, 284 S.E.2d 707, 709 (1981); *Lenz v. Ridgewood Assocs.*, 55 N.C. App. 115, 116, 284 S.E.2d 702, 703 (1981), *disc. rev. denied*, 305 N.C. 300, 290 S.E.2d 702 (1982).

95. *O'Neal*, 55 N.C. App. at 228, 284 S.E.2d at 710.

96. *Brooks*, 57 N.C. App. at 559-60, 291 S.E.2d at 891.

97. *Jackson*, 73 N.C. App. at 369, 326 S.E.2d at 299.

In *Miller*,⁹⁸ decided in 1987, the court of appeals first addressed the issue of contract remedies under the Residential Rental Agreements Act. The plaintiffs rented a house from defendant for over six years before bringing suit and sought recovery of rent paid for premises that defendant allegedly failed to maintain in a fit and habitable condition as required by the Act.⁹⁹ In reversing the trial court's grant of summary judgment for defendant, the *Miller* court held that the Act provides "an affirmative cause of action to a tenant for recovery of rent paid based on the landlord's noncompliance with [section] 42-42(a)."¹⁰⁰ The court based its damages formula on two premises. First, the court stated, "[t]he implied warranty of habitability entitles a tenant in possession of leased premises to the value of the premises *as warranted*, which may be greater than the rent agreed upon or paid."¹⁰¹ Second, the court ruled that a tenant in possession of defective housing is "liable only for the reasonable value, if any, of his use of the property in its defective condition."¹⁰² The court then declared the formula for damages "in the form of a rent abatement" to be the difference between the rental unit's fair rental value as warranted and its fair rental value "as is".¹⁰³

In another 1987 case, *Cotton v. Stanley*,¹⁰⁴ the court commented further on the application of the *Miller* formula for computing damages arising from a landlord's breach.¹⁰⁵ The court held that the illegality of renting housing that violates a housing code does not automatically establish a fair market value of zero for such housing in possession of a plaintiff tenant.¹⁰⁶ Dismissing the defendant's contention that plaintiffs must produce direct evidence of the two fair rental values in the damage formula, however, the court held that "[t]he fair rental value of property may be determined 'by proof of what the premises would rent for in the open market, or by evidence of other facts from which the fair rental value of the premises may be determined.'"¹⁰⁷ The court found that the rent to which the parties agreed was nonbinding evidence of the fair rental value as warranted, and that the jurors' personal experience with housing together with descriptions of the premises from plaintiffs and a building inspector were sufficient for the jury to determine a fair rental value "as is".¹⁰⁸ Thus,

98. *Miller v. C.W. Myers Trading Post*, 85 N.C. App. 362, 355 S.E.2d 189 (1987).

99. *Id.* at 364-65, 355 S.E.2d at 190-91. For the landlord's maintenance obligations under the Act, see *supra* note 21.

100. *Miller*, 85 N.C. App. at 368, 355 S.E.2d at 193. The court stated that "[t]he action for a rent abatement for breach of an implied warranty is wholly contractual." *Id.* at 371, 355 S.E.2d at 195.

101. *Id.* at 370, 355 S.E.2d at 194. The court explicitly held that renting unfit housing at a "fair" rental value for that property would not free a landlord from obligations under the Residential Rental Agreement Act. *Id.*

102. *Id.* at 370-71, 355 S.E.2d at 194.

103. *Id.* at 371, 355 S.E.2d at 194; see *supra* note 23 and accompanying text.

104. 86 N.C. App. 534, 358 S.E.2d 692, *disc. rev. denied*, 321 N.C. 296, 362 S.E.2d 779 (1987).

105. The court of appeals also affirmed the right of a group of tenants to bring a class action for breach of the implied warranty of habitability. *Id.* at 539, 358 S.E.2d at 696.

106. *Id.* at 538, 358 S.E.2d at 695. "The measure of the unit's fair rental value is not the price at which the owner could lawfully rent the unit to a new tenant in the open market, but the price at which he could rent it if it were lawful for him to do so." *Id.*

107. *Id.* at 539, 358 S.E.2d at 695 (quoting *Brewington v. Loughran*, 183 N.C. 558, 565, 112 S.E. 257, 260 (1922)) (emphasis added in *Cotton*).

108. *Id.*

Cotton provided broad leeway regarding the evidence that a tenant can introduce for the valuation of damages under the *Miller* formula.

Any action potentially affecting the damages recoverable in a consumer suit may also involve a claim of unfair trade practices.¹⁰⁹ The court of appeals has held that the rental of residential housing is in the nature of "trade or commerce" within the meaning of section 75-1.1 of the North Carolina General Statutes, the unfair trade practice statutes.¹¹⁰ In one case, a state district court judge found that the defendants committed an unfair trade practice when they "violated the North Carolina Residential Rental Agreements Act and the Tenant Security Deposit Act."¹¹¹ The trial judge assessed damages of \$240, which was the amount of the security deposit, and trebled the damages to \$720.¹¹²

In *Surratt* and *Allen* the North Carolina Court of Appeals added a rent-paid ceiling to its previously stated formula for contract damages arising from a landlord's violation of the implied warranty of habitability, while addressing for the first time the effect of rent withholding on the tenant's rights in an implied warranty action. The *Surratt* court established the rent paid by the tenant as a cap on the damages otherwise collectible under the court's "benefit of the bargain" formula.¹¹³ The court determined that it had prescribed such a cap by its prior characterization of a tenant's action as "'recovery of rent paid.'" ¹¹⁴

The *Surratt* court's imposition of an agreed rent ceiling on the damages formula might have been influenced by the *Miller* court's use of the "rent abatement" label in the formula.¹¹⁵ Generally, rent abatement and benefit of the bargain measures are distinct implied warranty damage measures.¹¹⁶ One might view the court's new formula, however, as a merger of the two concepts. In this merger, the court has chosen to use agreed rent as a variable in the "rent abatement" formula while abandoning the "percentage of rent" damage measure nor-

109. N.C. GEN. STAT. § 75-1.1 (1988). The court will award treble damages when those damages result from an unfair trade practice. *Id.* § 75-16.

110. *Love v. Pressley*, 34 N.C. App. 503, 516, 239 S.E.2d 574, 583 (1977), *disc. rev. denied*, 294 N.C. 441, 241 S.E.2d 843 (1978). For a discussion of the unfair trade practices statute, see *supra* note 53.

111. *Borders v. Newton*, 68 N.C. App. 768, 769, 315 S.E.2d 731, 731 (1984) (renting a dwelling that was under a condemnation order). The court of appeals dealt only indirectly with the unfair trade practice concept, ruling that the plaintiffs could not recover fraud damages because they had already recovered treble damages for the same conduct under the unfair trade practices statute. The court thereby implied that a violation of the Residential Rental Agreements Act could constitute an unfair trade practice. See *id.* at 770, 315 S.E.2d at 732.

The Tenant Security Deposit Act, ch. 914, 1977 N.C. Sess. Laws 1237, is codified as amended at N.C. GEN. STAT. § 42-50 (1984).

112. *Borders*, 68 N.C. App. at 770, 315 S.E.2d at 731.

113. *Surratt*, 99 N.C. App. at 406, 393 S.E.2d at 560; see *supra* text accompanying notes 23-25.

114. *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560 (quoting *Miller v. C.W. Myers Trading Post*, 85 N.C. App. 362, 368, 355 S.E.2d 189, 193 (1987)). To at least one commentator, however, such a conclusion is not inevitable. See Note, *Miller v. C.W. Myers Trading Post: North Carolina Adopts Expansive Tenant Remedies for Violations of the Implied Warranty of Habitability*, 66 N.C.L. REV. 1276, 1285 (1988) ("[T]he actual rent paid is not considered in the damage formula" in *Miller*).

115. *Miller*, 85 N.C. App. at 366, 355 S.E.2d at 191; see *supra* text accompanying note 23.

116. See R. SCHOSHINSKI, *supra* note 2, § 3:25, at 141-43.

maximally associated with rent abatement.¹¹⁷ If the "as warranted" rental value exceeds the agreed rent, the tenant might recover damages in excess of those awarded under a traditional rent abatement formula, but, as under traditional rent abatement, damages will not exceed the agreed rent.

Use of an "as warranted" benchmark in computing damages is consistent with the *Miller* court's concept that the implied warranty of habitability gives the tenant a statutory entitlement to habitable premises.¹¹⁸ This recognition of the tenant's legitimate expectation interest¹¹⁹ parallels the Uniform Commercial Code's basic contract principle that "the aggrieved party [should] be put in as good a position as if the other party had fully performed."¹²⁰ The "as warranted" measure of damages raises the possibility that the damages would exceed the agreed rent when the landlord and tenant agree to the rental of unfit housing at a rent significantly below the "as warranted" value.¹²¹ One might view this result as a reasonable consequence of renting unfit housing or, alternatively, as "the patently absurd result that the landlord would have to pay the tenant for occupying the unit."¹²² Under similar circumstances, a damage measure of agreed rent minus fair rental value "as is" would yield zero damages, "at least where the [unfit] condition of the dwelling unit has not worsened since the beginning of the tenancy."¹²³ This result would appear to be a clear violation of section 42-42(b), which bars an explicit or implicit waiver of the implied warranty by the tenant.¹²⁴

The court's agreed-rent cap on the damage formula is a compromise between an unrestricted benefit of the bargain formula and a formula providing damages only to the extent that the "as is" value falls below the agreed rent. The court's seemingly flexible evidence requirements for proving fair rental values in *Cotton*¹²⁵ and *Surratt*¹²⁶ create for landlords uncertainty concerning their potential liability, even under the *Surratt* compromise formula. To reduce this uncertainty by requiring expert testimony would add to the litigation costs of tenant-plaintiffs, many of whom have low incomes.¹²⁷ The percentage diminution approach, a rent abatement damage measure based on the percentage reduction in use and enjoyment resulting from the landlord's breach applied to the agreed rent, eliminates much of the uncertainty and cost of using fair rental

117. *Id.* at 142-43.

118. See *Miller*, 85 N.C. App. at 370, 355 S.E.2d at 194 (the tenant is entitled to the as-warranted value of the leased housing).

119. E. FARNSWORTH, *supra* note 5, § 12.1, at 840.

120. U.C.C. § 1-106(1) (1989).

121. R. CUNNINGHAM, W. STOEBUCK & D. WHITMAN, *supra* note 66, § 6.42, at 338.

122. *Id.*

123. *Id.* at 337.

124. N.C. GEN. STAT. § 42-42(b) (1984) ("The landlord is not released of his obligations under any part of this section by the tenant's explicit or implicit acceptance of the landlord's failure to provide premises complying with this section . . .").

125. See *Cotton v. Stanley*, 86 N.C. App. 534, 539, 358 S.E.2d 692, 695, *disc. rev. denied*, 321 N.C. 296, 362 S.E.2d 779 (1987); *supra* text accompanying notes 104-08.

126. See *Surratt*, 99 N.C. App. at 408, 393 S.E.2d at 561; *supra* text accompanying notes 31-33.

127. See R. SCHOSHINSKI, *supra* note 2, § 3:25, at 143.

values.¹²⁸

A broader perspective from which to evaluate damages for breach of the implied warranty of habitability would be that of public housing policy. One commentator, for example, has argued against adopting a percentage diminution approach by asserting that it would not discourage the practice of renting unfit housing at rents below as-warranted values.¹²⁹ From a policy standpoint, the potential tradeoff resulting from alternative damage formulas rests between a higher rate of compliance with the implied warranty obligations, obtained by using a liberal damages formula, and a more ample stock of housing for low income persons, albeit with a lower warranty compliance rate, obtained through a damages formula rendering more modest awards.¹³⁰ Because the Residential Rental Agreements Act is only one small piece in the array of federal, state, and local housing laws and programs, however, the North Carolina courts are better situated to apply contract law principles in fashioning a damages formula than to incorporate economic analysis of the housing market into their choice of remedies.¹³¹

The extent to which the *Surratt* damages formula fills in the picture concerning tenant remedies for landlord breach is diminished further by Judge Eagles's distinction between rent abatement and damages for breach of implied warranty.¹³² Although Judge Greene appropriately criticized this distinction in his concurring opinion in *Surratt*,¹³³ Judge Eagles stated the distinction in even stronger terms in his opinion for the *Allen* court.¹³⁴ Perhaps Judge Eagles simply was referring to the concluding phrase of the *Miller* damage formula statement, which allowed, in addition to rent abatement, "any special or consequential damages alleged and proved."¹³⁵ Subsequent opinions have quoted this provision,¹³⁶ but an alternative interpretation of Judge Eagles's opinion is that the court remains open to alternatives or additions to its rent abatement formula.

A potential avenue for expanding tenant remedies presents itself when a landlord's actions in violation of the Residential Rental Agreements Act rise to the level of an unfair trade practice. In *Borders v. Newton*¹³⁷ the court of ap-

128. R. CUNNINGHAM, W. STOEBCUK & D. WHITMAN, *supra* note 66, § 6.42, at 338.

129. Note, *Property Law—A Fresh Look at Contractual Tenant Remedies Under the North Carolina Residential Rental Agreements Act*—Miller v. C.W. Myers Trading Post, Inc., 10 CAMPBELL L. REV. 167, 191 (1987).

130. See Note, *supra* note 114, at 1287-89.

131. Government income maintenance programs, such as rent subsidy programs and direct income redistribution programs, also enter into the broader housing policy milieu. See *id.* at 1289.

132. *Surratt*, 99 N.C. at 409, 393 S.E.2d at 561 (claim was for "rent abatement and . . . not . . . damages for breach of the covenant of habitability"); see *supra* text accompanying notes 39-40.

133. *Surratt*, 99 N.C. at 411, 393 S.E.2d at 562 (Greene, J., concurring in result) (rent abatement is a measure of damages for breach of implied warranty); see *supra* text accompanying notes 41-42.

134. *Allen*, 99 N.C. App. at 641, 394 S.E.2d at 482 (raising possibility that a tenant bringing action for breach of implied warranty "may also seek rent abatement"); see *supra* text accompanying notes 57-58.

135. Miller v. C.W. Myers Trading Post, 85 N.C. App. 362, 371, 355 S.E.2d 189, 194 (1987).

136. *Allen*, 99 N.C. App. at 641, 394 S.E.2d at 482; *Surratt*, 99 N.C. App. at 406, 393 S.E.2d at 560; Cotton v. Stanley, 86 N.C. App. 534, 537, 358 S.E.2d 692, 694 (1987).

137. 68 N.C. App. 768, 315 S.E.2d 731 (1984); see *supra* notes 111-12 and accompanying text.

peals left undisturbed a district court holding in support of such an action.¹³⁸ Similarly, the *Allen* court found that the tenant had presented evidence from which a jury could find an unfair trade practice.¹³⁹ The court left unanswered, however, the question of which damages would be subject to the treble damages provision of section 75-16.¹⁴⁰ The district court in *Borders* trebled a security deposit as damages.¹⁴¹ Arguably, the court-ordered return of a security deposit is analogous to court-ordered rent abatement. This analogy would support the court trebling damages calculated by the *Miller* rent abatement formula when a landlord's violation of the Residential Rental Agreements Act rises to the level of an unfair trade practice. By allowing such an award, the court would penalize significantly those landlords who egregiously violate the Act.

The *Surratt* and *Allen* cases involved tenants who ceased paying rent and became defendants in summary eviction actions prior to vacating their rental units. The court held both times that the tenants' withholding of rent did not preclude their recovery of damages for the landlords' breaches of the implied warranty of habitability.¹⁴² Landlord *Allen* invoked the Act's prohibition against unilateral rent withholding prior to judicial action,¹⁴³ asserting that Simmons "waived her right to bring any action which arose out of her tenancy."¹⁴⁴ Professor Fillette has argued that such a broad interpretation of the prohibition would "largely negate" the Act's section 42-41, which states that the obligations of the tenant and the landlord are mutually dependent.¹⁴⁵

The court of appeals qualified its statement of the recovery rights of the tenant who has withheld rent by noting that a tenant cannot collect rent abatement damages for a period when she did not pay rent.¹⁴⁶ The court essentially defined rent abatement as a refund of rent; one obviously cannot get a refund of something one did not pay. This approach leaves unanswered, however, the question of the potential liability of the tenant who has withheld rent in response to his landlord's breach. Judge Greene's view that a tenant can assert the landlord's violation of the implied warranty of habitability to seek an abatement of his overdue rent obligation¹⁴⁷ is consistent with general law elsewhere.¹⁴⁸ It would be incongruous to hold that a landlord's attempt to collect rent for unfit

138. *Borders*, 68 N.C. App. at 770, 315 S.E.2d at 732.

139. *Allen*, 99 N.C. App. at 645, 394 S.E.2d at 484; see *supra* notes 59-64 and accompanying text.

140. See N.C. GEN. STAT. § 75-16 (1988) (providing for trebling of damages assessed for violation of unfair trade practice statutes).

141. *Borders*, 68 N.C. App. at 770, 315 S.E.2d at 731.

142. *Allen*, 99 N.C. App. at 642, 394 S.E.2d at 482; *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560.

143. See N.C. GEN. STAT. § 42-44(c) (1984).

144. Plaintiff Appellee's Brief at 7, *Allen* (No. 8921DC1155).

145. Fillette, *supra* note 65, at 789; see N.C. GEN. STAT. § 42-41 (1984).

146. *Allen*, 99 N.C. App. at 642, 394 S.E.2d at 482; *Surratt*, 99 N.C. App. at 407, 393 S.E.2d at 560.

147. *Surratt*, 99 N.C. App. at 411, 393 S.E.2d at 562 (Greene, J., concurring in result); see *supra* notes 41-42 and accompanying text.

148. See R. SCHOSHINSKI, *supra* note 2, § 3:22, at 136; UNIF. RESIDENTIAL LANDLORD AND TENANT ACT § 4.105, 7B U.L.A. 485 (1985).

housing could be an element of an unfair trade practice action,¹⁴⁹ and yet hold a tenant who has withheld rent liable for more than the fair rental value of substandard housing.

If the tenant withholding rent without judicial approval is not thereby limiting his claim to damages, except for the obvious lack of a claim for a refund of rent, the prohibition on unilateral rent withholding must be, by process of elimination, applicable to the tenant's right to possession. One commentator has suggested that the unilateral withholding prohibition complements the requirement in section 42-25.6 that prohibits landlords from evicting tenants without resort to a judicial summary ejectment proceeding.¹⁵⁰ Viewed from this perspective, the two provisions complement each other by denying either party possession rights if she attempts self-help remedies. North Carolina landlord and tenant law dealing with retaliatory eviction bars the retaliatory eviction defense to a summary ejectment claim when the tenant has failed to pay rent.¹⁵¹ This explicit linkage of the tenant's possession right and his duty to pay rent, in spite of possible wrongdoing by the landlord, lends further support to interpreting the unilateral withholding prohibition as creating a similar link.

Professor Fillette has argued for a narrower interpretation of the phrase "unilaterally withhold rent."¹⁵² He has suggested that, based on the mutuality of landlord and tenant obligations under section 42-41, a tenant's withholding of rent cannot be considered unilateral if the landlord previously breached the implied warranty of habitability.¹⁵³ Professor Fillette also has stated that "[t]he terms 'wrongful' and 'unilateral' could be read synonymously in this context."¹⁵⁴ This substitution of terms would essentially leave section 42-44(c) reading, "It is unlawful to withhold rent unlawfully." Thus, either by process of elimination or by following Professor Fillette's reasoning to its logical conclusion, section 42-44(c), if it has substantive content, denies rights of possession to the tenant who withholds rent before the landlord has been found by a court to have breached the implied warranty of habitability.

Although failing to clarify the unilateral rent withholding prohibition, the North Carolina Court of Appeals, in *Surratt* and *Allen*, has presented a formula for computing the contract damages of a tenant who has occupied unfit rental housing. The landlord is liable for the period of occupancy for which the court finds her to have breached the implied warranty of habitability under section 42-42(a).¹⁵⁵ Damages are the lesser of the rental unit's fair rental value as warranted minus the fair rental value of the rental unit in its defective condition, and the agreed rent. The court has derived a reasonable measure of damages in applying a lease-based, agreed-rent ceiling to a "benefit of the bargain" measure

149. *Allen*, 99 N.C. App. at 645, 394 S.E.2d at 484.

150. Note, *supra* note 114, at 1281; see N.C. GEN. STAT. § 42-25.6 (1984).

151. N.C. GEN. STAT. § 42-37.1(c)(1) (1984).

152. Fillette, *supra* note 65, at 790.

153. *Id.*

154. *Id.*

155. While the court also acknowledged the possibility of special and consequential damages, it did not discuss these types of damages.

of damages. The court was right to assert that withholding rent does not strip a tenant of the right to bring an action in breach of warranty against the landlord, but the court should have used the opportunity to clarify its interpretation of the statutory prohibition on unilateral withholding of rent.

