IN MEMORIAM

HENRY SAWYER: ADVOCATE FOR THE UNPOPULAR

ARLIN M. ADAMS†

Ralph Waldo Emerson declared that “the true test of civilization is, not the census, nor the size of cities, nor the crops,—no, but the kind of [person] the country turns out.” If America is a great civilization, and I believe it is, it is because of citizens like Henry W. Sawyer, III.

I first met Henry in 1946 at the University of Pennsylvania Law School after we both had returned from service in the U.S. Navy during World War II. Henry had been in a prior class, but I had been re-

† B.S. 1941, M.A. 1950, Temple University; J.D. 1947, LL.D. (Hon.) 1999, University of Pennsylvania; Retired Judge, United States Court of Appeals for the Third Circuit.

1 RALPH WALDO EMERSON, SOCIETY AND SOLITUDE AND POEMS 31 (1921).
leased a little sooner—apparently having served overseas somewhat longer. Since we both were active on the Law Review, I had many opportunities to discuss with him specific cases as well as overall legal philosophy.

I never doubted his brilliance; he had compiled a fine academic record at Chestnut Hill Academy and had been elected to Phi Beta Kappa while an undergraduate at Penn. Henry quickly and cogently could analyze cases, the law, and concepts governing those cases. He was quite articulate in expressing his views and intellectually honest at all times.

Although he had come from a somewhat conservative background, he was disappointed with the entrenched Republican political machine then dominating Philadelphia. Consequently, even while in law school, he became an ardent supporter of the Clark-Dilworth reform movement, which was making great strides locally. He enthusiastically campaigned for them, and they were elected mayor and district attorney, respectively. Henry, himself, was elected to the Philadelphia City Council in 1956.

Henry's real calling was the law, however, and in 1960 he returned to full-time practice with the distinguished firm of Drinker Biddle & Reath, where he remained until his death. In 1953, he agreed to serve as one of the defense attorneys in United States v. Kuzma, a criminal prosecution brought against a group of Communist sympathizers under the Smith Act. A number of attorneys had volunteered for service at the behest of the Chancellor of the Philadelphia Bar Association, who believed that even unpopular defendants were entitled to appropriate legal representation. Although the jury returned a verdict of guilty, the Court of Appeals for the Third Circuit reversed that result.

The courageous effort of these attorneys—led by Tom McBride, later a justice on the Pennsylvania Supreme Court—further

---

2 249 F.2d 619 (3d Cir. 1957).
4 In speaking for a unanimous court, Judge Hastie stated: Nor are we unmindful of the burden upon defense counsel who in this case are volunteers donating months of professional services in a difficult and unpopular cause in token of their high sense of responsibility of the bar to see that all defendants in criminal cases receive competent professional representation.

249 F.2d at 622.
roused Henry's interest in civil rights matters generally, and on behalf of less affluent defendants in particular. In that regard, he represented plaintiffs and achieved victories in two seminal law and religion cases before the Supreme Court: *School District of Abington v. Schempp* and *Lemon v. Kurtzman.* The Schempps were parents of two children who challenged Abington High School's mandatory Bible reading policy as violative of the Establishment Clause. At issue in *Lemon* was Pennsylvania's practice of making public funds available to sectarian schools.

So effective were Henry's arguments before the Court in *Schempp* and *Lemon* that Justice Brennan wrote: "Henry has made a historic contribution to the nation and civil rights law, and few lawyers who have appeared before the Supreme Court during my time have matched his abilities."

Later, Henry agreed to serve as counsel in *Gilfillan v. City of Philadelphia,* which involved an Establishment Clause attack on a proposed expenditure by the City of Philadelphia to construct a vast altar topped by a 60-foot cross at Logan Circle in downtown Philadelphia. Proponents intended the altar to enable the Pope, who was visiting Philadelphia, to conduct a mass. Admission to the area surrounding the altar was to be controlled by the Archdiocese.