The court of appeals has decided only four cases concerning contract damages for breach of the implied warranty of habitability in rental housing; none of these cases has reached the North Carolina Supreme Court. Future holdings should seek to reduce the remaining uncertainty concerning the legal obligations and potential liability of landlords and tenants. Holdings adhering to the three recommended principles that follow would add significantly to the certainty and equity of landlord and tenant law in North Carolina. First, the as-warranted fair rental value should be presumed equal to the agreed rent, if the tenant fails to prove that the premises were unfit when first occupied by the tenant. This approach would reduce both subjectivity and uncertainty of outcome in resolving valuation issues. Second, the courts should clarify the actions constituting unfair trade practices and the scope of damages to be trebled. Treble damages can be an appropriate and effective punitive remedy only if applied in a consistent manner to egregious circumstances. Third, the prohibition on unilateral rent withholding should have consequences only for the tenant's right to possession of the premises.

The North Carolina courts have addressed the issue of contract damages for which a landlord is liable when in violation of the implied warranty of habitability. If future holdings adhere to the three principles presented above, landlords and tenants will operate in a more certain and predictable legal environment. This element of stability would be of particular value in the fragile market for low income housing.

DANIEL B. HILL

A Clear Judicial Day in North Carolina—*Johnson v. Ruark Obstetrics* Smooths the Way for Plaintiffs' Claims for Negligent Infliction of Emotional Distress

"[T]here are clear judicial days on which a court can foresee forever,"¹ and in North Carolina, such a day has arrived. In *Johnson v. Ruark Obstetrics & Gynecology Associates*,² the North Carolina Supreme Court conclusively rejected the long-held and widely followed view that a plaintiff could not recover for negligent infliction of emotional distress absent the showing of some related physical injury, either as a cause of the distress or as a consequence of it.³ By way of replacement, the court decreed that "an allegation of ordinary negligence will suffice,"⁴ so long as the plaintiff also "allege[s] that severe emotional distress was the foreseeable and proximate result of such negligence."⁵ By rejecting the imposition of bright-line rules or standards,⁶ the *Johnson* decision is the judicial equivalent of a clean slate.

This Note chronicles the North Carolina courts' treatment of tort claims for negligent infliction of emotional distress and examines the varying limitations and restrictions imposed by the courts in their attempts to develop workable rules. It analyzes the *Johnson* court's selective application of precedent and questions both the court's disinclination to establish more specific standards regarding availability of the claim and the wisdom of dismissing several decades of contrary case law as "erroneous." The Note then explores the potential effects of *Johnson*. The Note concludes that the court should have set forth clearer standards to better guide the trial courts in determining whether to allow the

1. *Johnson v. Ruark Obstetrics & Gynecology Assocs.*, 327 N.C. 283, 309, 395 S.E.2d 85, 100 (Meyer, J., dissenting) (quoting *Thing v. La Chusa*, 48 Cal. 3d 644, 668, 771 P.2d 814, 830, 257 Cal. Rptr. 865, 881 (1989) ("with apologies to [lyricist] Bernard Witkin")).

2. 327 N.C. 283, 395 S.E.2d 85 (1990).

3. The court found it necessary to overrule a number of cases that "require[d] a plaintiff to show—in addition to mental or emotional injury—a physical impact, physical injury, or a physical manifestation of emotional distress to succeed on a claim of negligent infliction of emotional distress." *Id.* at 301, 395 S.E.2d at 95. The court overruled *Hinnant v. Tide Water Power Co.*, 189 N.C. 120, 126 S.E. 307 (1925), *overruled on other grounds* in *Nicholson v. Hugh Chatham Memorial Hosp., Inc.*, 300 N.C. 295, 266 S.E.2d 818 (1980). The court also disapproved *Stanback v. Stanback*, 297 N.C. 181, 254 S.E.2d 611 (1979), *disapproved* in *Dickens v. Puryear*, 302 N.C. 437, 276 S.E.2d 325 (1981); *Williamson v. Bennett*, 251 N.C. 498, 112 S.E.2d 48 (1960); and *Kimberly v. Howland*, 143 N.C. 398, 55 S.E. 778 (1906), to the extent those cases implicated a physical injury requirement.

In addition, *Johnson* overruled seven decisions of the North Carolina Court of Appeals. The overruled decisions were *Edwards v. Advo Systems, Inc.*, 93 N.C. App. 154, 376 S.E.2d 765 (1989); *Ledford v. Martin*, 87 N.C. App. 88, 359 S.E.2d 505 (1987), *disc. rev. denied*, 321 N.C. 473, 365 S.E.2d 1 (1988); *Campbell ex rel. McMillan v. Pitt County Memorial Hosp., Inc.*, 84 N.C. App. 314, 352 S.E.2d 904, *aff'd on other grounds*, 321 N.C. 260, 362 S.E.2d 273 (1987); *Woodell v. Pinehurst Surgical Clinic, P.A.*, 78 N.C. App. 230, 336 S.E.2d 716 (1985), *aff'd*, 316 N.C. 550, 342 S.E.2d 523 (1986) (*per curiam*); *Craven v. Chambers*, 56 N.C. App. 151, 287 S.E.2d 905 (1982); and *Wesley v. Greyhound Lines, Inc.*, 47 N.C. App. 680, 268 S.E.2d 855, *disc. rev. denied*, 301 N.C. 239, 283 S.E.2d 136 (1980).

4. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

5. *Id.*

6. *Id.* at 291, 395 S.E.2d at 89. The court explained that "our law includes no arbitrary requirements to be applied mechanically to claims for the negligent infliction of emotional distress." *Id.*

claim to go to a jury, and recommends different standards of recovery for directly injured victims and for bystanders. The Note proposes alternative limitations for each; specifically, the Note argues that in order to state a claim for negligent infliction of emotional distress, directly injured plaintiffs should be required to forecast evidence that, if believed, would support an award of punitive damages.⁷ This standard emphasizes the nature of the defendant's conduct rather than the extent of the plaintiff's emotional injury. The Note agrees with the *Johnson* court's affirmation of bystanders' ability to recover for negligently inflicted emotional distress, but suggests that they should be permitted to do so only in accordance with a modified version of the test recently established by the California Supreme Court in *Thing v. La Chusa*.⁸

From March through October of 1983, expectant parent Barbara Johnson received prenatal medical care from Ruark Obstetrics and Gynecology Associates, P.A.⁹ Mrs. Johnson's pregnancy progressed normally through the morning of October 3, 1983;¹⁰ later that afternoon, she began experiencing contractions and was admitted to Wake Medical Center.¹¹ Testing undertaken at the hospital revealed the absence of any fetal heart tones, and the Johnsons were informed of the fetal death at approximately 8:00 p.m.¹² Mrs. Johnson remained in labor and delivered the stillborn fetus in the early hours of October 4, 1983.¹³

On behalf of the fetus, Glenn Johnson filed a wrongful death action against defendants.¹⁴ Included in the complaint were claims for negligent infliction of emotional distress, brought by both Glenn and Barbara Johnson "in their individual capacities as father and mother of the fetus."¹⁵ The complaint alleged, in sum, that Mrs. Johnson's doctors negligently failed to treat Mrs. Johnson's diabetic condition properly, thereby causing the fetus to die of malnutrition.¹⁶

Defendants denied any negligence and moved to dismiss for failure to state

7. See *infra* notes 167-71 and accompanying text. A "directly injured plaintiff" is any party who brings suit to recover for mental injury negligently inflicted upon her by the defendant.

8. 48 Cal. 3d 644, 771 P.2d 814, 257 Cal. Rptr. 865 (1989). See *infra* notes 173-80 and accompanying text. Discussions of "bystander" recovery refer to recovery by a plaintiff who alleges mental suffering brought about by the plaintiff's concern for a third party who was directly injured by defendant's negligent conduct.

9. *Johnson*, 327 N.C. at 286, 395 S.E.2d at 87. Glenn and Barbara Johnson brought suit against Ruark Obstetrics and Gynecology Associates, P.A. (formerly The Ruark Clinic), and against the four individual doctors employed by the defendant who managed Mrs. Johnson's prenatal health care program.

10. *Id.* at 286-87, 395 S.E.2d at 87. Mrs. Johnson reported feeling fetal movement during the evening of October 2, 1983. *Id.* According to the complaint, defendant Dr. Edgerton reported the presence of a fetal heart tone at 9:30 a.m. on October 3, 1983, when Mrs. Johnson visited The Ruark Clinic to report contractions. Plaintiffs' Brief at 5, *Johnson* (No. 8610SC942).

11. *Johnson*, 327 N.C. at 286-87, 395 S.E.2d at 87.

12. *Id.* at 287, 395 S.E.2d at 87.

13. *Id.*

14. *Id.* Mr. Johnson was authorized by statute to bring the action as administrator of the fetal estate. N.C. GEN. STAT. § 28A-18-2 (1984).

15. *Johnson*, 327 N.C. at 287, 395 S.E.2d at 87.

16. *Johnson v. Ruark Obstetrics & Gynecology Assocs.*, 89 N.C. App. 154, 156, 365 S.E.2d 909, 910 (1988), *aff'd on other grounds*, 327 N.C. 283, 395 S.E.2d 85 (1990). The infant's death

a claim upon which relief could be granted.¹⁷ Defendants also requested, and received, summary judgment as to all claims.¹⁸ The court of appeals reversed the trial court's dismissal of both the wrongful death and the negligent infliction of emotional distress claims.¹⁹ In so doing, the court rejected defendants' contention that North Carolina prohibits recovery for mental anguish prompted by concern for another person.²⁰ The court relied on *Nicholson v. Hugh Chatham Memorial Hospital, Inc.*²¹ to hold that "where some intimate relationships are affected, there is no longer any *absolute* prohibition against compensating emotional distress damages arising from injuries to others."²² The court recognized, however, that availability of the claim must be tempered by the policy interest in limiting negligence liability.²³

Having determined that the claim was not barred as a matter of law, the court of appeals considered defendants' argument that plaintiffs still had no standing to sue, not having suffered any physical injuries themselves.²⁴ Citing *Williamson v. Bennett*,²⁵ the court found that "absent some [physical] impact, the emotional distress claimant must manifest some resulting physical injury."²⁶ The court held that both Glenn and Barbara Johnson had alleged proper claims because Barbara Johnson did in fact suffer physical injury and because the court could not say as a matter of law that Glenn Johnson had not suffered physical manifestations of his mental distress.²⁷

certificate listed the cause of death as "placental insufficiency." Plaintiffs' Record on Appeal to the North Carolina Court of Appeals, Exhibit B at 10, *Johnson* (No. 8610SC942).

In relevant part, the complaint alleged:

Past, present and future pain and suffering and emotional distress of enduring the labor, with the knowledge that their unborn child was dead, and the delivery of a dead child.

Past, present and future mental distress and anguish resulting from the dramatic circumstances surrounding the stillbirth of their child.

Plaintiffs' Complaint at 7, *Johnson* (No. 8610SC942).

17. *Johnson*, 327 N.C. at 287, 395 S.E.2d at 87; see N.C.R. Civ. PRO. 12(b)(6).

18. *Johnson*, 89 N.C. App. at 156-57, 365 S.E.2d at 911; see N.C.R. Civ. PRO. 56(e).

19. *Johnson*, 89 N.C. App. at 171, 365 S.E.2d at 919.

20. *Id.* at 163, 365 S.E.2d at 915. The defendants relied on *Williamson v. Bennett*, 251 N.C. 498, 508, 112 S.E.2d 48, 55 (1960), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85, and *Hinnant v. Tide Water Power Co.*, 189 N.C. 120, 126 S.E.2d 307 (1925), *overruled on other grounds in Nicholson v. Hugh Chatham Memorial Hosp., Inc.*, 300 N.C. 295, 266 S.E.2d 818 (1980), and *overruled in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

21. 300 N.C. 295, 266 S.E.2d 818 (1980).

22. *Johnson*, 89 N.C. App. at 163, 365 S.E.2d at 915.

23. *Id.* at 164, 365 S.E.2d at 915. The court explained that the policy issues involved generally relate to questions of remoteness and proximate cause. *Id.* According to the court, "the definition of proximate cause necessarily includes . . . whether the tort-feasor's liability should as a matter of public policy extend to [the] injuries." *Id.*

24. *Id.* at 162, 365 S.E.2d at 914.

25. 251 N.C. 498, 112 S.E.2d 48 (1960), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85. In *Williamson* the North Carolina Supreme Court held that "[i]t is almost the universal opinion that recovery may be had for mental or emotional disturbance in ordinary negligence cases where, coincident in time and place with the occurrence producing the mental stress, some actual physical impact or genuine physical injury also resulted directly from defendant's negligence." *Id.* at 503, 112 S.E.2d at 52.

26. *Johnson*, 89 N.C. App. at 165, 365 S.E.2d at 916.

27. *Id.* Barbara Johnson alleged two physical injuries, the first resulting from the negligent treatment of her diabetes and the second resulting from the injury to the fetus; the court observed that since "the fetus is normally attached to the mother's uterine wall, we fail to see how a physical

The North Carolina Supreme Court affirmed the court of appeals' reinstatement of the Johnsons' claims, albeit on very different grounds.²⁸ Writing for the majority, Justice Mitchell described the physical injury requirement as a miscategorization of North Carolina law and acknowledged that "varying and at times inconsistent analyses used by our courts have apparently buttressed such misconceptions."²⁹ Overruling or disapproving a number of North Carolina cases that had included such a requirement,³⁰ the court held that "neither a physical impact, a physical injury, nor a subsequent physical manifestation of emotional distress is an element of the tort of negligent infliction of emotional distress."³¹

In its review of the tort's "long and winding history in every state,"³² including North Carolina, the court considered the theories at work in other jurisdictions. The court resolved, however, to base any final conclusions exclusively on North Carolina law.³³ To state a valid claim for negligent infliction of emotional distress, the court concluded that "a plaintiff must allege that (1) the defendant negligently engaged in conduct, (2) it was reasonably foreseeable³⁴ that such conduct would cause the plaintiff severe emotional distress (often referred to as mental anguish), and (3) the conduct did in fact cause the plaintiff severe emotional distress."³⁵ Thus, the North Carolina courts' treatment of such traditional concepts as foreseeability and proximate cause drew much of the *Johnson* court's attention. The application of these concepts, the court said, "must be determined under all the facts presented, and should be resolved on a case-by-case basis by the trial court and, where appropriate, by a jury."³⁶

The court drew a distinction between mere "fright," which by itself is not compensable,³⁷ and "severe emotional distress," for which a plaintiff might re-

impact or injury to the fetus would not normally be an injury or impact to the mother." *Id.* at 167, 365 S.E.2d at 917.

Regarding Glenn Johnson's claim, the court concluded that the pleadings revealed no facts that would necessarily "prohibit him from later more specifically forecasting or introducing evidence that his alleged mental distress resulted in the necessary physical injury." *Id.* at 168-69, 365 S.E.2d at 918.

28. *Johnson*, 327 N.C. at 286, 395 S.E.2d at 86-87.

29. *Id.* at 291, 395 S.E.2d at 89.

30. See cases cited *supra* note 3.

31. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

32. *Id.* at 288, 395 S.E.2d at 88.

33. The dissent characterized as error the majority's reluctance to "[seek] guidance from the experience of other jurisdictions with less expansive doctrines of recovery." *Id.* at 308, 395 S.E.2d at 99 (Meyer, J., dissenting). For a survey of several other jurisdictions' attempts to define the required elements of the tort, see cases cited *infra* note 59.

34. The differing interpretations of "foreseeability" and its usefulness as a restrictive aspect of the claim are discussed thoroughly by the court and debated vigorously in Justice Meyer's dissent. For a summary of Justice Meyer's concerns, see *infra* notes 43-62 and accompanying text. For a collection of foreseeability standards applied by other jurisdictions, see *infra* note 59.

35. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

36. *Id.* at 305, 395 S.E.2d at 98 (citing *Kanoy v. Hinshaw*, 273 N.C. 418, 160 S.E.2d 296 (1968); *Toone v. Adams*, 262 N.C. 403, 137 S.E.2d 132 (1964)).

37. *Id.* at 303-04, 395 S.E.2d at 97. This restriction is consistent with the court's prior holdings, reviewed in detail by the majority. *Id.* at 294, 395 S.E.2d at 91-92. For example, in *Kimberly v. Howland*, 143 N.C. 399, 403, 55 S.E. 778, 780 (1906), the court observed that "[a]ll the courts agree that mere fright, unaccompanied by or followed by physical injury, cannot be considered as an element of damage." Although the *Kimberly* statement was dictum because the plaintiff had suffered contemporaneous physical injury, later cases often cited it. See, e.g., *Williamson v. Bennett*,

cover.³⁸ The court recognized that severe emotional distress could encompass a wide range of disorders so long as the condition alleged by the plaintiff was of a sort generally recognized as a genuine disorder by trained medical personnel.³⁹

The *Johnson* court rejected any requirement of heightened negligence.⁴⁰ The court also established identical standards for the plaintiff who brings an action based upon a mental or emotional injury inflicted directly upon him by the defendant, and for the "bystander" plaintiff who alleges mental suffering brought about by the plaintiff's concern for another party injured by the defendant's negligent conduct.⁴¹ Applying these standards, the court concluded that the Johnsons' allegations of emotional distress were sufficient to support their cause of action.⁴²

The *Johnson* decision was accompanied by a vigorous dissent,⁴³ which characterized the Johnsons' alleged injuries as the understandable distress occasioned by the loss of a child rather than as anguish caused by any specific acts of negligence on the part of the defendants.⁴⁴ Condemning the rule adopted by the

251 N.C. 498, 503-04, 112 S.E.2d 48, 52 (1960) ("mere fright caused by ordinary negligence does not give a cause of action and may not be considered an element of damages"), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; Kirby v. Jules Chain Stores Corp., 210 N.C. 808, 812, 188 S.E. 625, 627 (1936) ("fright alone is not actionable," but allegations of ensuing "nervous disorders of serious proportions" will preclude dismissal or nonsuit); Arthur v. Henry, 157 N.C. 438, 439, 73 S.E. 211, 212 (1911) ("mere fright is not actionable"); see Byrd, *Recovery for Mental Anguish in North Carolina*, 58 N.C.L. REV. 435, 437-45 (1980).

For a discussion of the North Carolina courts' treatment of fright as an element of a claim for the negligent infliction of emotional distress, see *infra* notes 81-85 & 87 and accompanying texts.

38. *Johnson*, 327 N.C. at 303-04, 395 S.E.2d at 97.

39. *Id.* at 304, 395 S.E.2d at 97. Specifically, the court spoke to "any emotional or mental disorder, such as, for example, neurosis, psychosis, chronic depression, phobia, or any other type of severe and disabling emotional or mental condition which may be generally recognized and diagnosed by professionals trained to do so." *Id.*

40. *Id.*

41. *Id.* Recovery for emotional injury to directly injured victims is qualitatively different from recovery for emotional injury to bystanders. Standards of recovery should reflect that difference. See *infra* notes 166 & 172-80 and accompanying texts.

42. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97. In the instant case, both Glenn and Barbara Johnson were within the generally recognized definition of "bystanders" insofar as their claim was related to the injury to their child (a third party). Alternatively, Mrs. Johnson's claim could have been based upon mental and emotional injuries that she herself endured while undergoing labor. See *Johnson v. Ruark Obstetrics & Gynecology Assocs.*, 89 N.C. App. 154, 166-67, 365 S.E.2d 909, 916-17 (1988), *aff'd on other grounds*, 327 N.C. 283, 395 S.E.2d 85 (1990).

43. Justice Meyer authored a lengthy dissent that included a survey of standards applied by other jurisdictions and a critical examination of the court's definition of foreseeability. See *infra* notes 44-62 and accompanying text.

Justice Webb dissented separately on grounds that the case represented a marked departure from precedent, as evidenced by the number of cases overruled, and that the cases overruled were in fact accurate interpretations of valid law. *Johnson*, 327 N.C. at 318, 395 S.E.2d at 106 (Webb, J., dissenting). Despite the admittedly "arbitrary" nature of the physical injury requirement, Justice Webb argued that the rule served valid policy purposes by providing a needed limitation on liability. *Id.* (Webb, J., dissenting). Justice Webb would hold that the plaintiffs did not state claims for negligent infliction of emotional distress. *Id.* (Webb, J., dissenting).