A number of very active civil rights lawyers were asked to take on this controversial assignment. Each declined because of the possible negative "fallout." When the Board of Directors of the American Civil Liberties Union approached Henry, however, he did not hesitate to take the case. So far as he was concerned, the validity of an important constitutional issue was at stake.

When I taught the course on religion and the law at the Penn Law School, I always called upon Henry to lead the class when we reached *Schempp* and *Lemon,* and Henry never disappointed. He kept the students mesmerized, recounting his trial strategy and the manner in which he answered difficult questions from the judges.

But the religion clauses of the Constitution were not Henry's only civil rights concern. He was troubled by the treatment of blacks

---

6 403 U.S. 602 (1971).
in the South, and in response he agreed to go to Selma, Alabama in 1965 to defend those unjustly charged with breaking local ordinances. He knew that his security would be threatened, as it eventually was, but a higher calling urged him to go forward.

Henry handled an important case involving Philadelphia's failure to hire a fair number of black police officers. He pressed this matter over a period of nine years before victory was achieved. One issue on appeal was the appropriateness of the counsel fee, which was provided by the applicable statute. Henry and his colleagues—including Robert Reinstein, the present Dean of the Temple Law School—requested only a portion of the payment they ordinarily would have been entitled to claim.

At the time Henry did most of his civil rights work, it was not popular for lawyers practicing with large law firms to become so involved. Nonetheless, he devoted to these matters an enormous amount of his own time with little or no compensation, and in addition, ran the risk of alienating his own clients as well as clients of the firm. He never flinched, however, and his firm backed him to the hilt.

Henry established himself as one of Philadelphia's premier trial lawyers, representing many national corporations in complicated and protracted civil litigation. In all of these, he conducted himself in accordance with the highest professional standards. He was scrupulously honest with his adversaries as well as with the courts. Although aggressive, like most great trial lawyers, he always was courteous and thoughtful with his opponents, both in and out of the courtroom.

Henry was able to see the critical question or questions in a case and to confine himself primarily to those issues without pursuing blind excursions and irrelevant matters. His clients were devoted to him, his opponents admired him, and the courts respected him.

Aside from being a loyal and loving husband to his wonderful wife, Grace, and an ardent supporter to his three children, Hal, Jonathan, and Rebecca, he had scores of friends. And these friends were from diverse areas—geographically and philosophically. Anyone visiting in his home most certainly would become engaged in discussing the critical political and social issues of the time, in addition to being entertained by his extensive collection of jazz recordings.

* * *

---

When I considered a way to end this essay, I first thought of words from Shakespeare's *The Tempest*, words that had special meaning for Henry. But then I believed it more apt to quote the language used by Thomas Jefferson, in writing the *A Bill for Establishing Religious Freedom*:

*I*t is time enough for the rightful purposes of civil government for its officers to interfere when principles break out into overt acts against peace and good order; ... [for] truth is great and will prevail if left to herself; ... she is the proper and sufficient antagonist to error, and has nothing to fear from the conflict unless by human interposition disarmed of her natural weapons, free argument and debate; errors ceasing to be dangerous when it is permitted freely to contradict them.\(^{11}\)

Henry Sawyer represented the very best in America, both in the law and in society in general. He brought to his profession an originality of vision which was based on careful thinking and a keen sensitivity, as well as a passionate belief in the rights of all citizens regardless of their color, gender, or religion.

---

\(^{10}\) See William Shakespeare, *The Tempest*, act 4, sc. 1 ("Our revels now are ended. These our actors . . .").

A FRIEND'S PORTRAIT OF HENRY W. SAWYER, III

WILLIAM T. COLEMAN, JR.