44. *Id.* at 307, 395 S.E.2d at 99 (Meyer, J., dissenting). For an interesting early discussion of the need to distinguish between these two sources of injury, see *Young v. Western Union Tel. Co.*, 107 N.C. 370, 382, 11 S.E. 1044, 1048 (1890). See also *Hancock v. Western Union Tel. Co.*, 137 N.C. 498, 501, 49 S.E. 952, 953 (1905) (expressing concern that jurors "may possibly confound the mental anguish naturally arising from the loss of a near relative with that which grows from the defendant's negligence"); *Cashion v. Western Union Tel. Co.*, 123 N.C. 267, 274, 31 S.E. 493, 494 (1898) (same).

court as "overbroad,"⁴⁵ Justice Meyer criticized the majority for "go[ing] beyond even *Dillon's*⁴⁶ broad approach"⁴⁷ without adequate justification and for failing to provide practical standards or workable restrictions to the lower courts.⁴⁸ Specifically, Justice Meyer expressed strong reservations concerning the court's treatment of the concepts of duty,⁴⁹ foreseeability,⁵⁰ and proximate cause.⁵¹

The dissent interpreted the majority's opinion as holding that "a defendant has a duty not to cause serious emotional distress in any person who might foreseeably suffer such distress from proximate negligence. This duty is limited only by the foreseeability that such harm may occur."⁵² Justice Meyer argued that this theory presumes the existence of a duty without first providing any analysis of the duty's foundation.⁵³ In particular, Justice Meyer questioned whether defendants' alleged negligent failure to treat, thereby causing the death of the fetus, created any duty not to cause serious emotional distress flowing from the defendants to the Johnsons.⁵⁴

Further, Justice Meyer observed that without more specific restrictions on the standards of recovery the *Johnson* decision "stands for the proposition that no risk of serious emotional distress is acceptable."⁵⁵ This premise, he contended, would be contrary to such well-established axioms of tort law as Judge Learned Hand's famous cost-benefit equation⁵⁶ and would have detrimental effects on the availability and price of social necessities, including medical care

45. *Johnson*, 327 N.C. at 313, 395 S.E.2d at 102 (Meyer, J., dissenting).

46. *Dillon v. Legg*, 68 Cal. 2d 728, 441 P.2d 912, 69 Cal. Rptr. 72 (1968) (en banc). For a description of the *Dillon* test, see *infra* notes 144-45 and accompanying text.

47. *Johnson*, 327 N.C. at 308, 395 S.E.2d at 99 (Meyer, J., dissenting).

48. *Id.* at 312-13, 395 S.E.2d at 102 (Meyer, J., dissenting). Justice Meyer criticized the court for failing to establish "any limitations whatsoever on this duty not to negligently inflict foreseeable serious emotional distress," because "[i]n adopting a rule, it should not be so vague that it provides no guidance to the judges and juries that must implement it." *Id.* (Meyer, J., dissenting).

49. See *infra* notes 52-54 and accompanying text.

50. See *infra* notes 58-62 and accompanying text.

51. *Johnson*, 327 N.C. at 309-14, 395 S.E.2d at 100-03 (Meyer, J., dissenting). For Justice Meyer's discussion of the majority's treatment of proximate cause, see *infra* notes 55-62 and accompanying text.

52. *Johnson*, 327 N.C. at 309, 395 S.E.2d at 100 (Meyer, J., dissenting).

53. *Id.* (Meyer, J., dissenting).

54. *Id.* (Meyer, J., dissenting).

55. *Id.* at 312, 395 S.E.2d at 102 (Meyer, J., dissenting).

56. *Id.* (Meyer, J., dissenting). Judge Hand suggested an equation that evaluates the sensibility of taking certain risks in *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947).

[Judge] Hand described the duty of an actor to protect against the resulting injuries as being a function of three variables: (1) the probability (*P*) of injury occurring, (2) the gravity (*L*) of resulting injury, and (3) the burden (*B*) of adequate precautions. Hand described this relationship algebraically as an inquiry as to whether $B < PL$.

Johnson, 327 N.C. at 312 n.3, 395 S.E.2d at 102 n.3 (Meyer, J., dissenting) (construing *Carroll Towing*, 159 F.2d at 173).

Justice Meyer observed that "[v]irtually all conduct is risk creating" and argued that the *Johnson* court failed to differentiate between acceptable and unacceptable risks. *Id.* at 312, 395 S.E.2d at 102 (Meyer, J., dissenting). He concluded: "Today's decision, drawing no such distinction, stands for the proposition that no risk of serious emotional distress is acceptable." *Id.* (Meyer, J., dissenting).

and insurance.⁵⁷

Stating that "the universe of plaintiffs contemplated by the majority's rule is infinite indeed,"⁵⁸ the dissent also called for more circumscribed definitions of foreseeability and proximate cause based upon "the relationship of the plaintiff, the proximity of perception, and the severity of the injury that would give rise to a bystander's cause of action for serious emotional distress."⁵⁹ These standards resemble the factors enumerated in *Dillon v. Legg*, an oft-cited California case which established the threshold requirements for recovery for negligently inflicted emotional distress later imposed by several jurisdictions.⁶⁰ Justice Meyer

57. Justice Meyer said:

[T]he impact of this rule on the availability of medical care, particularly that of obstetrics, will be to further discourage qualified physicians from practicing. The risk of liability and the escalated premium for insurance to cover the liability are already seriously affecting the delivery of obstetrical care to this state, particularly to the rural areas and to the poor. . . . I cannot think that our state will benefit from a rule that discourages such risk-taking activity without regard to the costs society might pay or the benefits society might derive therefrom.

Johnson, 327 N.C. at 312, 395 S.E.2d at 102 (Meyer, J., dissenting).

58. *Id.* at 311, 395 S.E.2d at 101 (Meyer, J., dissenting). Justice Meyer concluded that "[l]iability without limitation adversely affects three distinct groups: tort-feasors, the physically injured primary and secondary victims, and society as a whole." *Id.* at 311, 395 S.E.2d at 101-02 (Meyer, J., dissenting).

59. *Id.* at 313, 395 S.E.2d at 102-03 (Meyer, J., dissenting). For the requirement that the bystander plaintiff have a specific relationship to the victim, see, e.g., *Thing v. La Chusa*, 48 Cal. 3d 644, 667-68, 771 P.2d 814, 829-30, 257 Cal. Rptr. 865, 880 (1989) (mother of victim is "closely related"); *Elden v. Sheldon*, 46 Cal. 3d 267, 273, 758 P.2d 582, 586-87, 250 Cal. Rptr. 254, 258 (1988) (unmarried cohabitant denied recovery); *James v. Leib*, 221 Neb. 47, 55, 375 N.W.2d 109, 117 (1985) (child who witnessed death of sibling could recover; court placed greatest weight on relationship between victim and bystander and required marital or close familial relation); *Gates v. Richardson*, 719 P.2d 193, 199 (Wyo. 1986) (limiting recovery to spouses, children, parents, and siblings for social policy purposes).

For limitations based on the proximity of perception or the plaintiff's location in the "zone of danger," see, e.g., *La Chusa*, 48 Cal. 3d at 669, 771 P.2d at 830, 257 Cal. Rptr. at 881 (denied recovery to victim's mother, who was neither present at scene nor aware son was being injured); *Kelley v. Kokua Sales & Supply, Ltd.*, 56 Haw. 204, 209, 532 P.2d 673, 676 (1975) (physical proximity to scene of tort is determinative of liability); *Rickey v. Chicago Transit Auth.*, 98 Ill. 2d 546, 555, 457 N.E.2d 1, 5 (1983) (child who witnessed serious injury to brother on escalator required to prove that he too was in zone of danger); *Stadler v. Cross*, 295 N.W.2d 552, 553, 555 (Minn. 1980) (parents witnessed severe injury to their child but denied recovery because not within zone of danger); *Wilder v. City of Keene*, 131 N.H. 599, 604, 557 A.2d 636, 639 (1989) (no recovery for plaintiff who did not show geographic and temporal proximity to accident). *But cf.* *Dillon v. Legg*, 68 Cal. 2d 728, 733, 441 P.2d 912, 915, 69 Cal. Rptr. 72, 75 (1968) (en banc) (relief should not be conditioned on plaintiff's location within a matter of yards of accident; case "exposes the hopeless artificiality of the zone-of-danger rule").

For the requirement that the bystander plaintiff witness a severe injury to the victim, see, e.g., *Portee v. Jaffee*, 84 N.J. 88, 100-01, 417 A.2d 521, 527-28 (1980) (severe mental distress not a usual result when bystander perceives only less serious harm); *Gates v. Richardson*, 719 P.2d 193, 199 (Wyo. 1986) (injury to victim must be objectively severe; court reasoned that "people recover from serious shock quickly if it turns out to be a false alarm").

For the requirement that the bystander plaintiff himself suffer severe emotional distress as a result of the injury to the victim, see, e.g., *La Chusa*, 48 Cal. 3d at 668, 771 P.2d at 829-30, 257 Cal. Rptr. at 880-81 (requiring "serious emotional distress"); *Lejeune v. Rayne Branch Hosp.*, 556 So. 2d 559, 570 (La. 1990) (mental distress must be severe and debilitating). The *Johnson* court required the plaintiff's mental distress to be "severe." *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

60. The *Johnson* factors are comparable to the factors enumerated in *Dillon*, discussed *infra* in text accompanying note 145. *Dillon* itself was restricted recently by the California Supreme Court in

dismissed the *Johnson* court's requirements as "low hurdles"⁶¹ and called for "limits on the class of bystander plaintiff[s]."⁶²

The *Johnson* court devoted a great deal of its attention to an extensive review of North Carolina's treatment of claims alleging negligent infliction of emotional distress. The effects of conflicting policy goals and varying fact patterns turned what was originally a fairly clear doctrine into a confusing hybrid characterized, the court observed, by "erroneous" and "unfortunate" misstatements of law.⁶³ Because a sequential presentation of the cases would serve only as a literal illustration of this confusion, this Note orders the discussion of the cases according to their thematic rather than chronological development.⁶⁴

The North Carolina Supreme Court first confronted the question whether a plaintiff could recover for the negligent infliction of emotional distress at all in *Young v. Western Union Telegraph Company*.⁶⁵ This case and others like it⁶⁶

Thing v. La Chusa, 48 Cal. 3d 644, 771 P.2d 814, 257 Cal. Rptr. 865 (1989). The *La Chusa* court concluded that *Dillon* was unworkable; see *infra* notes 146-49 and accompanying text.

For scholarly measurements of the success of the *Dillon* test, see generally Diamond, *Dillon v. Legg Revisited: Towards a Unified Theory of Compensating Bystanders and Relatives for Intangible Injuries*, 35 HASTINGS L.J. 477, 504 (1984) (criticizing *Dillon* for "using foreseeability to limit compensation in the first instance"); see also Pearson, *Liability to Bystanders for Negligently Inflicted Emotional Harm—A Comment on the Nature of Arbitrary Rules*, 34 U. FLA. L. REV. 477, 490-501 (1982) (characterizing *Dillon* as at least as arbitrary and unsatisfactory as the zone-of-danger rule, if not more so); Comment, *Negligent Infliction of Mental Distress: A Jurisdictional Survey of Existing Limitation Devices and Proposal Based on an Analysis of Objective Versus Subjective Indices of Distress*, 33 VILL. L. REV. 781, 803-17 (1988) (detailed discussion of many jurisdictions' treatments of *Dillon*); Note, *Bystander Recovery: A Policy Oriented Approach*, 32 N.Y.L. SCH. L. REV. 877, 888-98 (1987) (discussing difficulties encountered by California courts interpreting *Dillon*).

61. *Johnson*, 327 N.C. at 310, 395 S.E.2d at 101 (Meyer, J., dissenting). Justice Meyer said the court's implication that psychiatric testimony might be required to substantiate or verify that the plaintiff suffered severe mental distress was a "totally ineffective barrier" because diagnosable responses to traumatic events are common, and possibly even a statistical likelihood. *Id.* (Meyer, J., dissenting).

For a listing of the various elements often cited in support of medical diagnoses of severe mental distress, see DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 236-38 (3d rev. ed. 1987) [hereinafter DIAGNOSTIC AND STATISTICAL MANUAL]. See *infra* note 119 for a short listing of elements.

62. *Johnson*, 327 N.C. at 311, 395 S.E.2d at 101 (Meyer, J., dissenting).

63. *Id.* at 294-95, 395 S.E.2d at 91-92. (The "Kimberly opinion was the first opinion of this Court to characterize, unfortunately, emotional injury as a type of physical injury . . . Such mischaracterizations of emotional distress [are now recognized to be] unnecessary and erroneous terminology.").

64. First, the Note discusses the North Carolina Supreme Court's initial recognition of negligent infliction of emotional distress claims when the claims were coupled with other causes of action. See *infra* notes 65-79. The Note then addresses the development of the physical injury requirement and attempts to clarify the often overlooked distinctions between mere fright and genuine mental anguish, and the differing standards applicable to each. See *infra* notes 80-92. The Note also analyzes the cases pertaining to the collateral issue of bystander recovery. See *infra* notes 93-108.

65. 107 N.C. 370, 11 S.E. 1044 (1890). The *Young* court noted that the issue was one of "first impression in this State." *Id.* at 383, 11 S.E. at 1048.

The *Young* decision was also the first of many "negligent delivery of a telegram" cases in which the court routinely allowed recovery for mental distress occasioned by a late or missing message. In this context, the court stated in *Hancock v. Western Union Tel. Co.*, 137 N.C. 498, 500-01, 49 S.E. 952, 953 (1905), that "[t]he right to recover damages for purely mental anguish, not connected with or growing out of a physical injury, is the settled law of this State, and it is too late now to question it."

66. The cases that first established the tort of negligent infliction of emotional distress in North Carolina generally date to the period between 1800 and 1920. For a comprehensive survey of the

focused not only on the injuries for which a plaintiff could recover, but also on the nature of the cause of action itself. Plaintiffs generally brought claims for negligent infliction of emotional distress in conjunction with other claims,⁶⁷ in tort or in contract. In *Young*, for example, plaintiff alleged that he had "suffered great pain, mental anguish, and distress by reason of the [defendant's] gross negligence and delay in transmitting . . . [a] telegram"⁶⁸ that, if delivered, would have given plaintiff notice of his wife's impending death and afforded him the opportunity to be with her and to attend her funeral.⁶⁹ Confirming that these mental injuries were indeed compensable, the court cited Cicero's Eleventh Philippic against Anthony as instructive: "For, as the power of the mind is greater than that of the body, in the same way the sufferings of the mind are more severe than the pains of the body."⁷⁰

The *Young* court also emphasized the importance of defendant's status as a public servant and of the duties it owed to the public as a result of that relationship.⁷¹ Though plaintiff's cause of action was based in contract, the court also recognized that the claim was "in reality in the nature of tort for the negligence, and that . . . the plaintiff is entitled to recover . . . for the actual damages done him, and that mental anguish is actual damage."⁷²

In other early cases alleging mental anguish, the court also routinely recognized a right to recover for mental distress occasioned by negligent transportation or mishandling of dead bodies where the plaintiff was a close relative of the deceased person.⁷³ In allowing these plaintiffs to recover, the court frequently

tort of negligent infliction of emotional distress and its development in North Carolina, see Byrd, *supra* note 37.

67. For example, plaintiffs sought recovery for negligent infliction of emotional distress in connection with missing or late telegraphs, negligent actions by common carriers, negligent mishandling of dead relatives' bodies, and the negligent termination of telephone services. Byrd, *supra* note 37, at 452-55 & n.17.

68. *Young*, 107 N.C. at 371, 11 S.E. at 1044.

69. *Id.*

70. *Id.* at 385, 11 S.E. at 1048-49. Cicero, the court observed, "certainly may be quoted as an authority among lawyers." *Id.* at 385, 11 S.E. at 1048.

The severity of mental injuries and their genuine ability to incapacitate has been recognized consistently by the North Carolina courts. See, e.g., *Hargis v. Knoxville Power Co.*, 175 N.C. 31, 34, 94 S.E. 702, 703 (1917) ("As all pain is mental and centers in the brain, it follows that as an element of damage for personal injury the injured party is allowed to recover for actual suffering of mind and body"); *Kimberly v. Howland*, 143 N.C. 398, 404, 55 S.E. 778, 780 (1906) ("[T]he general principles of the law of torts support a right of action for physical injuries resulting from negligence . . . none the less strongly because the physical injury consists of a wrecked nervous system instead of lacerated limbs."), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *Hunter v. Western Union Tel. Co.*, 135 N.C. 458, 467, 47 S.E. 745, 748 (1904) (Clark, C.J., concurring) ("[m]ental suffering is as real as physical").

71. *Young*, 107 N.C. at 383, 11 S.E. at 1048. The court explained that

[i]n failing to promptly deliver [a] telegram the telegraph company negligently fails to perform a duty which it owes to the sender of [the] telegram, and should be held liable for whatever injury follows as the proximate result of its negligent conduct. It is not a mere breach of contract, but a failure to perform a duty which rests upon it as a servant of the public.

Id. at 377, 11 S.E. at 1046.

72. *Id.* at 385, 11 S.E. at 1048.

73. *Parker v. Quinn-McGowen Co.*, 262 N.C. 560, 138 S.E.2d 214 (1964) (denying recovery but recognizing a quasi-property right in the body); *Stephenson v. Duke Univ.*, 202 N.C. 624, 163 S.E.

relied on the "rationale that the cause of action [was] based on a quasi-property right in the body."⁷⁴

In 1916, the court held in *Bailey v. Long*⁷⁵ that a plaintiff-husband could recover not only the medical expenses he incurred as a result of the defendant-doctor's negligent breach of a contract to provide skillful medical care for plaintiff's wife, but also could recover damages for the mental anguish he sustained when she died as a result of the doctor's negligence.⁷⁶ *Bailey* suggested that the underlying contract claim might not be crucial to plaintiff's cause of action for negligent infliction of emotional distress; in short, it suggested that the tort claim could stand alone. The court reasoned that "if the husband can recover damages from a telegraph company for mental anguish for delay in delivering a telegram informing him of his wife's illness, he should . . . recover for the mental anguish occasioned by witnessing her suffering and death against the alleged author of such suffering and death."⁷⁷

A number of decisions that emphasized the related or underlying claims present in the successful early negligent infliction of emotional distress cases, however, quickly curtailed *Bailey*'s suggestion that a negligent infliction of emotional distress claim might be valid even without an underlying claim based in contract or on a breach of public duty. After holding that the basis for liability in *Bailey* was a contractual duty owed to plaintiff by defendant,⁷⁸ the North Carolina Supreme Court began to scale back the availability of the negligent infliction of emotional distress claim by returning to its earlier emphasis on the presence of separate underlying tort or contract claims.⁷⁹

The physical injury requirement itself was first identified as a valid underlying cause of action in *Kimberly v. Howland*.⁸⁰ In *Kimberly*, plaintiff brought an action for negligently inflicted emotional distress to recover for the fright she sustained when defendants negligently conducted blasting operations near her home, causing a large rock to come crashing through the roof.⁸¹ The North Carolina Supreme Court observed: "All the courts agree that mere fright, unac-

698 (1932) (recovery limited to next-of-kin); *Floyd v. Atlantic Coast Line Ry. Co.*, 167 N.C. 55, 83 S.E. 12 (1914) (same); see Byrd, *supra* note 37, at 454-55 & nn.120-27.

74. Byrd, *supra* note 37, at 455.

75. 172 N.C. 661, 90 S.E. 809 (1916).

76. *Id.* at 661-62, 90 S.E. at 809. For further discussion of the *Bailey* facts and its significance in the context of bystander recovery, see *infra* notes 94-98 and accompanying text.

77. *Bailey*, 172 N.C. at 663, 90 S.E. at 810. For a discussion of the many cases allowing recovery for negligent infliction of emotional distress caused by a late or missing telegraph, see Byrd, *supra* note 37, at 452-54.

78. "The foundation of liability in *Bailey's Case* was the contractual relation existing between the plaintiff and the defendant; the action was not prosecuted by a stranger or a third party." *Hinnant v. Tide Water Power Co.*, 189 N.C. 120, 125, 126 S.E. 307, 310 (1925), *overruled on other grounds in Nicholson v. Hugh Chatham Memorial Hosp., Inc.*, 300 N.C. 295, 266 S.E.2d 818 (1980), and *overruled in Johnson*, 327 N.C. 283, 395 S.E.2d 85. The *Johnson* court characterized this interpretation of *Bailey* as "erroneous" and "simply . . . wrong." *Johnson*, 327 N.C. at 297, 395 S.E.2d at 93.

79. See *Williamson v. Bennett*, 251 N.C. 498, 112 S.E.2d 48 (1960), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *Hinnant*, 189 N.C. 120, 126 S.E. 307.

80. 143 N.C. 398, 55 S.E. 778 (1906), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

81. *Id.* at 401-02, 55 S.E. at 779.

accompanied by physical injury, cannot be considered as an element of damage."⁸² When the fright became manifested physically, however, the plaintiff could recover for the physical injury; the court held that "the general principles of the law of torts support a right of action for physical injuries resulting from negligence, whether wilful or otherwise, none the less strongly because the physical injury consists of a wrecked nervous system instead of lacerated limbs."⁸³

The *Kimberly* decision precipitated much of the later confusion regarding a "requirement" of physical injury either occurring contemporaneously with the event causing the emotional distress or appearing as a later physical manifestation of the distress before a plaintiff could recover for its negligent infliction. The *Kimberly* court faced only a claim based on fright, not emotional distress, and allowed recovery for fright only upon a showing that the fright resulted in some physical injury.⁸⁴ In dictum, the court went on to discuss mental upsets generally, which may have sparked the application of a physical injury requirement to cases for severe mental distress as well as those for fright. *Kimberly* also recast the claim by characterizing the mental upset itself as a distinct type of physical injury, rather than as an emotional injury and an additional, separate physical injury.⁸⁵ The decision presumably would allow recovery for pure mental anguish, absent any actual physical injury caused directly by defendant's negligence, so long as that anguish was manifested in such a way as to prove its authenticity.⁸⁶

In contrast, a plaintiff could recover for distress arising from negligent infliction of fright only if the plaintiff could show a literal physical injury either occurring contemporaneously with, or as a direct result of, the negligent act. The differences between mere fright and mental distress gradually became blurred,⁸⁷ however, and the requirement of an accompanying physical injury in claims that arose out of negligently inspired fright was attributed to mental

82. *Id.* at 403, 55 S.E. at 780. The court went on to note that "where the fright occasions physical injury, not contemporaneous with it, but directly traceable to it, the courts are hopelessly divided." *Id.*

83. *Id.* at 404, 55 S.E. at 780.

84. The *Kimberly* court confirmed that although fright and nervousness could not themselves be considered an injury within the court's usage of the term, any detrimental health effects brought about "naturally and directly" by the fright or nervousness would be compensable. *Id.* (quoting unreported opinion of court below).

85. *Id.* at 403-04, 55 S.E. at 780. For example, the *Kimberly* court held that a physical injury may consist of a "wrecked nervous system." *Id.* Professor Byrd notes that

[p]hysical injury, as incorporated in the rule, is not used in the sense of an injury to a specific part of the body, such as a cut, broken bones, or damage to an internal organ. . . .

Impairment of health, loss of bodily power, or sickness, without proof of any specific injury, has been held to constitute a physical injury.

Byrd, *supra* note 37, at 458.

86. The physical injury requirement originally served as a "vehicle used by the court to distinguish harm of [great] magnitude from less serious interferences which, if a multitude of suits are to be avoided, everyone must be left to absorb to some degree." Byrd, *supra* note 37, at 458.

87. See, e.g., *Williamson v. Bennett*, 251 N.C. 498, 507, 112 S.E.2d 48, 54 (1960) (court characterized case as belonging to "category [of] cases of fright, anxiety, and other emotional distress, unaccompanied by physical injury"), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

Even when claims for mental anguish are distinguished from those for fright, the mental anguish claims often receive generalized treatment. "There is a tendency in the decisions to treat all mental anguish claims alike, and as a result distinctions are seldom made between intentional con-

anguish claims as well.⁸⁸ Eventually, the courts came to apply a "physical injury requirement"⁸⁹ to almost all cases involving some form of unquantifiable emotional distress.⁹⁰

By this time, a crucial distinction had been lost. Though the courts later returned to *Bailey's* early suggestion that negligently inflicted emotional distress warranted compensation even when unaccompanied by an underlying claim, they retained the physical injury requirement. Claims prompted by the negligent *frightening* of a plaintiff could not succeed without a contemporaneous or resultant physical injury, the claim itself being founded upon the physical injury. Although claims for negligently inflicted emotional distress once had been considered qualitatively different from claims prompted by "mere fright," the separation between the two faded away. In *Williamson v. Bennett*,⁹¹ the court held that "it is almost the universal opinion that recovery may be had for mental or emotional disturbance in ordinary negligence cases where, coincident in time and place with the occurrence producing the mental stress, some actual physical injury also resulted directly from the defendant's negligence."⁹²

Other cases addressed the collateral issue of bystander recovery, as differentiated from a mental or emotional injury inflicted by the tortfeasor directly upon the victim.⁹³ In *Bailey v. Long*⁹⁴ the court allowed plaintiff-husband to recover for the "great pain and mental anguish . . . to his feelings and sympathies"⁹⁵ induced by "witnessing the agony and suffering of his said wife,"⁹⁶ when his wife

duct and negligence, or between the determination of the extent of liability and the decision if a cause of action exists at all." Byrd, *supra* note 37, at 442.

88. See, e.g., *Stanback v. Stanback*, 297 N.C. 181, 198-99, 254 S.E.2d 611, 623 (1979) ("clear that plaintiff must show some physical injury resulting from the emotional disturbance caused by defendant's alleged conduct"), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *King v. Higgins*, 272 N.C. 267, 269, 158 S.E.2d 67, 69 (1967) (citing *Williamson*, 251 N.C. 498, 112 S.E.2d 48, and disallowing recovery absent evidence of "any disfiguring injury"); *Williamson*, 251 N.C. at 503, 112 S.E.2d at 52 (recovery for mental or emotional distress allowed when accompanied by simultaneous physical impact or injury), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *Woodell v. Pinehurst Surgical Clinic, P.A.*, 78 N.C. App. 230, 232, 336 S.E.2d 716, 718 (1985) (explaining that "non-permanent discomfort (physical pain and suffering, mental anguish and emotional distress) with no physical injury [is not compensable]"), *aff'd*, 316 N.C. 550, 342 S.E.2d 523 (1986) (per curiam), *overruled in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *Wesley v. Greyhound Lines*, 47 N.C. App. 680, 690, 268 S.E.2d 855, 862 (1980) (requiring physical impact or injury in conjunction with mental trauma or emotional disturbance); *McDowell v. Davis*, 33 N.C. App. 529, 537, 235 S.E.2d 896, 901 (applying *Williamson* standard), *disc. rev. denied*, 293 N.C. 360, 237 S.E.2d 848 (1977), *overruled in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

89. See *Williamson*, 251 N.C. 498, 112 S.E.2d 48, *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85; *Wyatt v. Gilmore*, 57 N.C. App. 57, 290 S.E.2d 790 (1982).

90. See, e.g., cases cited *supra* note 3.

91. 251 N.C. 498, 112 S.E.2d 48 (1960), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

92. *Id.* at 503, 112 S.E.2d at 52.

93. For a discussion of cases involving relational interests or "bystander recoveries," see Byrd, *supra* note 37, at 448-452. See also cases cited *supra* note 59 (listing the major restrictions and standards applicable to bystander recovery in other jurisdictions).

94. 172 N.C. 661, 90 S.E. 809 (1916). The *Bailey* decision may be read as bridging the gap between permitting recovery for limited factual situations in which plaintiff's claim arose from a contractual relationship with defendant, and more liberal reasoning that focuses more specifically on the nature of the injury itself, instead of its source. For a discussion of the contractual relationship in *Bailey*, see *supra* notes 75-78 and accompanying text.

95. *Bailey*, 172 N.C. at 662, 90 S.E. at 809.

96. *Id.* at 663, 90 S.E. at 810.

died from illnesses traceable to defendant-doctor's negligent failure to provide a habitable room for her.⁹⁷ The court held that plaintiff could recover not only the sums he expended on her medical services but also for his own mental distress.⁹⁸

The *Bailey* decision was followed by *Hipp v. E.I. Dupont de Nemours & Co.*,⁹⁹ in which the court reviewed *Bailey* and confirmed that a spouse could recover for her own personal mental injuries caused by a third party's negligent physical injury to the other spouse.¹⁰⁰ The court limited availability of the claim to spouses and placed it in the context of damages relating to loss of consortium; "children or other dependent relatives" could not have access to the claim because "[t]he wife's cause of action [arose] from the nature of the relationship created by the contract of marriage."¹⁰¹

The court sharply abridged this apparent ability of a bystander to recover for his own mental injuries, albeit occasioned by a more direct injury to another person, in *Hinnant v. Tide Water Power Co.*¹⁰² In *Hinnant* the North Carolina Supreme Court addressed both the general standards for recovery for negligently inflicted emotional distress and the availability of the claim to a third-party bystander. The court held that plaintiff-wife could not recover for the mental distress and "serious nervous shock" she experienced after seeing her husband in a "broken [and] mashed" condition after a train collision caused by the defendant's negligence.¹⁰³ The husband died of his injuries, causing plaintiff's nervous system to be "permanently impair[ed] and weaken[ed]."¹⁰⁴ The *Hinnant* court conclusively denied the availability of bystander damages, holding that

[i]n the law mental anguish is restricted, as a rule, to such mental pain and suffering as arises from an injury or wrong to the person himself, as distinguished from that form of mental suffering which is the accompaniment of sympathy or sorrow for another's suffering, or which

97. *Id.* at 661-62, 90 S.E. at 809.

98. *Id.* at 663, 90 S.E.2d at 810. The plaintiff in *Bailey* was a "bystander" in the sense that the physical injury negligently and directly inflicted upon his wife caused his mental distress.

99. 182 N.C. 9, 108 S.E. 318 (1921).

100. *Id.* at 19, 108 S.E. at 322-23. The plaintiff-wife's husband became permanently incapacitated while working for defendant. Plaintiff alleged that due to her increasing indebtedness and the despair she felt while watching her husband suffer, "her own nerves and health [were] seriously and permanently shocked, weakened, and impaired; and that by reason of the physical and mental condition of her husband she still continue[d] to suffer in mind and body." *Id.* at 11, 108 S.E. at 319.

101. *Id.* at 19, 108 S.E. at 323.

102. 189 N.C. 120, 126 S.E. 307 (1925), overruled on other grounds in *Nicholson v. Hugh Chatham Memorial Hosp., Inc.*, 300 N.C. 295, 266 S.E.2d 818 (1980), and overruled in *Johnson*, 327 N.C. 283, 395 S.E.2d 85.

103. *Id.* at 120, 126 S.E. at 308. The majority of the court's later decisions refused to allow such recovery. See *Helmstetter v. Duke Power Co.*, 224 N.C. 821, 824, 32 S.E.2d 611, 613 (1945) (Married Women's Act of 1913 eliminated right of either spouse to recover for injury to the other; without a cause of action for loss of consortium, "there is none for mental anguish"), overruled in *Nicholson*, 300 N.C. 295, 266 S.E.2d 818; *Michigan Sanitarium & Benevolent Ass'n v. Neal*, 194 N.C. 401, 403, 139 S.E. 841, 841-42 (1927) (mother could not recover for her mental anguish occasioned by injuries to her son because damages "too remote").

104. *Hinnant*, 189 N.C. at 120, 126 S.E. at 308. The claim can be read as alleging a physical injury, but the court apparently did not find the allegation sufficient.

arises from the contemplation of wrongs committed on the person of another.¹⁰⁵

As a general rule, the court said, "mental suffering, unrelated to any other cause of action, is not alone a sufficient basis for the recovery of substantial damages."¹⁰⁶ More recently, in *Williamson v. Bennett*,¹⁰⁷ the North Carolina Supreme Court refused to allow a plaintiff to recover for mental distress brought on by her fright and fear that she had driven over a child, noting that "this Court has held that there can be no recovery for fright and anxiety, and resultant neurosis, which arises for the safety and well-being of another."¹⁰⁸

In *Johnson* the North Carolina Supreme Court took stock of all the confusion and conflict in this area of law and chose to define a plaintiff's ability to recover by principles and standards of foreseeability and proximate cause without additional limitation.¹⁰⁹ The holding is arguably in accord with prior North Carolina case law, insofar as the earliest cases are concerned; however, by discounting several decades of subsequent case law as erroneous or unfounded, the court adopted a markedly revisionist approach to precedent.¹¹⁰ The court

105. *Id.* at 129, 126 S.E. at 312.

106. *Id.* The *Hinnant* court went on to list a number of exceptions to this rule under which plaintiffs could bring claims for mental or emotional damages without an underlying or supporting cause of action; among these exceptions were claims for a breach of a promise to marry, and the negligent delivery of a telegram. *Id.*

107. 251 N.C. 498, 112 S.E.2d 48 (1960), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85.

108. *Id.* at 508, 112 S.E.2d at 55. The *Williamson* court denied recovery to the plaintiff because the mental injury was not proximately caused by any negligent act of the defendant. While the defendant did negligently collide with plaintiff's car, the injury itself was caused by plaintiff's unfounded fear that she had injured a child; "[i]n short, she was not frightened by what actually happened but by what might have happened," and the defendant was under no duty to guard against such an unlikely reaction. *Id.*

109. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97. The application of these standards, without more, has been criticized extensively by numerous courts and scholars. *See, e.g.,* *Thing v. La Chusa*, 48 Cal. 3d 644, 662, 771 P.2d 814, 826, 257 Cal. Rptr. 865, 877 (1989) (foreseeability "not a realistic indicator of potential liability and does not afford a rational limitation on recovery"); *see also* Diamond, *supra* note 60, at 493 ("It is apparent that the use of a foreseeability standard to assess non-physical damages often is inconsistent with the kinds of injuries that courts, and society, are prepared to compensate. . . . [The courts are] not willing to 'open the floodgates' by compelling compensation for all foreseeable mental distress.").

110. Interview with Robert G. Byrd, Professor of Law, University of N. C. School of Law, in Chapel Hill, N.C. (Oct. 5, 1990). In its attempt to mold its prior decisions into a more cogent and logical progression, the court adopted what could be termed a "historical revisionist" view, according to Professor Byrd.

For example, the court explained the physical injury requirement enunciated in *Stanback v. Stanback* as follows:

While we said in *Stanback* that a showing of "physical injury" was required, we also relied upon our earlier statement in *Kimberly*, indicating that emotional distress is *one type* of physical injury, and held that the trial court's dismissal of the plaintiff's claim must be reversed. Thus, the statement in *Stanback* is, to some extent at least, at odds with its holding. Further, the awkward two-step analysis of *Stanback* and *Kimberly* — by which we implied that physical injury was required, but then defined emotional distress as a *type* of physical injury for which a plaintiff could recover — was entirely unnecessary in light of the analyses contained in our prior cases which *reached the same result* in a more straightforward and less cumbersome fashion. . . . [O]ur earlier cases did *not* require any physical impact or injury in *addition to* the mental or emotional injury itself; instead, our earlier cases simply treated emotional distress as any other type of injury — compensable if the plaintiff shows that the injury was foreseeably and proximately caused by the defendant's negligence.

aligned *Johnson* with North Carolina's earliest cases on the issue,¹¹¹ and accurately "reinstated" the original principles applied by the first courts allowing recovery.¹¹²

The court contended, in essence, that *Johnson* is not new law. "While admittedly some of our opinions have suggested contrary results," the court observed, "the overwhelming weight of this Court's opinions for the past one hundred years leads us to the conclusion that neither a physical impact, a physical injury, nor a subsequent physical manifestation of emotional distress is an element of the tort of negligent infliction of emotional distress."¹¹³ The implications of this view are far reaching. A practicing North Carolina attorney observed:

Since *Ruark* held that this holding has been the law in North Carolina since the *Hancock* decision in 1905, it must be understood that the holding and reasoning was that this had been our law for quite some time, despite court decisions to the contrary. Accordingly, every lawyer handling tort claims must now look into the past to determine how many emotional distress claims have had life breathed into them through the decision in *Ruark*. . . . My opinion is that there are emotional distress claims which have been revived by *Ruark* which are now buried in our "retired" files.

This raises issues of legal malpractice. . . . If our clients have had new claims occur because of *Ruark*, or have had an old claim that we did not prosecute revived, it is our ethical obligation to review our old files to the extent necessary and to determine which of our clients have emotional distress claims which now ethically must be prosecuted.¹¹⁴

Insurance companies are likely to feel keenly the problems occasioned by dormant claims for negligent infliction of emotional distress if the companies believe they had addressed and settled the claims of minors or, indeed, any party for whom the applicable statute of limitations has not yet run.¹¹⁵

Johnson, 327 N.C. at 302, 395 S.E.2d at 96 (discussing *Stanback v. Stanback*, 297 N.C. 181, 198-99, 254 S.E.2d 611, 623 (1979), *disapproved in Johnson*, 327 N.C. 283, 395 S.E.2d 85 (citations and footnote omitted)).

111. See cases cited *supra* note 65.

112. See cases cited *supra* note 70.

113. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97. The court suggested:

Such misstatements have led some to believe that an action for negligent infliction of emotional distress may not be maintained absent some physical impact, physical injury or subsequent physical manifestation of the emotional distress, and also that recovery may not be had for emotional distress caused by a plaintiff's concern for another person.

Id. at 290-91, 395 S.E.2d at 89. This observation probably is understated; see *McCain, Johnson v. Ruark Obstetrics: Land Mine For The Unwary*, 9 LML TODAY, Nov. 1990, at 2, 2.

114. *McCain*, *supra* note 113, at 2.

115. *Id.* The statute of limitations for actions based on personal injury is three years, commencing on the date on which the claim accrues. N.C. GEN. STAT. § 1-15 (1983). When a minor is involved, the statute does not begin to run until the minor reaches the age of 18. *Id.* § 1-17(a)(1) (Supp. 1990).

McCain warned that *Johnson* "also raises a spectre as to whether these insurance companies will look to their defense counsel for failure to obtain releases of all claims," when the insurance companies believed they had "bought their peace in previous settlements." *McCain*, *supra* note 113, at 2.

Read literally, *Johnson* appears to invite a flood of litigation,¹¹⁶ if the physical injury requirement in fact operated to exclude many cases in the first place.¹¹⁷ *Johnson* held, however, that plaintiffs must allege severe mental distress of the sort generally recognized by professionals trained in the diagnosis and treatment of mental disorders.¹¹⁸ Given that such diagnoses in all probability would depend on the presence of the physical ailments that commonly accompany severe mental distress,¹¹⁹ the rejection of a physical injury requirement may have only limited effect.¹²⁰

Johnson's most significant impact probably will be in the area of bystander recovery.¹²¹ The *Johnson* court held both that "the relationship between the plaintiff and the other person for whose welfare the plaintiff is concerned"¹²² is relevant to the question of foreseeability and that the plaintiff may "recover for his or her severe emotional distress arising due to concern for another person, if the plaintiff can prove that he or she has suffered such severe emotional distress as a proximate and foreseeable result of the defendant's negligence."¹²³ Ques-

116. The *Johnson* decision sparked a great deal of public interest, to say the least. *Johnson* was interpreted as "a ruling likely to increase the liability woes of doctors and enable plaintiffs to win damages for mental anguish in a variety of cases." Raleigh News & Observer, Aug. 30, 1990, at B1, col. 2. An attorney for the North Carolina Association of Defense Attorneys suggested: "What they may have done is open a Pandora's box for all sorts of tort actions, not just medical malpractice actions." *Id.* at B2, col. 3.

The criticism of *Johnson* took on partisan tones as November elections approached; in the running for Supreme Court judgeships were incumbent Justices Exum, Webb, and Whichard. Republican Governor James G. Martin "criticized Democratic judges for being too eager to rewrite state law and to make it too easy for plaintiffs to win large civil verdicts." *Id.*, Oct. 17, 1990, at B1, col. 2. News reports observed that "Mr. Martin also has singled out specific decisions [*Johnson* and a death penalty case] by the Democratic incumbents and has complained that the court has a liberal, activist approach to the law." *Id.*, Oct. 18, 1990, at B3, col. 1.

117. Plaintiffs have satisfied the physical injury requirement by alleging, for example, general "[i]mpairment of health, loss of bodily power, or sickness, without proof of any specific injury." Byrd, *supra* note 37, at 458 (citing *Sparks v. Tennessee Mineral Prods. Corp.*, 212 N.C. 211, 212, 193 S.E. 31, 32 (1937) (plaintiff suffered from loss of weight, nervousness, and other ailments); *Kimberly v. Howland*, 143 N.C. 398, 403, 55 S.E. 778, 780 (1906) (plaintiff affected with shocked nervous system), *dissapproved in Johnson*, 327 N.C. at 295, 302, 395 S.E.2d at 92, 96); *see also* *Arthur v. Henry*, 157 N.C. 438, 439, 73 S.E. 211, 212 (1911) (plaintiff "suffered in body and mind" and was "made sick"); *Watkins v. Kaolin Mfg. Co.*, 131 N.C. 536, 538, 42 S.E. 983, 984 (1902) (plaintiff became "almost helpless" and "could not go about her daily duties"). *But cf.* *Woodell v. Pinehurst Surgical Clinic, P.A.*, 78 N.C. App. 230, 232, 336 S.E.2d 716, 718 (1985) (neither nonpermanent discomfort nor physical pain satisfies physical injury requirement), *aff'd*, 316 N.C. 550, 342 S.E.2d 523 (1986) (per curiam), *overruled in Johnson*, 327 N.C. at 301, 395 S.E.2d at 95.