Henry W. Sawyer, III and his wife, Grace, were great art lovers. Does not law, like art, seek to accommodate change within the framework of continuity to bring heresy and heritage into fruitful tension? As Alfred North Whitehead observed:

[A] society maintains its civilization by preserving its symbolic code while giving expression to forces that, repressed, could break a society asunder. And so the basic dilemmas of art and law are, in the end, not dissimilar, and in their resolution—the resolution of passion and pattern, of frenzy and form, of convention and revolt, of order and spontaneity—lies the clue to creativity that will endure.¹

It is daunting to write about Henry, particularly for the Law Review of Henry's law school—a school he revered, and to which he gave so much honor. His classmates, colleagues, and the judges before whom he appeared, received from Henry so much life, inspiration, introspection, direction, and fun, that to think of him no longer here brings tears to our eyes as when the last bars of Wagner's Parsifal fade into silence. Many at the Philadelphia bar could write of Henry with as much knowledge, relish, and relevance as I—in fact, even more.

I have, however, several significant regrets. The prime one, of course, is that Henry no longer is here in person. Young lawyers never again will be in his presence and never will share Henry's élan and intellectual verve. They will be deprived of his path of inspiration to affect the law in its highest function—to make this grand democracy under law work for all in a civilized fashion. Finally, I regret that thinking about Henry reveals a basic defect in the Constitution. His very presence calls into question a clause in article I, section 9—the one providing that "[n]o title of nobility shall be granted by the

† Senior Partner and Senior Counsellor, O'Melveny & Myers, LLP; A.B., University of Pennsylvania 1941; LL.B., Harvard Law School 1946; Law Clerk, Herbert F. Goodrich, United States Court of Appeals for the Third Circuit; Law Clerk, Mr. Justice Felix Frankfurter; Chairman, NAACP Legal Defense and Education Fund, 1961-1975 and 1997-98; U.S. Secretary of Transportation, 1975-1977; Presidential Medal of Freedom, awarded by President William J. Clinton in 1995.

¹ PAULA A. FREUND, ON LAW AND JUSTICE 1 (1968).
United States.” As Thomas Jefferson once wrote to John Adams: “For I agree with you that there is a natural aristocracy among men. The grounds of this are virtue and talents.” Some suggest that John Marshall, Alexander Hamilton, and James Madison would have added a third attribute of “natural aristocracy”—courage. They, unlike Jefferson, fought in the Revolutionary War, just as Henry fought for this country in World War II and the Korean War. In any event, perhaps during a more credulous time, Henry W. Sawyer, III would have been confirmed into this rare class of persons.

I truly enjoyed the time spent working through my thoughts about Henry. For it is an absolute, ultimate joy to relive time spent with Henry as a social friend, public figure, jazz expert, and ally or foe—be it in politics or litigation. I think of riding the Chestnut Hill Local, working through legal arguments to present to Thurgood Marshall and Lou Pollak for the Brown v. Board of Education brief. I picture the family man, surrounded by Grace and the children, being guests at each other’s homes, drinking, eating, and trying to make sense out of the political process, as well as the practice of law. I remember the strong disagreements—which we had many times—yet always keeping humor, respect, and friendship.

There is special affection for a human being who was—by any 1953 definition—a Philadelphia Swell or Philadelphia Brahmin, yet knew that the A Train winds up in Harlem on 125th Street and Seventh Avenue, and has the courage to enjoy with me the night life in that special part of New York. Or knows—as Henry did by experience—that life and spirit resonate below Pine Street in Philadelphia. Or that Temple University is not the only place in North Philadelphia to learn how high and wide the human spirit can soar aided by jazz, by courage, or by the inspirational preaching of a Leon Sullivan or a Bill Gray, Sr. on a Sunday morning. Or to slip off to a nightclub in Atlantic City when attending the Judicial Conference of the Third Circuit, and have me realize that someone knows and understands Lou Rawls much better than even Bill Hastie did.

---

2 U.S. CONST., art. I, § 9, cl. 8.
4 Federal Judge, United States District Court for the Eastern District of Pennsylvania; Dean of the University of Pennsylvania Law School (1975-78).
6 William H. Hastie was a Judge for the United States Court of Appeals for the Third Circuit from 1950-71, and Chief Judge from 1968-71.
Henry's distinctive mark, or genre, is that of a great gentleman: lettered, urbane, charming, witty, affable, and caring. He was someone who dealt with friend and foe without condescension and always was approachable. A French expression—escalator de mots—refers to the brilliant reply you should have made to the raconteur at the cocktail party, but only thought of going down the stairway when leaving. Often in a social, political, or legal conversation, and especially in court, I envied the fact that Henry came up with the bright thought in the right canter or pace and the right words when I was still stumbling.