Professor Byrd observed that "[u]nder these holdings it is probable that a 'physical' injury can be shown in any case in which significant mental or emotional harm has occurred." Byrd, *supra* note 37, at 458.

118. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

119. *See* DIAGNOSTIC AND STATISTICAL MANUAL, *supra* note 61, at 235-39. Post traumatic stress disorders may be manifested by such things as nightmares, hyperalertness, difficulty in concentrating, sleep disorders, and diminished responsiveness to external stimuli. *Id.* at 248.

120. Interestingly, the language of *Johnson* suggests that although expert psychiatric testimony is not required in order to substantiate a claim, its use might well become the accepted standard. Thus, plaintiffs could be obligated to incur the additional expenses of psychiatric examination and testimony, which might also serve to curb any potential increase in fraudulent claims.

121. The dissent contended that *Johnson* represented the "addition of [a] new layer of liability to bystanders." *Johnson*, 327 N.C. at 312, 395 S.E.2d at 102 (Meyer, J., dissenting).

122. *Id.* at 305, 395 S.E.2d at 98.

123. *Id.* at 304, 395 S.E.2d at 97.

tions of foreseeability traditionally are reserved to the trier of fact; thus, unless a claim is so unsound on its face as to warrant dismissal by the trial judge, the case is potentially assured of reaching a jury.¹²⁴ Prior to *Johnson*, the "physical injury requirement" served as a clearly defined, if merely arbitrary, screening device.¹²⁵

Given the widespread perception that *Johnson* opened new doors to more extensive liability,¹²⁶ the decision may prompt an increased effort to recover for negligent infliction of emotional distress and, naturally, a corresponding increase in litigation.¹²⁷ Because claims must be examined on their facts, plaintiffs must overcome few preliminary obstacles.¹²⁸ Plaintiffs need allege only that a defendant acted negligently, that the foreseeable result was severe emotional distress or mental anguish on the part of the plaintiff, and that the plaintiff did suffer severe emotional distress as a result.¹²⁹

Characterizing *Johnson* as exemplary of the "[new] policy of this jurisdiction to extend an infinite responsibility to everyone who has suffered,"¹³⁰ Justice Meyer noted in dissent that an increase in litigation and a heightened awareness of liability on the part of service providers, and insurance companies in particular, could impose severe societal costs.¹³¹ The danger of escalating insurance premiums and decreasing availability of obstetric care are valid concerns. In addition, the risk of an excessive recovery by plaintiffs is real.¹³² Damages for emotional distress are regularly compensated in wrongful death actions, even if not specifically requested as an element of damages and despite instructions to

124. The court held that "[q]uestions of foreseeability and proximate cause must be determined under all the facts presented, and should be resolved on a case-by-case basis by the trial court and, where appropriate, by the jury." *Id.* at 305, 395 S.E.2d at 98. As a result of *Johnson*, plaintiffs who bring negligent infliction of emotional distress claims may have enhanced bargaining power during settlement negotiations.

125. As one commentator noted:

[The] so-called "physical manifestation" requirement served two major purposes. First, it served to limit the potential liability of defendants since a cause of action was only available in those cases where physical injury could be established. Second, it operated as an element of proof that the plaintiff suffered actual mental distress, and thus that the suit was not trivial or fraudulent.

Comment, *supra* note 60, at 795.

126. The North Carolina legal community's reaction to *Johnson* was dominated by concerns that the case ushered in unlimited liability on the part of defendants. The decision became a popular topic for various political campaigns underway in October and November of 1990; critics frequently pointed to *Johnson* as an example of the need for judicial reform in North Carolina and as a reason to elect new justices to the North Carolina Supreme Court. See news articles cited *supra* note 116.

127. For a discussion of the potential revival of negligent infliction of emotional distress claims that were left unlitigated due to the absence of physical injury, see *supra* notes 113-15 and accompanying text.

128. As discussed earlier, however, the severity of the mental distress may be subject to expert psychiatric evaluation. See *supra* notes 118-19 and accompanying text.

129. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

130. *Id.* at 312, 395 S.E.2d at 102 (Meyer, J., dissenting).

131. *Id.* (Meyer, J., dissenting). See *supra* notes 55-57 and accompanying text.

132. In his dissent, Justice Meyer suggested that when a plaintiff brings an action for both emotional distress and wrongful death, the claims should be joined to curtail the risk of "inconsistent verdicts and double recoveries for the same loss." *Johnson*, 327 N.C. at 314-15, 395 S.E.2d at 103 (Meyer, J., dissenting).

the jury not to include them.¹³³ Assuming that *Johnson* does represent a significant expansion of liability, the financial burden on defendants may be increased if plaintiffs are permitted to bring negligent infliction of emotional distress claims independently of other related claims they may have.¹³⁴

Fortunately, *Johnson's* rejection of the physical injury requirement will promote more honest pleadings. The requirement originally evolved as the courts sought a means of differentiating between serious and trivial claims.¹³⁵ When a plaintiff genuinely does suffer serious physical effects as a result of negligent infliction of emotional distress, the plaintiff will continue to allege such factors to demonstrate the extent of his emotional harm. Such "physical manifestations" as sleeplessness, irritability, and inability to concentrate, however, have been found to satisfy the physical injury requirement when the plaintiff's claim was otherwise valid.¹³⁶ These allegations were often only tangentially related to the basis of the claim and were included in the pleadings for the sole purpose of fulfilling the physical injury requirement. *Johnson*, then, may curb the artificiality formerly involved in alleging negligent infliction of emotional distress. Repudiation of the physical injury requirement should have little, if any, effect on the risk of fraudulent claims; at the very least, its absence should not increase their likelihood.¹³⁷

The long-term import of *Johnson* is less clear. By conclusively rejecting a physical injury requirement and dismissing it as "arbitrary," the majority left the lower courts with no specific threshold requirements or rules of law by which to screen claims for validity. The majority reviewed the tests followed by other jurisdictions¹³⁸ and noted the absence of any "single clear doctrine to which it can be said that a majority of states adhere."¹³⁹ A number of states, however, reject the approach of the "new" North Carolina standard in favor of more definitive guidelines.¹⁴⁰

Illustrative of states that have rejected a standard similar to that established in *Johnson* is a recent California case, *Thing v. La Chusa*.¹⁴¹ In *La Chusa* the

133. *Id.* (Meyer, J., dissenting). Several supreme court opinions have addressed the possibility of "double recovery" in light of the difficulties faced by juries as they try to apportion damages between the distress caused by negligence and the distress caused by the loss of, or injury to, a relation. See cases cited *supra* note 44.

134. Because any award for negligent infliction of emotional distress must first be predicated on a finding of negligence, joinder of these related claims would be eminently sensible. See N.C.R. Civ. PRO. 18(a); see also C. WILSON, NORTH CAROLINA CIVIL PROCEDURE § 18-2, at 356 (1989) ("[Rule 18(a) does not] permit claim-splitting in violation of the common law principle that all damages incurred as a result of a single injury must be recovered in one lawsuit.").

135. See Byrd, *supra* note 37, at 458.

136. See, e.g., cases cited *supra* note 117.

137. One commentator addressed the risk of fraudulent claims as follows: "If recovery is limited to instances where it would be generally viewed as appropriate and not excessive, then, by definition, the defendant's liability is commensurate with the damage that the defendant's conduct caused. Further, the judicial system would not be overburdened by administering fair and proper claims." Comment, *supra* note 60, at 819.

138. *Johnson*, 327 N.C. at 288-90, 395 S.E.2d at 88-89.

139. *Id.* at 290, 395 S.E.2d at 89.

140. For a listing of common restrictions applied in other jurisdictions, see, e.g., cases cited *supra* note 59.

141. 48 Cal. 3d 644, 771 P.2d 814, 257 Cal. Rptr. 865 (1989).

California Supreme Court declared the *Dillon v. Legg*¹⁴² test unworkable due to its tendency to create "ever widening circles of liability."¹⁴³ The *Dillon* decision spawned the "foreseeable plaintiff" test, which facilitated recovery on the basis of "the neutral principles of foreseeability, proximate cause and consequential injury that generally govern tort law."¹⁴⁴

The *Dillon* test required the California courts to take certain factors "into account" to determine foreseeability: specifically, (1) whether the plaintiff was in close proximity to the scene of the accident, (2) whether the plaintiff's emotional distress was the result of the "sensory and contemporaneous observance of the accident," and (3) whether the plaintiff was "closely related" to the victim.¹⁴⁵ The *Dillon* test did not call for a physical injury to the plaintiff, but was otherwise more restrictive regarding the class of plaintiffs eligible to recover than is *Johnson*.

In place of *Dillon*, the *La Chusa* court set forth a list of requirements that could be waived or excused only under "exceptional circumstances."¹⁴⁶ The court held that

a plaintiff may recover damages for emotional distress caused by observing the negligently inflicted injury to a third person if, but only if, said plaintiff: (1) is closely related to the injured victim; (2) is present at the scene of the injury-producing event at the time it occurs and is then aware that it is causing injury to the victim; and (3) as a result suffers serious emotional distress—a reaction beyond that which would be anticipated in a disinterested witness and which is not an abnormal response to the circumstances.¹⁴⁷

The court explained that reliance on the concepts of foreseeability and proximate causation was unsatisfactory because "reliance on foreseeability of injury alone in finding a duty, and thus a right to recover, is not adequate when the damages sought are for an intangible injury."¹⁴⁸ Significantly, the *Dillon* test condemned as unworkable by the *La Chusa* court held that the plaintiff's right to recover was limited only by the "general rules of tort law . . . long applied to all other types of injury."¹⁴⁹ The abandoned *Dillon* standards are generally consistent with those recently adopted in *Johnson*.¹⁵⁰

The *Johnson* court attempted to rest a controversial holding on undeniable

142. 68 Cal. 2d 728, 441 P.2d 912, 69 Cal. Rptr. 72 (1968) (en banc).

143. *La Chusa*, 48 Cal. 3d at 653, 771 P.2d at 819, 257 Cal. Rptr. at 870.

144. *Dillon*, 68 Cal. 2d at 737, 441 P.2d at 918, 69 Cal. Rptr. at 78.

145. *Id.* at 740-41, 441 P.2d at 920, 69 Cal. Rptr. at 80.

146. *La Chusa*, 48 Cal. 3d at 668 n.10, 771 P.2d at 829 n.10, 257 Cal. Rptr. at 880 n.10.

147. *Id.* at 667-68, 771 P.2d at 829-30, 257 Cal. Rptr. at 880-81 (footnotes omitted).

148. *Id.* at 664, 771 P.2d at 826, 257 Cal. Rptr. at 877.

149. *Dillon*, 68 Cal. 2d at 746, 441 P.2d at 924, 69 Cal. Rptr. at 84.

150. *Johnson*, 327 N.C. at 304-05, 395 S.E.2d at 97-98. The *Johnson* standard is as follows:

[T]o state a claim for negligent infliction of emotional distress, a plaintiff must allege that (1) the defendant negligently engaged in conduct, (2) it was reasonably foreseeable that such conduct would cause the plaintiff severe emotional distress (often referred to as 'mental anguish'), and (3) the conduct did in fact cause the plaintiff severe emotional distress.

Id. at 304, 395 S.E.2d at 97.

principles of fairness. Under traditional and certainly well-established tort principles of recovery, any party that shows his injury to be the foreseeable or proximate result of the unreasonable negligence of another is entitled to be compensated for that injury.¹⁵¹ Under this doctrine, the mental anguish caused by the stillbirth of a child and attributable to the negligence of professionals entrusted to provide competent medical care qualifies as a compensable injury.¹⁵² Against this precept, however, the courts should have weighed the interests of society in being free from unreasonable liability and in being permitted to undertake activities that may involve risk.¹⁵³ In that obligation, the *Johnson* decision falls short. The repercussions of the case range far beyond the limited holding warranted by the narrow facts before the *Johnson* court; indeed, the same standards apply to a mother who loses her unborn child to professional negligence as to a bystander who suffers emotional distress as a result of an injury to a third party caused by only ordinary negligence.

The very nature of the tort of negligent infliction of emotional distress mandates a limitation on the ability to recover.¹⁵⁴ Standards of "foreseeability" and "proximate cause" are rarely applied in their purest sense, but traditionally are tempered by policy considerations and the common-sense notion that after the reaching of an admittedly arbitrary line, forcing defendants to assume the risk of liability simply becomes unfair. When the chain of causation is real but nonetheless grossly attenuated, no principles of deterrence may be served, and the ability to guard against causing injury to another is so negligible as to be reduced to a question of luck. The experience of other jurisdictions warns that standards of foreseeability and proximate cause in negligent infliction of emotional distress cases foster uncertainty:

It is apparent that reliance on foreseeability of injury alone in finding a duty, and thus a right to recover, is not adequate when the damages sought are for an intangible injury. In order to avoid limitless liability out of all proportion to the degree of a defendant's negligence, and against which it is impossible to insure without imposing unacceptable costs on those among whom the risk is to be spread, the right to recover for negligently caused emotional distress must be limited.¹⁵⁵

The need for fact-specific restrictions is made all the more compelling by the court's assurance that "ordinary negligence will suffice."¹⁵⁶ Further, the

151. W. PROSSER, LAW OF TORTS § 30, at 164-65 (5th ed. 1984).

152. The Johnsons' claim was dismissed prior to trial, and at the time of this writing no finding of negligence or other wrongdoing has yet been lodged against defendants as a result of the trial.

153. See *United States v. Carroll Towing Co.*, 159 F.2d 169, 173 (2d Cir. 1947). In *Carroll Towing*, Judge Learned Hand introduced his famous equation distinguishing those risks that are worth taking from those that are not. For the text of the equation and Justice Meyer's application of the principle to *Johnson*, see *supra* notes 55-57.

154. This is particularly true with respect to bystander recovery cases that turn on the relational interest between the bystander plaintiff and the directly injured victim; such situations "create the possibility of liability to a large number of people Under these circumstances the fear of an indefinite liability is a legitimate one, and the need to impose reasonable limits upon the extent of a defendant's responsibility clearly exists." Byrd, *supra* note 37, at 448.

155. *Thing v. La Chusa*, 48 Cal. 3d 644, 664, 771 P.2d 814, 826-27, 257 Cal. Rptr. 865, 877-78 (1989).

156. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

court stipulated that the validity of negligent infliction of emotional distress claims must be determined case by case;¹⁵⁷ however, an ad hoc approach surely will foster further uncertainty in this area of law, resulting in the disparate treatment of cases.¹⁵⁸ The *Johnson* court also explained that "our trial courts have adequate means available to them for disposing of improper claims for negligent infliction of emotional distress and for adjusting excessive or inadequate verdicts."¹⁵⁹ The court quoted *Chappell v. Ellis*:¹⁶⁰

But it is urged that the principle [of recovery for mental anguish,] . . . if carried out to its fullest extent, would directly lead to the recovery of damages for all kinds of mental suffering. It may be, but we feel compelled to carry out a principle only to its necessary and logical results, and not to its furthest theoretical limit, in disregard of other essential principles.¹⁶¹

The experience of other jurisdictions suggests that the court is overly optimistic. *Dillon v. Legg*¹⁶² enumerated specific factors to be considered by the trial judges just as the *Johnson* court did, but in practice the standard proved to be too expansive. The recent *La Chusa*¹⁶³ decision recognized the failings of *Dillon*, and spoke to the very hazards that the North Carolina Supreme Court invites in *Johnson*:

The *Dillon* court anticipated and accepted uncertainty in the short term in application of its holding, but was confident that the boundaries of this [negligent infliction of emotional distress] action could be drawn in future cases. In sum . . . the *Dillon* court was satisfied that trial and appellate courts would be able to determine the existence of a duty because the court would know it when it saw it. Underscoring the questionable validity of that assumption, however, was the obvious and unaddressed problem that the injured party, the negligent tortfeasor, their insurers, and their attorneys had no means short of

157. *Id.* at 305, 395 S.E.2d at 98.

158. This approach is consistent with the North Carolina Supreme Court's traditional treatment of negligent infliction of emotional distress claims. Professor Byrd writes that

a tendency exists to view cases involving mental anguish claims on an ad hoc basis, an approach that the court itself suggested in the *Williamson* case. Apart from the general unsatisfactory nature of this approach in deciding important legal issues, it involves a real danger that the sensible and sound development in the law that has occurred will unintentionally be undermined in later decisions, and the uncertainties that arise out of the *Stanback* case illustrate this danger.

Byrd, *supra* note 37, at 468. The "danger" predicted by Professor Byrd was realized in *Johnson*, which disapproved both *Stanback v. Stanback*, 297 N.C. 181, 254 S.E.2d 611 (1979), *disapproved on other grounds in* *Dickens v. Puryear*, 302 N.C. 437, 276 S.E.2d 325 (1981), and *Williamson v. Bennett*, 251 N.C. 498, 112 S.E.2d 48 (1960). *Johnson*, 327 N.C. at 299-304, 396 S.E.2d at 94-97.

159. *Johnson*, 327 N.C. at 306, 395 S.E.2d at 98.

160. 123 N.C. 259, 31 S.E. 709 (1898).

161. *Johnson*, 327 N.C. at 306, 395 S.E.2d at 98 (quoting *Chappell*, 123 N.C. at 263, 31 S.E. at 711). *Contra* *Thing v. La Chusa*, 48 Cal. 3d 644, 673, 771 P.2d 814, 833, 257 Cal. Rptr. 865, 884 (1989) (Kaufman, J., concurring). The *La Chusa* court observed that "*Dillon's* confident prediction that future courts would be able to fix just and sensible boundaries on bystander liability has been found to be wholly illusory—both in theory and in practice." *Id.* (Kaufman, J., concurring).

162. 68 Cal. 2d 728, 441 P.2d 912, 69 Cal. Rptr. 72 (1968) (en banc). For a listing of the *Dillon* factors, see *supra* text accompanying note 145.

163. 48 Cal. 3d 644, 771 P.2d 814, 257 Cal. Rptr. 865 (1989).

suit by which to determine if a duty such as to impose liability for damages would be found in cases other than those that were "on all fours" with *Dillon*.¹⁶⁴

The *Johnson* decision suggests that bringing every case to trial is acceptable; because the case offers no rules or standards by which a claim might be challenged, the effect of *Johnson* may be to discourage the worthwhile goal of out-of-court settlement.¹⁶⁵

Moreover, the *Johnson* court erroneously elected to recognize the same standards for recovery for both directly injured plaintiffs and bystander plaintiffs.¹⁶⁶ The court should have examined the two issues separately and imposed reasonable judicial limits on claims for negligent infliction of emotional distress. The court declined to develop viable alternatives to the rejected physical injury requirement, although other options would permit recovery without opening the door to indeterminate liability. For example, the court could have held that the standards applicable to recovery for punitive damages apply by analogy to recovery for directly injured plaintiffs. Further, the court could have used a modified version of the *La Chusa* test for bystander plaintiffs.

Although imposing meaningful limitations on the right of directly injured victims of negligently inflicted emotional distress to recover for merely part of their injuries rather than to be fully compensated is nearly impossible, an alternative does exist. Instead of focusing the standards of recovery on the nature of the plaintiff's injury, the nature of the defendant's harmful and injurious conduct should be determinative.¹⁶⁷ A standard allowing extensive liability for "ordinary negligence"¹⁶⁸ is too easily met; instead, the court should require plaintiffs to forecast evidence that, if believed, would prove defendant's conduct to be gross, wanton, or willful negligence.¹⁶⁹ Trial courts are already familiar with this standard. Indeed, the standards for recovery for punitive damages are set forth in the North Carolina pattern jury instructions:

[Punitive damages] may only be awarded when the jury finds that the

164. *Id.* at 655, 771 P.2d at 821, 257 Cal. Rptr. at 872.

165. An alternative argument could be made that the absence of a physical injury requirement will in fact encourage settlement by making trial more likely, and summary judgment for defendants less likely. *Contra La Chusa*, 48 Cal. 3d at 655, 771 P.2d at 821, 257 Cal. Rptr. at 872 (discussing the "problem that the injured party, the negligent tortfeasor, their insurers, and their attorneys had no means short of suit by which to determine if a duty such as to impose liability for damages would be found" in cases falling under *Dillon*).

166. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

167. Interview with Samuel O. Southern, Esq., Poyner & Spruill, in Raleigh, N. C. (Jan. 28, 1991). Mr. Southern, expressing the concerns of many practicing attorneys, advocated adoption of a more reasonable limitation to the scope of this tort. Mr. Southern's suggestion: "Plaintiff states a claim for negligent infliction of emotional distress only where the facts alleged in the complaint, if believed, would give rise to a claim for punitive damages." *Id.*

168. *Johnson*, 327 N.C. at 304, 395 S.E.2d at 97.

169. For example, suppose plaintiff brings a claim for negligent infliction of emotional distress on grounds that she suffered severe mental distress after being "almost struck" by defendant's automobile. "One result: if [plaintiff's] claim alleged that she suffered emotional injury when she was 'almost struck' by [defendant's] automobile which he was negligently operating, she would not get to the jury. But if the pleadings and evidence established that [defendant] was drunk and doing 80 mph in a school zone, she would carry her case to the jury." Letter from Samuel O. Southern, Esq. to Tracy L. Hamrick (Jan. 30, 1991).

conduct of the defendant is so outrageous as to justify punishing *him* or making an example of *him*. In a case of alleged negligence, punitive damages may be awarded upon a showing that the negligence was gross, willful, or wanton. Negligence is gross, willful or wanton when the wrongdoer acts with a conscious and intentional disregard of, and indifference to, the rights and safety of others.

Upon a showing of gross, willful or wanton negligence, whether to award punitive damages and, within reasonable limits, the amount to be awarded are matters within the sound discretion of the jury.¹⁷⁰

As a result, the trial courts could apply this standard easily and consistently.

Professor Prosser provides further comment on the nature of gross, wanton, or willful negligence. He would apply the terms to "conduct which is still, at essence, negligent, rather than actually intended to do harm, but which is so far from a proper state of mind that it is treated in many respects as if it were so intended."¹⁷¹ This standard strikes a workable balance between plaintiffs' interests in recovering for emotional injury and the simple inability of the average citizen to proceed through life without ever doing the sort of negligent acts that will, under *Johnson*, permit recovery by emotionally injured plaintiffs.

In the area of bystander recovery, the physical injury requirement was equally ineffective, although the need for a better, more consistent limitation on liability was apparent.¹⁷² The *Johnson* court, by refusing to recognize either the qualitatively different nature of bystander recovery or its very real potential to foster liability beyond all reasonable bounds, leaves to the state courts a complex question and no clues to the answer.

The standards recently adopted by the California Supreme Court in *La Chusa* are reasonable and fair, and offer a valuable blueprint for liability limitation as it pertains to bystander recovery. At the least, they offer a more sensible launching point, if the perimeters of the tort are to be redefined.¹⁷³ The *La Chusa* court recounted the difficulties brought about by *Dillon's* broad rule and concluded that plaintiffs may recover damages for emotional distress incurred during the observation of a negligently inflicted injury to a third person if the bystander plaintiff (1) is "closely related" to the directly injured victim; (2) is "present at the scene of the injury-producing event at the time it occurs and is then aware that it is causing injury to the victim"; and (3) personally suffers

170. NORTH CAROLINA PATTERN INSTRUCTIONS — CIVIL 810.01 (Feb. 1986).

171. W. PROSSER, *supra* note 151, § 34, at 212-13.

172. See *supra* notes 154-58 and accompanying text.

173. One commentator observed that:

The failure of the common law courts to recognize and utilize their ability to change procedural rules and to adjust remedies may contribute significantly to the difficulties they encounter when called upon to expand or contract substantive rights. In such situations they tend to act as if their only choice is between full recovery or none at all, with the burden of proof remaining the same as in most other civil actions. The effect may well be to retard needed reform, to prevent the courts from experimenting with techniques designed to allay fears of the catastrophes changes might bring, and sometimes . . . to replace old problems with equally troubling new ones.

Miller, *The Scope of Liability for Negligent Infliction of Emotional Distress: Making "The Punishment Fit the Crime,"* 1 HAW. L. REV. 35, 36 (1979).

serious emotional distress as a result of the event.¹⁷⁴

The *La Chusa* court would limit recovery to "[close] relatives residing in the same household, or parents, siblings, children, and grandparents of the victim," unless "exceptional circumstances" otherwise justify recovery.¹⁷⁵ The legal relationship between the parties, however, should be given a more generous interpretation than that allowed by the *La Chusa* court. Instead, non-family members should be permitted to overcome the presumption against their recovery if they can conclusively establish the presence of a relationship between themselves and the directly injured victim that is essentially equivalent to familial or marital ties.

The additional requirement that the bystander plaintiff apprehend the injury to the victim while present at the scene of the injury derives from the need to assure that a bystander plaintiff who recovers for negligent infliction of emotional distress does so as a result of his personal reaction to the distressing event, as differentiated from grief or suffering prompted solely by sympathy for the victim. The *La Chusa* court pointed out that "[t]he impact of personally observing the injury . . . [is likely to] distinguish[] the plaintiff's resultant emotional distress from the emotion felt when one learns of the injury or death of a loved one from another, or observes pain and suffering but not the traumatic cause of the injury."¹⁷⁶ In a sense, a bystander plaintiff is compensated for being forced by a defendant's negligence to share in the distressing experience itself. The requirement does not limit a bystander plaintiff's ability to pursue any other tort claims that may arise on the facts.¹⁷⁷

In sum, trial courts deserve more than ambiguous "factors,"¹⁷⁸ although they should be allotted enough discretion to validate claims brought by the bystander plaintiff who may not squarely meet the *La Chusa* requirements but whose claim is nonetheless compelling. Insofar as the physical injury requirement limited recovery by bystander plaintiffs, as it did by directly injured plaintiffs, the added "protection" of the physical injury requirement was so unfairly arbitrary as to outweigh any beneficial function the rule might have served.

The *Johnson* court discredits these concerns and imposes on the trial courts

174. *Thing v. La Chusa*, 48 Cal. 3d 644, 667-68, 771 P.2d 814, 829-30, 257 Cal. Rptr. 865, 880-81. The *La Chusa* court defined serious emotional distress as "a reaction beyond that which would be anticipated in a disinterested witness and which is not an abnormal response to the circumstances." *Id.*

175. *Id.* at 667 n.10, 771 P.2d at 829 n.10, 257 Cal. Rptr. at 880 n.10. The *La Chusa* court limited bystander recovery to persons related by blood or marriage on the presumption that they are more likely to sustain severe emotional distress as a result of the injury to their loved one than would a disinterested witness. The court conceded that "[s]uch limitations are indisputably arbitrary since it is foreseeable that in some cases unrelated persons have a relationship to the victim or are so affected by the traumatic event that they suffer equivalent emotional distress" but argued that "drawing arbitrary lines is unavoidable if we are to limit liability and establish meaningful rules for application by litigants and lower courts." *Id.* at 666, 771 P.2d at 828, 257 Cal. Rptr. at 879.

176. *Id.* at 666 & n.9, 771 P.2d at 828 & n.9, 257 Cal. Rptr. at 879 & n.9.

177. The facts may support a wrongful death claim, for example. For the argument that negligent infliction of emotional distress claims should be joined to related tort claims, see *supra* note 134 and accompanying text.

178. *Johnson*, 327 N.C. at 305, 395 S.E.2d at 98. The "factors" named by the court are readily understandable, but their relative weight is left undefined. *Id.*

the duty to establish new, more meaningful standards for recovery as the tort of negligent infliction of emotional distress develops. The court offers no reason why the most relevant limitations are still to be discovered, or why the job falls to the trial courts. This lack of explanation is particularly disturbing given that the North Carolina Supreme Court's experience with the tort of negligent infliction of emotional distress is extensive.¹⁷⁹

The *Johnson* court's conclusion that the physical injury requirement operated in an arbitrary fashion is supported on varying grounds. The most obvious source of support is the disproportionality between allowing recovery for mental distress damages when accompanied by an insignificant injury as compared to the denial of recovery for truly severe mental anguish when a plaintiff neglected to characterize the symptoms of his distress as "physical" injury.¹⁸⁰ Though the physical injury requirement promoted a valid evidentiary purpose in that it helped to distinguish severe mental distress from momentary or insignificant emotional upsets, its tendency to validate or invalidate emotional distress claims on unfairly arbitrary grounds greatly outweighed its usefulness. The court should be commended for rejecting such a rule, but faulted for not replacing it with a better one.

TRACY L. HAMRICK

179. Justice Meyer observed: "The majority opinion is exceedingly (and in my view unnecessarily) critical of the care this Court has previously exercised in this area. Besides being inaccurate, these statements do nothing to instill confidence in this Court's opinions." *Johnson*, 327 N.C. at 315, 395 S.E.2d at 104 (Meyer, J., dissenting). Justice Webb expressed the same view: "I do not believe the Court of Appeals has been wrong in the way it has interpreted our cases." *Id.* at 318, 395 S.E.2d at 106 (Webb, J., dissenting).

The tone of the majority opinion is surprisingly censorious in its discussion of what it repeatedly identified as miscategorizations of North Carolina law. In light of the court's concession that this area of law is not clear, its insistence on deciding negligent infliction of emotional distress cases on an ad hoc basis without any threshold requirements, and its own multiple contributions to the prevailing uncertainty, this posture is unwarranted.

180. The disproportionality of the requirement has often been criticized in scholarly journals. For example, one commentator noted that "[t]he physical manifestation requirement has been criticized as being overinclusive, in that it permits recovery for demonstrably trivial mental distress accompanied by physical symptoms, and underinclusive, since serious distress is noncompensable absent the happenstance of subsequent physical symptoms." Comment, *supra* note 60, at 801-02.

This criticism has been emphasized by the jurisdictions that have declined to adopt a physical injury requirement. See, e.g., *Culbert v. Sampson's Supermarkets, Inc.*, 444 A.2d 433, 437 (Me. 1982) (physical injury requirement is over- and underinclusive); *James v. Leib*, 221 Neb. 47, 58, 375 N.W.2d 109, 116 (1985) (same). For other criticisms of the requirement, see, e.g., *Taylor v. Baptist Medical Center, Inc.*, 400 So. 2d 369, 374 (Ala. 1981) (rejecting requirement because current medical technology allows assessment of mental injury without reliance on "procrustean principles which have little or no resemblance to medical realities"); *Sinn v. Burd*, 486 Pa. 146, 159, 404 A.2d 672, 678 (1979) (same; medical advances have "discredited these hoary beliefs").

***Ellis v. Northern Star Co.*: Libel in a Business Setting Subject to Mandatory Treble Damages Under North Carolina General Statutes Sections 75-1.1 and 75-16**

In the United States, commercial activities take place in a market system in which private buyers and sellers trade money for goods and services. Federal and state governments provide structure and order to the market system through selective regulation. The realm of regulation includes prohibitions against unfair competition and trade practices. In North Carolina, General Statutes section 75-1.1¹ stands at the center of the law of unfair competition and trade practices. Section 75-1.1, enacted in 1969,² is enforceable in a private damage action brought by an aggrieved competitor or consumer; the accompanying section 75-16 provides automatic trebling of damages awarded in such an action.³ To establish a violation of section 75-1.1, a plaintiff must prove that the defendant engaged in “[u]nfair methods of competition” or “unfair or deceptive acts or practices in or affecting commerce.”⁴ The statutory language of section 75-1.1 does not delineate the scope of its protection; rather, the courts define the parameters of the statute through the process of “judicial inclusion and exclusion.”⁵

In *Ellis v. Northern Star Co.*⁶ the North Carolina Supreme Court broadened the realm of protection provided by section 75-1.1 by holding that libel per se in a business setting constitutes an unfair or deceptive practice in or affecting commerce in violation of the section.⁷ The significance of this holding stems from the automatic treble damages that accompany a violation of section 75-1.1,⁸ whereas the common-law remedy for libel per se consists merely of actual and punitive damages.⁹ This extension of the protective shield of section 75-1.1 oc-

1. N.C. GEN. STAT. § 75-1.1 (1988). Section 75-1.1 provides, in pertinent part, that “[u]nfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are declared unlawful.” *Id.* § 75-1.1(a). For a thorough discussion of § 75-1.1, see Aycock, *North Carolina Law on Antitrust and Consumer Protection*, 60 N.C.L. REV. 205, 210-23 (1982).

2. Act of June 12, 1969, ch. 833, 1969 N.C. Sess. Laws 930 (codified as amended at N.C. GEN. STAT. § 75-1.1 (1988)).

3. N.C. GEN. STAT. § 75-16 (1988). Section 75-16 provides:

If any person shall be injured or the business of any person, firm or corporation shall be broken up, destroyed or injured by reason of any act or thing done by any other person, firm or corporation in violation of the provisions of this Chapter, such person, firm or corporation so injured shall have a right of action on account of such injury done, and if damages are assessed in such case judgment shall be rendered in favor of the plaintiff and against the defendant for treble the amount fixed by the verdict.

Id.

4. *Id.* § 75-1.1(a). For the text of section 75-1.1(a), see *supra* note 1.

5. Federal Trade Comm’n v. Raladam Co., 283 U.S. 643, 648 (1931); see *infra* text accompanying note 58.

6. 326 N.C. 219, 388 S.E.2d 127 (1990).

7. *Id.* at 225, 388 S.E.2d at 131. For the North Carolina common-law definition of libel per se, see *infra* text accompanying notes 47-48.

8. See *supra* note 3 and accompanying text.

9. *Ellis*, 326 N.C. at 227, 388 S.E.2d at 132.

curred with minimal justification by the court,¹⁰ and despite a strong dissent that the communication in question was not even a libel *per se*.¹¹

This Note analyzes the *Ellis* decision's broadening of section 75-1.1 in light of prior case law concerning libel *per se* in a business context and unfair trade practices. The analysis includes a comparison of the North Carolina unfair trade practices scheme with that used in several major jurisdictions. This Note concludes that the expansion of section 75-1.1 in *Ellis* was not well supported, and was a step towards over-restriction of commercial transactions.

Ellis Brokerage Company, Inc. (Ellis Brokerage) was a North Carolina food broker that served as a middleman between large-quantity food buyers, such as hospitals and school systems, and food producers.¹² The defendant, Northern Star Company (Northern Star), was a potato processor based in Minnesota and a client of Ellis Brokerage since 1981.¹³ Between 1981 and 1986 Ellis Brokerage increased Northern Star's annual sales in eastern North Carolina from zero to approximately \$640,000.¹⁴

On June 20, 1986, Earl Ellis, the only full-time employee of Ellis Brokerage, received Northern Star potato pricing information over the telephone from Thomas Kenney, Northern Star's senior vice president for sales.¹⁵ On June 23 Ellis sent price lists based on the information to several potential buyers.¹⁶ By a letter dated August 29, 1986, Kenney notified Ellis that Northern Star was terminating its brokerage arrangement with Ellis Brokerage.¹⁷ Finally, on September 5 Kenney sent a letter to several of the buyers who received the June 23 price list from Ellis Brokerage.¹⁸ The letter contained the following language:

We have recently received copies of a price list sent to you from Ellis Brokerage Company regarding pricing on Northern Star potato products. These prices were noted for *bids only*, delivered by Northern Star.

We at Northern Star Company did not authorize such a price list and therefore cannot honor the prices as quoted on June 23, 1986.¹⁹

On the basis of this letter, Earl Ellis and Ellis Brokerage sued Northern Star and Thomas Kenney individually.²⁰

The plaintiffs' original action asserted that the letter of September 5, 1986, was "libelous *per se* and an unfair or deceptive act affecting commerce" in viola-

10. *Id.* at 225-26, 388 S.E.2d at 131.

11. *Id.* at 228-31, 388 S.E.2d at 132-34 (Meyer, J., dissenting).

12. *Id.* at 221, 388 S.E.2d at 128.

13. *Id.* at 221, 388 S.E.2d at 128-29.

14. *Id.* at 221-22, 388 S.E.2d at 129. Ellis Brokerage received a 3% commission for orders brokered on behalf of Northern Star. Defendant-Appellants' Brief at 4, *Ellis* (No. 192PA89). Commissions from Northern Star constituted "at least 80% of [the] commissions" of Ellis Brokerage. *Id.*

15. *Ellis*, 326 N.C. at 221-22, 388 S.E.2d at 128-29.

16. *Id.* at 222, 388 S.E.2d at 129.

17. *Id.*; Plaintiffs' Exhibit 2, *Ellis* (No. 192PA89).

18. *Ellis*, 326 N.C. at 222, 388 S.E.2d at 129.

19. *Id.*; Plaintiffs' Exhibit 14, *Ellis* (No. 192PA89).

20. *Ellis*, 326 N.C. at 222, 388 S.E.2d at 129.

tion of section 75-1.1.²¹ The plaintiffs amended their complaint to include allegations of "breach of a covenant of good faith, breach of contract through unreasonable termination, tortious interference with business relations, and unjust enrichment The defendants counterclaimed for breach of fiduciary duty and breach of contract."²² Earl Ellis's testimony at trial included a discussion he had with a customer of Ellis Brokerage who received the Northern Star letter, stating that the customer "was going to look for other sources to get his potatoes because he didn't know whether he could trust me or Northern Star."²³ The jury found that the defendants had libeled Ellis Brokerage and awarded \$32,500 in compensatory damages and \$12,500 in punitive damages.²⁴ The trial court, however, granted the defendants' motions for directed verdicts on all other claims, including the defendants' alleged violation of section 75-1.1.²⁵

Both plaintiffs and defendants appealed the decision of the trial court and the North Carolina Supreme Court considered two issues: whether the Northern Star letter of September 5, 1986, was libelous *per se*, as found by the jury, and whether libel *per se* of a plaintiff relating to its business "constitutes an unfair or deceptive act affecting commerce in violation of" section 75-1.1.²⁶ The supreme court affirmed the trial court's finding that the letter was libelous *per se*, rejecting the defendants' argument that the letter was "not defamatory at all or, alternatively . . . susceptible of both defamatory and nondefamatory interpretations."²⁷ Justice Mitchell, writing for the court, stated that "[w]hether a publication is one of the type that properly may be deemed libelous *per se* is a question of law to be decided initially by the trial court," and he asserted that the trial court properly decided that issue.²⁸ After that step, the jury must decide if the publication was actually libelous *per se*, and the supreme court agreed with the jury's affirmative resolution of that issue in this case.²⁹ The court, however, was not unanimous on the issue of libel *per se*. In a dissenting opinion, Justice Meyer, joined by Justice Whichard, stated that the "defendant's letter was clearly *not* defamatory *per se*, and the issue should *not* have been submitted to the jury."³⁰

The court then considered an issue of first impression: whether libel *per se* in a business setting is an unfair or deceptive trade practice in violation of section 75-1.1.³¹ In holding that libel *per se* of this type is within the realm of section 75-1.1, the court analogized libel *per se* in a business setting to false

21. *Id.*

22. *Id.* The parties settled the breach of contract claim and counterclaim prior to trial. *Id.*

23. *Id.* at 224, 388 S.E.2d at 130.

24. *Id.* at 222, 227, 388 S.E.2d at 129, 132. The jury found that "the defendants had not libeled the individual plaintiff Earl Ellis." *Id.* at 222, 388 S.E.2d at 129.

25. *Id.* at 222, 388 S.E.2d at 129.

26. *Id.* at 221, 388 S.E.2d at 128.

27. *Id.* at 224, 388 S.E.2d at 130. For "words to be libelous *per se* [they] must be susceptible of but one meaning." *Flake v. Greensboro News Co.*, 212 N.C. 780, 786, 195 S.E. 55, 60 (1938).

28. *Ellis*, 326 N.C. at 224, 388 S.E.2d at 130; *see Flake*, 212 N.C. at 785, 195 S.E. at 59.

29. *Ellis*, 326 N.C. at 224-25, 388 S.E.2d at 130-31.

30. *Id.* at 230, 388 S.E.2d at 134 (Meyer, J., dissenting).

31. *Id.* at 225-26, 388 S.E.2d at 131.

advertising and fraud, both of which the court previously had found to violate the statute.³² Because libel per se in a business setting did not fall within the category of "transactions already subject to pervasive and intricate statutory regulation,"³³ the court summarily deemed it to be an act in violation of section 75-1.1, justifying an award of treble damages under section 75-16.³⁴ The court applied this decision to the facts at hand and decided that the trial court erred in granting the defendants' motion for directed verdicts on the unfair trade practice claim.³⁵ Because the defendants' act violated section 75-1.1, and since the jury found that Ellis Brokerage suffered actual damages to its business reputation caused by the libel, the court directed the jury on remand to award Ellis Brokerage a choice of damages: \$32,500 in actual damages automatically trebled to a sum of \$97,500 under section 75-16, or the previously calculated libel award of \$45,000.³⁶

As a backdrop to analysis of the *Ellis* decision, North Carolina law covering trade libel, defamation, and unfair trade practices will be discussed. The doctrine of trade libel, or disparagement, provides a common-law cause of action for injured plaintiffs engaged in business in North Carolina, as well as a number of other jurisdictions.³⁷ One definition of the cause of action is as follows:

[D]isparagement . . . may consist of the publication of matter derogatory to the plaintiff's title to his property, or its quality, or to his business in general, or even to some element of his personal affairs, of a kind calculated to prevent others from dealing with him, or otherwise to interfere with his relations with others to his disadvantage.³⁸

The development of the trade libel cause of action in North Carolina reached its peak in the 1942 case of *Carolina Aniline & Extract Co. v. Ray*.³⁹ In *Carolina Aniline*, the plaintiff sold almost its entire production operation to the defendant, then purchased new equipment and continued its business at another location.⁴⁰ The defendant wrote the plaintiff's customers letters that explained the purchase, stated that it would "manufacture identically the same products under [its] own trade names," and directly compared its lower prices with those of the plaintiff.⁴¹ The plaintiff alleged that the letters were "false, deceptive and

32. *Id.* at 225, 388 S.E.2d at 131; see *infra* notes 98-106 and accompanying text.