This portrait also reveals a young person who sought expression and meaning in a political party, the minority party in Philadelphia at the time, and clearly not the party of Philadelphia's Swells. He had the nerve to live in its distractions, indeed to take an active part in its activities. He leavened its coarseness, for example by insisting on an ordinance that required all new building projects to earmark one percent of its costs for art. Cities and counties across the country have adopted this Sawyer brain-child. As the United States Secretary of Transportation, I later adopted this policy for all projects financed in whole or part by that Department.

Legislation often is described as being like sausage—you don't want to see it being made. Can you imagine the taste or zest of the sausage when mixed by Jim Tate, Raymond Pace Alexander, and Henry Sawyer? History truly will suffer if the tales of the Philadelphia City Council from 1956-59 that Sawyer reveled in and retold over a Saturday beer at the Sawyer swimming pool are not among the records of Drinker Biddle & Reath.

Now for a moment let us put on the canvas, Henry in the profession he so dearly loved. A lawyer's life, like every other occupational life, has much drudgery—senseless bickering, stupid obstinacies, capricious petitfogging—all disguising and obstructing the only sane purpose which can justify the whole endeavor. These, of course, take an inordinate part of the lawyer's time; they harass and befog the unhappy wretch and at times nearly drive him from the particular work-

---

7 I refer here to the Democratic Party.
9 President of Philadelphia City Council, and later Mayor of the City of Philadelphia.
10 A leading member of the Philadelphia Bar, and later a judge in the Common Pleas Court of Philadelphia.
place where the work must be done. But near the end, when the trial almost is complete or the appellate argument is to be called, the turmoil must stop and craftsmanship must take over. Then the resultant joy in creativity begins. And out of this murk, the pattern emerges. For the superb advocate, like Henry, it must be one’s own pattern, and one’s own expressions. In that atmosphere, at that crucial moment, Henry reveals himself the master, as that rare lawyer who not only knows, but, even more important, knows why, and, still more important, knows how. The gifted, attractive, remarkable, capacious-minded Henry, with all-seeing eyes, lays aside efficiency and dispatch, as he never is shy of recondite learning, always on point.

For example, in 1961, when he argued Deutch v. United States, Henry knew how to work through the pitfalls. Just two years prior, the Court upheld a contempt of Congress citation in Barenblatt v. United States. Moreover, Watkins v. United States essentially had spelled out the almost absolute power of Congress. So what did Henry do? First he skillfully briefed a far more sweeping constitutional question—should a witness be compelled to inform on a friend? Years before Americans expressed their uneasiness over the Lewinsky-Tripp matter, Henry instinctively had grasped the discomfort of such a requirement. Any Court, even in 1961, would wish to avoid answering that difficult question and would seek resolution in a less controversial legal harbor. But first Henry had to share the unfairness of his client’s treatment by the court below. He made it crystal clear by suggesting that the person who had persuaded the defendant to stay in the Communist Party was an agent of the FBI, a fact not even mentioned in the D.C. Circuit opinion. Thereafter it became easier for the majority to determine that the Government had not shown the defendant’s responses about his associates to be pertinent to the congressional inquiry.