33. *Ellis*, 326 N.C. at 225, 388 S.E.2d at 131. The court previously applied this limitation to securities transactions in *Skinner v. E.F. Hutton & Co.*, 314 N.C. 267, 275, 333 S.E.2d 236, 241 (1985); see *infra* notes 74-76 and accompanying text.

34. *Ellis*, 326 N.C. at 226, 388 S.E.2d at 131.

35. *Id.*

36. *Id.* at 227-28, 388 S.E.2d at 132; see *supra* note 24 and accompanying text. Plaintiffs may elect to recover either punitive damages under a common-law claim or treble damages under § 75-16, but not both. See *Marshall v. Miller*, 47 N.C. App. 530, 542, 268 S.E.2d 97, 103 (1980), *modified and aff'd*, 302 N.C. 539, 276 S.E.2d 397 (1981).

37. See *infra* notes 121-24 and accompanying text. Other names for the tort of trade libel include "injurious falsehood," "commercial disparagement," "disparagement of property," and "slander of goods." W. KEETON, D. DOBBS, R. KEETON & D. OWEN, PROSSER AND KEETON ON THE LAW OF TORTS § 128, at 962-63 (5th ed. 1984) [hereinafter LAW OF TORTS].

38. LAW OF TORTS, *supra* note 37, § 128, at 967.

39. 221 N.C. 269, 20 S.E.2d 59 (1942).

40. *Id.* at 270, 20 S.E.2d at 60.

41. *Id.* at 270-71, 20 S.E.2d at 60.

intended to deceive and did deceive plaintiff's customers."⁴² In reversing the trial court's grant of nonsuit for the defendant, the court expressed the rationale that unfair competition occurs whenever "the public is likely to be deceived."⁴³ The court clarified very little about the trade libel cause of action, however, and did not even label the action by any of the traditional phrases such as "trade libel", "disparagement", or "injurious falsehood".⁴⁴ Since *Carolina Aniline*, the cause of action for disparagement has not developed and injured plaintiffs in similar circumstances instead have pursued a cause of action for defamation.⁴⁵

Defamation is that which tends "to diminish the esteem, respect, goodwill or confidence in which the plaintiff is held, or to excite adverse, derogatory or unpleasant feelings or opinions against him."⁴⁶ North Carolina courts recognize three categories of defamatory material:

- (1) Publications which are obviously defamatory and which are termed libels per se;
- (2) publications which are susceptible of two reasonable interpretations, one of which is defamatory and the other is not;
- and (3) publications which are not obviously defamatory, but which become so when considered in connection with innuendo, colloquium and explanatory circumstances.⁴⁷

Furthermore,

a publication is libelous per se, or actionable *per se*, if, when considered alone without innuendo: (1) It charges that a person has committed an infamous crime; (2) it charges a person with having an infectious disease; (3) it tends to subject one to ridicule, contempt, or disgrace; or (4) it tends to impeach one in his trade or profession.⁴⁸

The courts have interpreted the fourth category of libel per se, applied in *Ellis*, in a number of cases. In *Badame v. Lampke*,⁴⁹ the plaintiff alleged that the defendant, a direct competitor, spoke words over the telephone to a customer that implied that the plaintiff engaged in "shady deals," and thereby impaired the plaintiff's business reputation.⁵⁰ Holding that the words spoken by the defendant were actionable per se for "charg[ing] the plaintiff with a dishonorable

42. *Id.* at 270, 20 S.E.2d at 60.

43. *Id.* at 272, 20 S.E.2d at 61 (quoting 63 C.J. *Trade-Marks, Trade-Names, and Unfair Competition* § 112, at 415 (1933)).

44. One point of clarification made by the court was the following:

[I]n order to escape liability for unfair competition, statements made for the purpose of inducing a competitor's customers to purchase the advertiser's products by making the express statement that his products possess all the qualities of the products of another . . . must be true, or the injured party will be entitled to relief.

Id.

45. See Comment, *Unfair Competition—Law of Unfair Competition in North Carolina*, 46 N.C.L. REV. 856, 877-78 (1968). For a discussion of *Carolina Aniline* and the status of the disparagement cause of action in North Carolina, see *id.* at 874-78.

46. LAW OF TORTS, *supra* note 37, § 111, at 773.

47. *Flake v. Greensboro News Co.*, 212 N.C. 780, 785, 195 S.E. 55, 59 (1938).

48. *Id.* at 787, 195 S.E. at 60.

49. 242 N.C. 755, 89 S.E.2d 466 (1955).

50. *Id.* at 755-56, 89 S.E.2d at 467.

course of business conduct,"⁵¹ the supreme court noted that the words "must contain an imputation necessarily hurtful in its effect on [the plaintiff's] business."⁵² More recently, in *Matthews, Cremins, McLean, Inc. v. Nichter*,⁵³ the court of appeals held that letters sent by defendant media buying service to television stations, which asserted that the plaintiff advertising agency breached its contract and failed to pay its bills, were libelous per se.⁵⁴ The court's rationale was that the communication of such assertions to third parties "clearly tend[ed] to disparage plaintiff's integrity in its business dealings."⁵⁵

While the trade libel and defamation causes of action exist within the common law of North Carolina, the law of unfair trade practices flourishes within the confines of North Carolina General Statutes section 75-1.1. In 1977 the general assembly amended the original language of section 75-1.1⁵⁶ in order to conform to the exact wording of its federal counterpart, section 5 of the Federal Trade Commission Act.⁵⁷ Rather than enumerate a list of specific illegal acts, practices, and methods of competition, the general assembly chose to follow Congress's definition in "adopt[ing] a phrase which . . . does not 'admit of precise definition, but the meaning and application of which must be arrived at by what [the Supreme Court] elsewhere has called 'the gradual process of judicial inclusion and exclusion.' " "⁵⁸

The North Carolina Supreme Court examined section 75-1.1 for the first time in *Hardy v. Toler*.⁵⁹ In that case, the plaintiff alleged that the defendant automobile dealer made false representations concerning the condition of a car at the time of purchase.⁶⁰ In holding that as a matter of law the false representations made by defendant to plaintiff violated section 75-1.1,⁶¹ the court noted that "[s]ome guidance may be obtained by reference to federal decisions on appeals from the Federal Trade Commission since the language of [section] 75-1.1 closely parallels that of the Federal Trade Commission Act."⁶² The court also

51. *Id.* at 757, 89 S.E.2d at 468.

52. *Id.*

53. 42 N.C. App. 184, 256 S.E.2d 261, *disc. rev. denied*, 298 N.C. 569, 261 S.E.2d 123 (1979).

54. *Id.* at 188, 256 S.E.2d at 264.

55. *Id.* For other North Carolina cases dealing with the business impeachment category of libel per se, see, e.g., *Lay v. Gazette Publishing Co.*, 209 N.C. 134, 183 S.E. 416 (1936) (libel per se found when a newspaper incorrectly published that the plaintiff was the leader of a strike and had been arrested for trespassing); *Broadway v. Cope*, 208 N.C. 85, 179 S.E. 452 (1935) (statement by butcher that his competitor had slaughtered a cow bitten by a mad dog was defamatory per se); *U v. Duke Univ.*, 91 N.C. App. 171, 371 S.E.2d 701 (statements by defendant to plaintiff's colleague that plaintiff was a liar, deceitful, absolutely useless, and a fraud impeached plaintiff in his profession and were slanderous per se), *disc. rev. denied*, 323 N.C. 629, 374 S.E.2d 590 (1988); *Talbert v. Mauney*, 80 N.C. App. 477, 343 S.E.2d 5 (1986) (allegations that president of bank published statements imputing that plaintiff forged his letters of credit and was a drug dealer alleged slander per se).

56. Act of June 27, 1977, ch. 747, §§ 1-2, 1977 N.C. Sess. Laws 984 (codified as amended at N.C. GEN. STAT. § 75-1.1 (1988)). For the amended text of § 75-1.1(a), see *supra* note 1.

57. 15 U.S.C. § 45(a)(1) (1988).

58. *Federal Trade Comm'n v. R.F. Keppel & Bros.*, 291 U.S. 304, 312 (1934) (quoting *Federal Trade Comm'n v. Raladam Co.*, 283 U.S. 643, 648 (1931)).

59. 288 N.C. 303, 218 S.E.2d 342 (1975).

60. *Id.* at 304, 218 S.E.2d at 343.

61. *Id.* at 311, 218 S.E.2d at 347.

62. *Id.* at 308, 218 S.E.2d at 345.

established the bifurcated process used in cases under section 75-1.1, whereby "the jury . . . determine[s] the facts, and based on the jury's finding, the court . . . determine[s] as a matter of law whether the defendant engaged in unfair or deceptive acts or practices in the conduct of trade or commerce."⁶³

In *Johnson v. Phoenix Mutual Life Insurance Co.*,⁶⁴ plaintiff real estate developers alleged that the defendant mortgage broker and mortgagor "entered into a deliberate course of conduct which was designed to force [the plaintiff] into an untenable economic position so that it would be unable to complete" the development of a shopping center.⁶⁵ The court provided some guidance on the scope of section 75-1.1 by stating that "[w]hat is an unfair or deceptive trade practice usually depends upon the facts of each case and the impact the practice has in the marketplace."⁶⁶ Furthermore, a practice is unfair when it "offends established public policy as well as when the practice is immoral, unethical, oppressive, unscrupulous, or substantially injurious to consumers,"⁶⁷ or when it "amounts to an inequitable assertion of [a party's] power or position."⁶⁸ The court applied these guidelines to the defendants' conduct and concluded that it was not unfair or deceptive.⁶⁹

The court made two significant points in *Marshall v. Miller*,⁷⁰ an action by a consumer for misrepresentation. First, the court held that the intent or good faith belief of a party is irrelevant to the determination of the unfairness or deceptiveness of a particular act.⁷¹ Second, the court justified automatic treble damages. While discussing section 75-16 and treble damages, the court concluded that "[a]bsent statutory language making trebling discretionary with the trial judge . . . the Legislature intended trebling of any damages assessed to be automatic once a violation [of section 75-1.1] is shown."⁷² The rationale offered for the automatic trebling scheme under section 75-16 was that "it makes more economically feasible the bringing of an action where the possible money damages are limited, and thus encourages private enforcement" and "it increases the incentive for reaching a settlement."⁷³

The court limited the application of section 75-1.1 in *Skinner v. E.F. Hutton & Co.*,⁷⁴ when it held that because securities transactions "were already subject

63. *Id.* at 310, 218 S.E.2d at 346-47.

64. 300 N.C. 247, 266 S.E.2d 610 (1980).

65. *Id.* at 251, 266 S.E.2d at 614. The defendants' course of conduct included the tendering of a mortgage for the shopping center, which was conditioned on the plaintiffs' ability to secure an interim construction loan as well as particular lease commitments. *Id.* at 250, 266 S.E.2d at 613. When the plaintiffs failed to obtain the construction loan or a lease commitment from a bank, the defendant mortgagor terminated the mortgage. *Id.* at 251, 266 S.E.2d at 614.

66. *Id.* at 262-63, 266 S.E.2d at 621.

67. *Id.* at 263, 266 S.E.2d at 621. The court also noted that "[t]he concept of 'unfairness' is broader than and includes the concept of 'deception.'" *Id.*

68. *Id.* at 264, 266 S.E.2d at 622.

69. *Id.* at 266, 266 S.E.2d at 623.

70. 302 N.C. 539, 276 S.E.2d 397 (1981).

71. *Id.* at 548-49, 276 S.E.2d at 403.

72. *Id.* at 547, 276 S.E.2d at 402.

73. *Id.* at 549, 276 S.E.2d at 403-04.

74. 314 N.C. 267, 333 S.E.2d 236 (1985).

to pervasive and intricate regulation under the North Carolina Securities Act, as well as the Securities Act of 1933, and the Securities Exchange Act of 1934,' " they were beyond the scope of the statute.⁷⁵ The judicial exclusion of transactions subject to statutory regulation apart from section 75-1.1 reflected the court's concern for over-penalizing a party already subject to sanctions.⁷⁶

The court attempted to interpret "commerce," as used in section 75-1.1(b), in *Olivetti Corp. v. Ames Business Systems, Inc.*⁷⁷ For legislative intent, the court looked to the original version of section 75-1.1(b), which applied to "dealings between persons engaged in business, and between persons engaged in business and the consuming public."⁷⁸ The court noted that "individual consumers are not the only ones protected" by the language of the statute,⁷⁹ but found that "section [75-1.1(b)] is not broad enough . . . to encompass 'all forms of business activities.' "⁸⁰

The development of North Carolina law on libel and unfair trade practices converged in the *Ellis* decision. The first question answered by the court in *Ellis* concerned the libelous nature of the letter mailed by defendant Northern Star.⁸¹ Arguably, plaintiff Ellis Brokerage could have brought a common-law action for trade libel or disparagement. After all, the plaintiff's libel per se claim, which was premised on injury to its business relationships with customers allegedly caused by Northern Star's letter, appears to fit within the definition of disparagement.⁸² The failure of the plaintiff to raise this cause of action in an appropriate fact scenario testifies to the lifeless state of the disparagement cause of action in North Carolina,⁸³ despite its viability in other jurisdictions.⁸⁴

75. *Id.* at 275, 333 S.E.2d at 241 (quoting *Lindner v. Durham Hosiery Mills, Inc.*, 761 F.2d 162, 167 (4th Cir. 1985) (citations omitted)); see also *Bache Halsey Stuart, Inc. v. Hunsucker*, 38 N.C. App. 414, 420-21, 248 S.E.2d 567, 570 (1978) (commodities transactions not within the scope of § 75-1.1 due to the existence of a federal scheme of regulation), *disc. rev. denied*, 296 N.C. 583, 254 S.E.2d 32 (1979).

76. *Skinner*, 314 N.C. at 274, 333 S.E.2d at 241 ("[A]pplication of the statute . . . would expose a party violating the statute to a host of legislatively created sanctions in addition to those sought in the private action.").

77. 81 N.C. App. 1, 22-23, 344 S.E.2d 82, 94-95 (1986), *aff'd in part and rev'd in part on other grounds*, 319 N.C. 534, 356 S.E.2d 578 (1987). Section 75-1.1(b) provides: "For purposes of this section, 'commerce' includes all business activities, however denominated, but does not include professional services rendered by a member of a learned profession." N.C. GEN. STAT. § 75-1.1(b) (1988).

78. N.C. GEN. STAT. § 75-1.1(b) (1969). The original version of § 75-1.1(b) provided:

The purpose of this section is to declare, and to provide civil legal means to maintain, ethical standards of dealings between persons engaged in business, and between persons engaged in business and the consuming public within this State, to the end that good faith and fair dealings between buyers and sellers at all levels of commerce be had in this State.

Id.

79. *Olivetti*, 81 N.C. App. at 23, 344 S.E.2d at 95.

80. *Id.* at 22, 344 S.E.2d at 94 (quoting *Threatt v. Hiers*, 76 N.C. App. 521, 523, 333 S.E.2d 772, 773 (1985), *disc. rev. denied*, 315 N.C. 397, 338 S.E.2d 887 (1986)).

81. See *supra* text accompanying notes 27-30.

82. See *supra* text accompanying note 38.

83. See *supra* note 45 and accompanying text; see also UNITED STATES TRADEMARK ASSOCIATION, STATE TRADEMARK AND UNFAIR COMPETITION LAW NC-14 (1991) [hereinafter UNFAIR COMPETITION] ("There are no reported decisions on [the] subject [of trade disparagement or trade libel] by North Carolina courts.").

84. See *infra* text accompanying note 124.

In the absence of a body of common law on disparagement on which to rely, Ellis Brokerage brought a successful action for libel per se. The supreme court provided the following explanation for its affirmation of the trial court's finding of libel per se:

The language "[w]e at Northern Star did not authorize such a price list," taken in the context of the entire letter, can only be read to mean that Ellis Brokerage Company, acting in its capacity as broker for Northern Star, did an unauthorized act. Whether that act was publishing certain unauthorized prices within a price list or publishing the entire price list itself without authorization is of no import; either reading is defamatory and impeaches Ellis Brokerage in its trade as a food broker.⁸⁵

It is not entirely clear why this language is defamatory and impeaching. As pointed out by Justice Meyer in his dissenting opinion, the language in question, "when read by a typical recipient of [the] letter, could very reasonably be interpreted to mean that there was a simple breakdown in communications or an inadvertent mistake in the price list through the fault of either or both parties."⁸⁶ If the accusation that a food broker acted without authority connotes a lack of scrupulous business practices or integrity in the eyes of its customers, and the customers have come to expect those qualities from the broker, the case for libel per se is clear. If the defendant communicated an accusation of repeated mistakes in price listings or habitual unauthorized listings on the part of the plaintiff, there is also a sound basis for a libel per se claim.⁸⁷

In this case, however, Northern Star's letter "does not rise to the level of accusing [Ellis Brokerage] of incompetence or untrustworthiness, nor would a typical buyer automatically reach that conclusion."⁸⁸ Whereas the findings of libel per se in *Badame*, *Matthews*, and other North Carolina cases dealing with defamation in a business context⁸⁹ were based on language "contain[ing] an imputation which [was] necessarily harmful in its effect on plaintiffs' business,"⁹⁰ Justice Meyer presented a strong argument that the court's finding in *Ellis* did not rest on so solid a foundation.⁹¹ Conversely, Justice Mitchell, writing for the majority, provided scant justification for the holding, thereby accepting a minimal standard for finding libel per se in a business setting. While this part of the court's holding is of only secondary importance, it is noteworthy for the shadow that it casts over business communications.

The substance of *Ellis* lies in the supreme court's treatment of the question of first impression whether libel per se is an unfair or deceptive trade practice in violation of section 75-1.1. The court emphasized that the existence of "perva-

85. *Ellis*, 326 N.C. at 224, 388 S.E.2d at 130.

86. *Id.* at 229, 388 S.E.2d at 133 (Meyer, J., dissenting).

87. *Id.* (Meyer, J., dissenting).

88. *Id.* (Meyer, J., dissenting).

89. See *supra* notes 49-55 and accompanying text.

90. *Ellis*, 326 N.C. at 229, 388 S.E.2d at 133 (Meyer, J., dissenting).

91. *Id.* (Meyer, J., dissenting) ("[O]ur courts have tended to recognize more blatantly derogatory statements than the one at issue here as defamatory *per se* in the business context."); see *supra* note 55.

sive and intricate statutory regulation" for a particular type of transaction would provide an exception to the protective shield of section 75-1.1, but that no such limitation applied to libel per se of a type impeaching a party in its business.⁹² While the court did not explain the exception, the probable rationale is that an injured plaintiff would have a sufficient avenue of recourse under the specifically tailored statutory regulation and would not need the protections of section 75-1.1.⁹³ Although libel per se is not addressed by statute in North Carolina, it is the subject of a well-developed common-law cause of action.⁹⁴ As in *Ellis*, a plaintiff who can show that an injury was proximately caused by a libelous statement likely will receive compensatory and punitive damages.⁹⁵ Common law, in this instance, thus provides as effective and sufficient a remedy for the injured party as statutory regulation provides in the securities arena.⁹⁶ A party suing for libel has the option of selecting higher damages under section 75-1.1 and the accompanying section 75-16, while a plaintiff injured under a statutorily regulated transaction is limited to actual and compensatory damages.⁹⁷ The common-law protection from libel is pervasive, like that of a statutory scheme, and logically the court should exclude libel per se from protection under section 75-1.1.

As further justification for the categorization of libel per se in a business setting as an unfair or deceptive trade practice in violation of section 75-1.1, the *Ellis* court pointed out that both false advertising and fraud had been found to violate the statute.⁹⁸ The court cited a fraud case, *Hardy v. Toler*,⁹⁹ in which a plaintiff consumer recovered under sections 75-1.1 and 75-16 for misrepresentations made by the defendant automobile dealer at the time of purchase of a car.¹⁰⁰ The court also cited a false advertising case, *Winston Realty Co. v. G.H.G., Inc.*,¹⁰¹ in which the plaintiff corporation hired the defendant personnel agency for the purpose of filling a bookkeeping position.¹⁰² The plaintiff filled the position with someone who ultimately embezzled \$24,000 and committed

92. *Ellis*, 326 N.C. at 225, 388 S.E.2d at 131.