Another Henry signature style as an advocate was that he often presented a Rhadamanthine detachment despite the fact that many lawyers think that detachment can subvert a practical advocate’s zeal. In Greek mythology, after all, Rhadamanthus was a judge, not an advocate. Thus many think detachment to be more desirable in a judge

---

11 367 U.S. 456 (1961) (reversing a D.C. Circuit opinion that affirmed a contempt of Congress conviction for a congressional witness who, even though freely admitting all his own acts and meetings with reference to the Communist Party, refused to inform on others).
or scholar than in a trial lawyer in the pit or at the bar of an appellate
court. But in Freund's classic book, Justice Brandeis reminds us that
good opinions are made when the Court and the advocate are in
equal rhythm. In other words, as Jeremy Bentham said, "The Law is
not made by judge alone, but by judge and company." Thus a splen-
did technique in the hands of a master like Henry is to show detach-
ment—even though presenting the legal position quite sincerely,
quite effectively, and quite clearly—in order to get one to go along
with your contentious legal point. Henry often observed that a strong
but subtle suggestion often is better than a hit on the head to make
the key point in an argument very effectively. He recognized that the
great appellate judge does not wish to be accused, as Daniel Webster
often accused Chief Justice John Marshall, of just copying the advo-
cate's brief.

One cannot end this portrait without a dash about Henry's tre-
mendous sense of humor. His "all-seeing eye" gave him the capacity
to engage in and admire colleagues of sharply different philoso-
phies—whether right or left, rich or poor, black, red or white, dumb
or brilliant, Jew or Gentile—so long as they fairly engaged in discus-
sions and maintained a sense of humor. On one occasion, his love
and appreciation for the way that the Quakers managed the
Germantown Friends School conflicted with the fact that Henry twice
had been to war. Henry had heard the anger of enemy fire and thus
realized that, unfortunately, tyrants often could not be subdued by the
friendly thoughts of Quakers, their ability always to listen, never to
think ill of a human being, and to seek unanimity in the Meeting
House. As he left that one particular Quaker meeting, he said of an
opponent (known for voluminous comment), "He never took his eye
off the ball, for he never saw it." And when talking about Martin Dies,
Parnell Thomas, or Joe McCarthy—three of witch-hunt fame—he of-
ten remarked, "Why must there be so many bastards born in wed-
lock?"

Henry had an undying appreciation for Drinker Biddle & Reath—
the law firm that gave Henry the head, range, and support to do what
he thought necessary in civil rights and civil liberties cases. Henry be-
came an associate and a partner because of his brilliance as a lawyer
and his ability to represent large industrial corporations in the bet-the-
company cases, whether antitrust, securities, or patent matters. In a

---

14 PAUL FREUND, ON UNDERSTANDING THE SUPREME COURT (1950).
15 Id. at 78.
16 See id. at 79-80.
nationwide electrical antitrust suit, General Electric, the lead defendant, retained Henry as local counsel. Before long, Henry's impressive command of the law and facts of the case induced the defendant's General Counsel to appoint Henry as lead counsel, replacing a pillar of the D.C. trial bar and senior partner in a major D.C. firm. Earlier, Henry had cut his teeth in the *Viking Theatre* case, which introduced him to the feverish, competitive world of movie distribution where Louis Nizer was opposing lawyer. Henry also convinced Polish Communists that the American Courts' hallmark was fairness and justice. In the Polish Golf Cart case, where the United States was plaintiff, Henry achieved a complete victory for his Communist client.

Henry—because of his performance as a corporate litigation lawyer—became a partner in record time, even though his career was interrupted after one year in law school by five years active duty as a U.S. Navy Commander in the Pacific in World War II. After starting in practice, he served one year in the Korean War, and two years in Europe connected with NATO. He also handled more civil liberties and civil rights cases than did most lawyers who practice exclusively in these fields. Three were victories in the Supreme Court, but a computer search reveals over thirty-five cases in the lower courts. Many of Henry's causes were unpopular, and this occurred long before the "white-shoe" law firms felt such pro bono work fit within their strategy. In fact, many often felt their business, commercial, and governmental clients would disapprove. But Drinker was strikingly different. On the day in 1961 that Henry argued *Schempp*—challenging the Lord's Prayer recitation in public schools—Messrs. Drinker, Biddle and Van Dusen sat in the front row.

From the day Henry started practicing law—soon after his first year of law school, interrupted by five years in the Navy, and during his work on the *Law Review*—he always sought variety in everything. He believed in Goethe's famous passage to those of any family wealth:

---

18 *Viking Theatre Corp. v. Paramount Film Distrib. Corp.*, 320 F.2d 285 (3d Cir. 1963).
21 This term has nothing to do with race, but what the less fortunate of us who did not go to Princeton or Virginia called those men who, in the style of 1938, wore white shoes when dressed properly.
"You must labor to possess that which you have inherited." Henry once expressed (during one of our Chestnut Hill Local rides) his recognition that routine is a kind of narcotic. It keeps the mind so occupied that one thinks very little about what really matters, thus losing the roar of the waves to the fullness of the sea. After all, such challenging adventures are not important for a lot of mankind. Emerson put it right when he said "[m]ankind is as lazy as it dares to be." But Henry always has been driven to reject what Anatole France suggests in *M. Bergerat*: "Je comprends, c’est mon faiblessé, il y a beaucoup de face de ne pas comprendre." For Henry, it is always important to weigh the pros and cons, and always important to play a role in the great social and political issues of the day. Further, it always was important to Henry that the court understand the business issues of the client, or see that the wrongful uses of governmental power—which bring out the worst in man—be retarded. Senator William Borah, and John Greenleaf Whittier, though a century apart, help supply the last splashes to this portrait of Henry, as Jackson Pollock, an artist so admired by Henry and Grace Sawyer, would do. Senator Borah said:

The safeguards of our liberty are not so much in danger from those who openly oppose them as from those who, professing to believe in them, are willing to ignore them for their purposes. . . . The latter undermine the very first principles of our government and are far the more dangerous.

And Whittier, about Charles Sumner, wrote:

"Forego thy dreams of lettered ease,
Put thou the scholar’s promise by,
The rights of man are more than these."
He—heard, and answered: "Here am I!"
A VOICE FOR LIBERTY

STEWART DALZELL†

I first saw Henry W. Sawyer, III in action in 1973 when I served as his junior associate in an unfair competition case that had been transmuted into a Sherman Act section 1 claim. The defendants retained Henry after suffering what was for them a catastrophic verdict at the hands of a jury in a Philadelphia federal court. When I came into Henry's office, he explained to me that the five individual defendants were facing personal ruin as a result of the verdict and, more immediately, from execution on the seven-figure treble damage award.

Our urgent task was to stay execution on the judgment, while the plaintiff corporation, through its able counsel, was insisting that the individuals post their houses as security for the stay under Rule 62(d) of the Federal Rules of Civil Procedure. Since all five of these men were married, this meant that their wives' interests were at stake and, as one would expect, they were horrified at the prospect of losing their homes. We therefore filed a motion on behalf of the wives (as well as their husbands) for relief from this Draconian condition, and shortly thereafter, I for the first time, saw Henry Sawyer as an advocate, pressing our motion before my now-colleague, Judge Raymond Broderick.

What immediately struck me as I heard Henry arguing our motion was what a terrific, impressive voice he had. If memory serves correctly, we were seated in Judge Broderick's chambers at the old Federal Courthouse at Ninth and Market Streets in Philadelphia. Seemingly without raising the volume, Henry Sawyer's rich bass voice nevertheless filled the room and kept everyone—most importantly, Judge Broderick—in rapt attention. I still can hear the sound of that voice saying to Judge Broderick, at a time when the women's movement was still largely a gleam in feminists' eyes, "Your Honor, women are people. They have an existence and interests separate and apart from their husbands. These women had nothing in any way, shape, or

† United States District Court Judge for the Eastern District of Pennsylvania. B.S. 1965, J.D. 1969, University of Pennsylvania. Judge Dalzell was associated with the Philadelphia law firm of Drinker Biddle & Reath from 1970-1976, and then was a partner with that firm (with Henry W. Saywer, III) until leaving for the bench in 1991.
form to do with what the plaintiff claims in this case."