93. Additionally, it would be inequitable to penalize a defendant already subject to statutory sanctions. The court employed such reasoning in cases involving securities and commodities transactions. See *Skinner v. E.F. Hutton & Co.*, 314 N.C. 267, 274, 333 S.E.2d 236, 241 (1985) (securities transactions); *Bache Halsey Stuart, Inc. v. Hunsucker*, 38 N.C. App. 414, 420-21, 248 S.E.2d 567, 570 (1978) (commodities transactions), *disc. rev. denied*, 296 N.C. 583, 254 S.E.2d 32 (1979).

94. See *supra* text accompanying notes 47-48.

95. *Ellis*, 326 N.C. at 225, 388 S.E.2d at 130-31.

96. This result occurs notwithstanding the legislative intent behind the enactment of § 75-1.1, which noted that "common law remedies had proved often ineffective." See, e.g., *Marshall v. Miller*, 302 N.C. 539, 543, 276 S.E.2d 397, 400 (1981).

97. See *supra* note 75 and accompanying text.

98. *Ellis*, 326 N.C. at 225, 388 S.E.2d at 131.

99. 288 N.C. 303, 218 S.E.2d 342 (1975); see *supra* text accompanying notes 59-63.

100. *Hardy*, 288 N.C. at 311, 218 S.E.2d at 347. The misrepresentations included the following: The automobile had 79,000 miles on it instead of the 21,000 registered on the odometer; the vehicle had been sold twice before, but was represented as having only one previous owner; and the purchaser was not told that the automobile had been damaged in a collision. *Id.* at 304, 218 S.E.2d at 343.

101. 314 N.C. 90, 331 S.E.2d 677 (1985).

102. *Id.* at 92, 331 S.E.2d at 678.

other financial improprieties.¹⁰³ The defendant advertised that it "pre-screened" job applicants, but failed to conduct a reference check or background investigation with regard to any criminal activity by the job applicant.¹⁰⁴ The court held that the plaintiff could recover for the defendant's false advertising under section 75-1.1.¹⁰⁵ In both *Hardy* and *Winston Realty*, the injured plaintiff was a consumer of the defendant's goods or services. This is a significant distinction from *Ellis*, in which plaintiff Ellis Brokerage was not a consumer of defendant Northern Star's products, but rather a conduit for its products. In the cited cases the plaintiffs' direct reliance on the acts of the defendants necessitated the broad protections of section 75-1.1 to provide the plaintiffs with recourse. Conversely, in *Ellis* the alleged libel did not cause Ellis Brokerage to rely on Northern Star, and Ellis did have an available common-law remedy. The court did not address these differences and failed to explain its reliance on the cited examples in reaching its conclusion.¹⁰⁶ The court's silence calls into question the soundness of its holding.

Another concern prompted by the court's inclusion of libel per se in a business context within section 75-1.1 is the danger that communications found to be libelous may vary widely as to the degree of harm and injury they cause. Justice Meyer reflected this concern by pointing out that the language found to be libelous in *Ellis* was much less derogatory than statements the North Carolina courts traditionally have found to be libelous in a business context.¹⁰⁷ Viewing the libel in *Ellis* as occurring near the low end of the libel spectrum, in terms of degree of imputed harm, the supreme court has paved the way for lower courts to treat many business communications that, prior to *Ellis*, flowed with impunity, as unfair acts under section 75-1.1. Communications between persons engaged in business are undoubtedly activities within the realm of "commerce" and thereby trigger analysis under section 75-1.1.¹⁰⁸ The court's prior interpretation of unfair acts in *Johnson* as those that are "immoral, unethical, oppressive, unscrupulous, or substantially injurious to consumers,"¹⁰⁹ however, provides little in the way of limitations. With the exception of "substantially injurious to consumers," each of the delineated characteristics is subjective, thereby providing minimal guidance. In the case of libel per se, an extremely defamatory statement in a business context very well could fall under one or more of those labels and have a devastating effect on a plaintiff. Such a strong libel, however, certainly would be actionable for under a libel suit, which would obviate the need for the protection of section 75-1.1. *Ellis* does not present such a strong case and it is very difficult to label the *Ellis* language, "[w]e at Northern

103. *Id.* at 93, 331 S.E.2d at 679.

104. *Id.*

105. *Id.* at 97, 331 S.E.2d at 681.

106. The court's entire commentary was: "We have concluded, for example, that both false advertising and fraud violate [§ 75-1.1]. . . . [L]ike fraud and false advertising, a libel *per se* of a type impeaching a party in its business activities is an unfair or deceptive act in or affecting commerce in violation of [§] 75-1.1." *Ellis*, 326 N.C. at 225-26, 388 S.E.2d at 131.

107. *See supra* note 91.

108. *See supra* notes 77-80 and accompanying text.

109. *Johnson v. Phoenix Mut. Life Ins. Co.*, 300 N.C. 247, 263, 266 S.E.2d 610, 621 (1980).

Star did not authorize such a price list," as satisfying the *Johnson* interpretation.¹¹⁰ The court failed to apply the useful test of "the impact the practice has in the marketplace,"¹¹¹ which appears to be minimal in the case of libel, due to the infrequency of the transgression. Thus, the *Ellis* holding expands the applicability of section 75-1.1, but provides no analytical framework to guide the lower courts in determining what constitutes a violation of the statute.¹¹²

The holding that libel per se in a business context violates section 75-1.1 is not significant simply because it classifies libel as an unfair or deceptive trade practice. Rather, it is the attachment of automatic treble damages provided by section 75-16 that provides the real bite.¹¹³ If the court finds a violation of section 75-1.1, it must treble the compensatory damages established by the jury.¹¹⁴ The legislative intent behind section 75-16 included a desire to increase the economic feasibility of "bringing . . . an action where the possible money damages are limited, and thus encourage[] private enforcement."¹¹⁵ In a libel case such as *Ellis*, monetary damages would be "limited" to compensatory and punitive damages—a sufficient common-law remedy. Providing the plaintiff with the choice between a common-law remedy and automatically trebled actual damages under section 75-16¹¹⁶ exceeds the legislative intent. The automatic trebling does not take into account the nature of the particular transaction; in *Ellis* for example, the existence of libel was strongly questioned.¹¹⁷ Judicial determination of libel is too uncertain a process to subject to automatic treble damages; the imposition of such damages could inequitably penalize a defendant.

A brief survey illustrates where the North Carolina unfair or deceptive trade practices act stands in relation to the schemes of other states.¹¹⁸ The statutes of thirty-two states proscribe unfair or deceptive acts or practices.¹¹⁹ Thirty-four states list certain acts or practices as unlawful; North Carolina does not.¹²⁰ Statutory lists serve the important function of providing predictability as to potential liability for unfair trade practices. North Carolina is one of twenty

110. *Ellis*, 326 N.C. at 222, 388 S.E.2d at 129; see *Johnson*, 300 N.C. at 263, 266 S.E.2d at 621.

111. *Johnson*, 300 N.C. at 263, 266 S.E.2d at 621.

112. For a list of the types of activities to which § 75-1.1 has been applied, see Note, Olivetti Corp. v. Ames Business Systems, Inc.: *Recovery of Lost Profits for a Violation of North Carolina General Statutes Section 75-1.1*, 65 N.C.L. REV. 1169, 1178-79 (1987).

113. For the text of § 75-16, see *supra* note 3.

114. See *Marshall v. Miller*, 302 N.C. 539, 547, 276 S.E.2d 397, 402 (1981); N.C. GEN. STAT. § 75-16 (1988).

115. *Marshall*, 302 N.C. at 549, 276 S.E.2d at 403-04.

116. A plaintiff may elect either a common-law remedy including punitive damages or trebled actual damages under § 75-16, but not both. See *supra* note 36 and accompanying text. For an examination of the availability of punitive or treble damages, see Note, *Unfair Trade Practices and Unfair Methods of Competition in North Carolina: Are Both Treble and Punitive Damages Available for Violations of Section 75-1.1?*, 62 N.C.L. REV. 1139 (1984).

117. *Ellis*, 326 N.C. at 228-31, 388 S.E.2d at 132-34 (Meyer, J., dissenting); see *supra* notes 86-91 and accompanying text.

118. For purposes of this summary, the District of Columbia is treated as a state.

119. See Leaffer & Lipson, *Consumer Actions Against Unfair or Deceptive Acts or Practices: The Private Uses of Federal Trade Commission Jurisprudence*, 48 GEO. WASH. L. REV. 521, 531, 560-64 app. (1980).

120. See *id.* Some of the state statutes that list unlawful practices also include a catch-all phrase designed to include nonlisted practices. *Id.* at 531.

states that make available double or treble plaintiff's actual damages in private actions for injuries resulting from unfair or deceptive trade practices.¹²¹ Among the states that provide treble damages, only five, including North Carolina, provide mandatory treble damages.¹²² Eighteen states have statutory provisions specifically prohibiting disparagement or trade libel,¹²³ while fourteen states recognize a common-law disparagement cause of action.¹²⁴ In ten states, including North Carolina, a common-law libel cause of action allows plaintiffs to recover for disparagement-type injuries, while in several other states, libel statutes provide for recovery.¹²⁵ Overall, North Carolina's unfair or deceptive trade practices act, with its automatic treble damages and lack of statutory list, is broader and more generous than the parallel acts of many other states; its common-law remedy for disparagement or trade libel is on par with about half of the states.

Finally, application of the disparagement and unfair trade practice schemes of three leading commercial states—California, New York, and Texas—to the facts of *Ellis* yields significantly different results from those of the North Carolina approach. In California there is an established common-law cause of action for trade libel.¹²⁶ The cause of action requires a publication that induces others not to deal with the plaintiff, as well as a showing of special damages.¹²⁷ Applying the facts of *Ellis*, the plaintiff could successfully bring such a cause of action if it could prove that the letter from Northern Star injured its reputation and thereby caused customers to terminate their business relationships. Under this cause of action, Ellis Brokerage could recover specific lost sales and punitive damages upon proof of actual damages.¹²⁸ This result parallels the one under

121. See UNFAIR COMPETITION, *supra* note 83 (includes "Trade Disparagement or Trade Libel" subsections for each state).

122. The other four are California, Colorado, Hawaii, and New Jersey. See CAL. BUS. & PROF. CODE § 17082 (West 1987); COLO. REV. STAT. § 6-1-113(2)(a) (Supp. 1990) (treble damages or \$250, whichever is greater); HAW. REV. STAT. § 480-13(a)(1) (Supp. 1990) (treble damages or \$1000, whichever is greater); N.J. STAT. ANN. § 56:8-19 (West 1989). In states with treble damages that are not mandatory, such damages may be awarded at the discretion of the court, or for violations that are "knowing or intentional." Leaffer & Lipson, *supra* note 119, at 532.

123. See, e.g., ALASKA STAT. § 45.50.471(b)(7) (1986); ARK. STAT. ANN. § 4-88-106(a)(2) (1987); COLO. REV. STAT. § 6-1-105(1)(h) (Supp. 1990); DEL. CODE ANN. tit. 6, § 2532(a)(8) (1974); D.C. CODE ANN. § 28-3904(g) (1981); GA. CODE ANN. §§ 10-1-372(a)(8), -393(b)(8) (1989 & Supp. 1990); HAW. REV. STAT. § 481A-3(a)(8) (1985); IDAHO CODE § 48-603(8) (Supp. 1990); ILL. ANN. STAT. ch. 121½, para. 312, § 2(8) (Smith-Hurd Supp. 1990); ME. REV. STAT. ANN. tit. 10, § 1212(1)(H) (1964); MINN. STAT. ANN. § 325D.44(1)(8) (West Supp. 1991); NEB. REV. STAT. § 87-302(a)(8) (Supp. 1990); NEV. REV. STAT. § 598.410(8) (1985); N.M. STAT. ANN. § 57-12-2(D)(8) (Supp. 1990); OHIO REV. CODE ANN. § 4165.02(H) (Anderson 1991); OKLA. STAT. ANN. tit. 78, § 53(a)(8) (West 1987); OR. REV. STAT. § 646.608(1)(h) (1989); TEX. BUS. & COM. CODE ANN. § 17.46(b)(8) (Vernon 1987). A typical statute contains language that prohibits "disparaging the goods, services, or business of another by false or misleading representation of fact." IDAHO CODE § 48-603(8) (Supp. 1990). An example of a more explicit statute, which leaves less room for misinterpretation, is: "It shall be a violation of this chapter, whether or not any consumer is in fact misled, deceived or damaged thereby, for any person to . . . disparage the goods, services or business of another by false or misleading representations of material facts." D.C. CODE ANN. § 28-3904(g) (1981).

124. See UNFAIR COMPETITION, *supra* note 83.

125. *Id.*

126. *Id.* at CA-18.1.

127. *Id.* at CA-19.

128. *Id.*

the North Carolina common-law libel cause of action.¹²⁹ Application of California's unfair and deceptive trade practices act provides different results, however. The act contains a broad statement of scope,¹³⁰ like North Carolina, and it also provides mandatory treble damages.¹³¹ The important distinction, however, is that under California's law only consumers and direct competitors may bring actions.¹³² Thus, a California plaintiff in the situation of *Ellis* Brokerage would be entitled to equitable common-law damages but not statutory treble damages, unlike the result in North Carolina.

New York has a common-law cause of action for disparagement that requires a showing of malice, falsity of the statement, and special damages.¹³³ Since there is no indication that the statement made by Northern Star was false, it is unlikely that *Ellis* Brokerage would succeed under New York's disparagement cause of action. Like North Carolina, New York also has a broadly worded deceptive trade practices act that provides a private cause of action for both consumers and other persons injured by violations.¹³⁴ This consumer-oriented act, however, limits remedies to the greater of actual damages or fifty dollars, and places a cap on discretionary treble damages of \$1000.¹³⁵ Under the facts of *Ellis*, a best-case scenario for the plaintiff in New York would be recovery of \$32,500 in actual damages plus \$1000 in discretionary treble damages—well below the \$97,500 provided by North Carolina law.¹³⁶

Texas recognizes a common-law cause of action for disparagement that protects "the economic interests of the injured party against pecuniary loss."¹³⁷ Disparagement requires "publication . . . , falsity, malice, lack of privilege, and special damages."¹³⁸ The *Ellis* facts likely would not satisfy the element of falsity and therefore would not lead to recovery of damages for lost profits, which are available under the cause of action.¹³⁹

129. See *supra* text accompanying note 24.

130. CAL. BUS. & PROF. CODE § 17001 (West 1987). The statute provides in part that the purpose of the act is to "foster and encourage competition, by prohibiting unfair, dishonest, deceptive, destructive, fraudulent and discriminatory practices by which fair and honest competition is destroyed or prevented." *Id.*

131. *Id.* § 17082.

132. *Id.* §§ 17021, 17070; see *Harris v. Capitol Records Distrib. Corp.*, 64 Cal. 2d 454, 460-61, 413 P.2d 139, 143-44, 50 Cal. Rptr. 539, 543-44 (1966) (requiring that a competitor be in "primary" or direct competition with wrongdoer in order to bring an action under California's Unfair Practices Act).

133. UNFAIR COMPETITION, *supra* note 83, at NY-23.

134. N.Y. GEN. BUS. LAW § 349(a) (McKinney 1988).

135. Discretionary treble damages will only be awarded if a violation of the statute is willful or knowing. UNFAIR COMPETITION, *supra* note 83, at NY-8.

136. The best-case scenario is unlikely, as the courts have interpreted the scope of New York's deceptive practices act to exclude "private, non-consumer transactions of a non-recurring type without implications for the public." *Id.* at NY-7; see *Genesco Entertainment v. Koch*, 593 F. Supp. 743, 752 (S.D.N.Y. 1984). Arguably, a transaction like that of *Ellis* would not meet the limited interpretation of the scope of the act, since it was between business entities (food supplier, broker, and institutional buyers), and was a one-time communication. See *supra* text accompanying notes 12-19.

137. *Hurlbut v. Gulf Atl. Life Ins. Co.*, 749 S.W.2d 762, 766 (Tex. 1987). There is a separate cause of action for defamation, which is intended "to protect the personal reputation of the injured party." *Id.*

138. *Id.*

139. UNFAIR COMPETITION, *supra* note 83, at TX-38.1 to TX-38.2.

The Texas deceptive trade practices act includes a broad prohibition against any "[f]alse, misleading, or deceptive acts or practices in the conduct of any trade or commerce," as well as a list of twenty-four acts considered per se false, misleading, or deceptive.¹⁴⁰ Included among the enumerated acts is "disparaging the goods, services, or business of another by false or misleading representation of facts."¹⁴¹ The cause of action provided by the statute is available to "consumers," with "consumer" defined as "an individual, partnership, corporation . . . who seeks or acquires by purchase or lease, any goods or services."¹⁴² Remedies for an injured consumer include actual damages and discretionary treble damages "[i]f the trier of fact finds that the conduct of the defendant was committed knowingly."¹⁴³ Thus, unlike North Carolina's scheme, the Texas statutory scheme is explicit as to its inclusion of disparagement as a violation and does not mandate automatic treble damages. A plaintiff such as Ellis Brokerage could recover treble damages nearly equivalent to those provided by the North Carolina statute, but only if the defendant's conduct was intentional. The application of the *Ellis* scenario to appropriate law in California, New York, and Texas illustrates that North Carolina's statutory scheme is out of step, due to its combination of broadness and unconditional, unlimited treble damages.

When the general assembly enacted North Carolina General Statutes sections 75-1.1 and 75-16, it intended to "provide civil means to maintain ethical standards of dealings between persons engaged in business and the consuming public . . . and to enable a person injured by [unfair or] deceptive acts or practices to recover . . . damages from a wrongdoer."¹⁴⁴ The legislature created a repository into which the courts could selectively place practices deemed to be unfair or deceptive through the continuous process of judicial inclusion and exclusion. The legislature also provided a substantial remedy for injured parties in the form of treble damages. Standing apart from all but a handful of states,¹⁴⁵ the legislature implemented, and the courts have enforced, automatic treble damages. While this protection provides sufficient recourse for injured parties, and actually may improve the ethics of the marketplace, there is the potential for the protection to become too restrictive. The decision of the supreme court in *Ellis* is a step in that direction. With brevity and minimal justification, and without any analysis of marketplace impact, the court added libel per se in a business context to the stockpile of unfair acts. In so doing, the court displayed an eagerness to expand section 75-1.1 to include a practice that already had sufficient

140. TEX. BUS. & COM. CODE ANN. § 17.46 (Vernon 1987).

141. *Id.* § 17.46(b)(8).

142. *Id.* § 17.45(4). The term "consumer" excludes a "business consumer" with assets of \$25 million or more, or under the ownership or control of a corporation or entity with assets of \$25 million or more. *Id.*

143. *Id.* § 17.50(b)(1) (Vernon Supp. 1991). The statute defines the discretionary treble damages as "[no] more than three times the amount of actual damages in excess of \$1000." *Id.*

144. *Hardy v. Toler*, 24 N.C. App. 625, 630-31, 211 S.E.2d 809, 813, *modified on other grounds*, 288 N.C. 303, 218 S.E.2d 342 (1975). For the statement of this intent in the original version of § 75-1.1(b), see *supra* note 78.

145. See *supra* note 122 and accompanying text.

remedies under common law.¹⁴⁶ This willingness to summarily expand protections in commerce could ultimately have a chilling effect on business transactions and the marketplace in general. North Carolina should consider the protective, yet more moderate, unfair trade practice schemes of other states that are leaders in commerce.¹⁴⁷

One option for improving the effectiveness of section 75-1.1 would be for the North Carolina General Assembly to clarify the statute by expressly including or excluding of particular practices. By amending the statute in that way, the legislature would provide warning to potential transgressors as well as concrete guidance to the courts. Alternatively, the legislature might consider implementing discretionary treble damages instead of the automatic treble damages of section 75-16. Discretionary trebling would eliminate the harshness of uniformly applying treble damages to all unfair trade practices, regardless of the nature of a particular act, while retaining the penalty for egregious acts.

In the absence of any legislative action, the onus will fall on the North Carolina judiciary to effectively define the parameters of the statute. The courts should view each decision involving the statute as an opportunity to clarify the analytical framework for determining what constitutes an unfair or deceptive trade practice. The courts should include in this framework an analysis of the impact the alleged violation has upon the marketplace. This analysis should consider the availability of remedies through other causes of action, as well as the derivative effects of automatic treble damages; the *Ellis* court ignored these considerations. It is imperative that the courts carry out the process of judicial inclusion and exclusion with a high degree of care and strike a balance between over-inclusion and over-exclusion. Failure to do so will have a detrimental impact on the commercial climate in North Carolina.

PAUL M. SARACENI

146. The expansion occurred in a case in which the court's standard for libel was particularly lenient. See *supra* notes 88-91 and accompanying text.

147. See *supra* notes 126-43 and accompanying text.