To my amazement, and before my very eyes, the voice worked its magic. At least to the extent that we were given the opportunity to provide substitute security—as it turned out, for less than the full amount of the verdict—the marital homes were never again in danger. Indeed, the case ultimately settled on appeal, on terms that brought our clients back from the brink of bankruptcy and enabled them to put their energies and talents to rather rewarding business successes in later years. As one might imagine, they were convinced that Henry W. Sawyer, III had saved their business and personal lives. They were right.

I thereafter had the great good fortune of working closely with Henry Sawyer in other commercial litigation, most notably *Outboard Marine Corp. v. Pezetel*. In that case, we represented the then-state-owned Polish entities who made golf carts that sold so successfully in this country that the domestic manufacturers of those vehicles became rather upset. In *Pezetel* and in other cases, I heard that unforgettable voice, witnessed its crafty modulations, and beheld what it could do to many tribunals and gatherings of lawyers.

Indeed, over the course of the more than fifty years that Henry Sawyer practiced after his graduation from this law school in 1947, he used that highly sophisticated voice in many cases in courts around this country. As successful as Henry Sawyer was in tuning that voice to advance the interests of his paying clients, he always in my years of practice with him saw those cases, however zealously he argued them, as a means to a larger end. Successful as he was in these private, and

---

3 Interestingly, Henry was recommended to the Poles by Professor Louis B. Schwartz, who taught me criminal and antitrust law at Penn, and for whom I worked as a research assistant. Professor Schwartz remained as our co-counsel in the litigation before Judge Murray Schwartz (no relation) in the Delaware District Court and in the antidumping administrative proceeding in the United States Treasury Department.
4 A Westlaw search reveals over one hundred citations of reported cases in which Henry Sawyer appeared as counsel from 1951 to the early 1990s. Of course, for much of Henry's career, judicial memoranda and opinions were not transmitted routinely to Lexis and Westlaw. Thus, for example, there are no reported decisions in two significant Robinson-Patman Act cases Henry and I worked on (together with Edward M. Posner), *Joseph Walsh Tire Co. v. B.F. Goodrich Co.*, C.A. No. 74-1310 (E.D. Pa.) and *Hub Tire Co. v. B.F. Goodrich Co.*, C.A. No. 75-85 (E.D. Pa.), cited in Broyer v. B.F. Goodrich Co., 415 F. Supp. 193, 195 nn.1-2 (E.D. Pa. 1976).
largely commercial, cases, the reason this Law Review has gathered these tributes is because of Henry Sawyer's contribution in public interest litigation. And it was in that litigation that one always found Henry Sawyer's heart and, above all, his voice in its most resonant and powerful pitch.

Henry Sawyer usually is identified with his successes in two of the century's most important Establishment Clause cases, School District of Abington Township v. Schempp and Lemon v. Kurtzman. There is no question that both cases constitute major legal landmarks in First Amendment jurisprudence, as well as in the everyday life of any American who has ever set foot in a public or sectarian school. Given the lopsided votes in the Supreme Court, however—eight-to-one in Schempp and seven-to-one in Lemon—a detractor might contend that Henry Sawyer was just lucky enough to be in the right place at the right time and won two cases that really did not depend on the power and sound of that great voice. Although I would argue to the contrary—particularly given the reality that Henry lost before the Lemon three-judge court—it seems to me hard to dispose of such detraction, at least directly.

Another Supreme Court case, however, provides conclusive evi-

5 In private litigation, Henry Sawyer probably was best known for his work in the antitrust area, an expertise that began when he was (unsuccessful) trial counsel for the plaintiff in Viking Theatre Corp. v. Paramount Film Distributing Corp., 320 F.2d 285 (3d Cir. 1963), though on the briefs and the certiorari petition, the Supreme Court did not hear Sawyer's voice, but that of Edward Bennett Williams. The decision of the Court of Appeals was affirmed by an equally divided Court. See 378 U.S. 123 (1964). Henry always believed that his work for Viking Theatre led to General Electric's retaining him in the civil electrical conspiracy price-fixing cases, a litigation so complex it became the impetus for the later enactment of 28 U.S.C. § 1407. See, e.g., United States v. General Elec. Co., 209 F. Supp. 197 (E.D. Pa. 1962).

6 374 U.S. 203 (1963) (holding that state-mandated prayer and devotional Bible reading in public schools are unconstitutional under the First Amendment's Establishment Clause).

7 403 U.S. 602 (1971) (holding that state financial aid to nonpublic schools was an "excessive entanglement between government and religion" and therefore was unconstitutional under the First Amendment's Religion Clauses).

8 In Schempp, only Justice Stewart dissented, see 374 U.S. at 308-20, though Justices Douglas and Brennan wrote concurrences to Justice Clark's opinion for the Court. Lemon involved one Pennsylvania and two Rhode Island cases. Justice White concurred in the Pennsylvania case and dissented in the Rhode Island cases, see 403 U.S. 661-71, and Justices Douglas and Brennan wrote concurrences to Chief Justice Burger's opinion for the Court in the Rhode Island cases. Justice Marshall did not participate.

dence that Henry Sawyer's voice indeed made a decisive difference, not only in those two Establishment Clause cases, but also for the cause of liberty in this country.

I refer to the case of *Deutch v. United States*. At the time young Bernhard Deutch came to Henry Sawyer's office, the nation was consumed in what we now call the McCarthy Era. The facts of Bernhard Deutch's case demonstrate how far the Zeitgeist had taken our political and legal life from its constitutional moorings—as well as from the common "decency" to which Joseph Welch referred in his famous colloquy with Senator McCarthy.

Two months before Welch's rhetorical questions about Senator McCarthy's "decency," Bernhard Deutch was working in the Physics Building of the University of Pennsylvania. He was served with a subpoena that commanded him to appear in Albany, New York two days later before a subcommittee of the House Committee on Un-American Activities. Though still only a graduate student, Deutch had the good sense to choose Henry Sawyer as his lawyer.

The subcommittee in its grace granted Henry's request for a continuance, and the appearance occurred three days later than scheduled, and in Washington, D.C. As Henry later described the Kafkaesque scene in his Brief for Petitioner, "Upon appearance at the committee's office in the House Office Building in Washington, [Deutch] and his counsel were directly shown into an office in which there were seated several unidentified men; he was forthwith sworn and without preamble the questioning commenced."

The committee counsel informed Deutch that:

[D]uring hearings at Albany last week, the committee heard testimony regarding the existence of a Communist Party group or cell operating among undergraduates at Cornell University, among certain graduates at Cornell and in the city of Ithaca.

---


11 The great Boston lawyer's words to Senator McCarthy, after McCarthy's gratuitous slander of Welch's young associate, bear repeating to give the flavor of the era in which Sawyer represented political outcasts like Deutch: "Until this moment, Senator, I think I never really gaged [sic] your cruelty or your recklessness. . . . Let us not assassinate this lad further, Senator. You have done enough. Have you no sense of decency, sir, at long last? Have you left no sense of decency?" *Special Senate Investigation on Charges and Countercharges Involving Secretary of the Army Robert T. Stevens et al.: Hearing Before the Special Subcomm. Investigations of the Comm. on Gov't Operations*, 83d Cong., at 2429 (1954) (statement of Joseph N. Welch to Senator Joseph McCarthy).

In connection with that testimony, the committee was informed that you were a member of one or more of those groups.\textsuperscript{13}

Whereupon, Deutch explained that he had met a charismatic black law student at Cornell named Ross Richardson, who eventually persuaded Deutch to join a small Communist Party cell. Richardson would drive Deutch to the meetings and collect dues from him.

Although Deutch answered most of the subcommittee's questions, he declined to answer five, all relating to people other than himself. For example, the first question he declined to answer was:

1. The committee was advised that a witness by the name of Ross Richardson has stated that you acted as liaison between a Communist Party group on the campus and a member of the faculty at Cornell, and that you knew the name of the member of that faculty, who was a member of the Communist Party. Will you tell us who that member of the faculty was?\textsuperscript{14}

As Henry Sawyer later told the Supreme Court, Deutch “respectfully told the committee that he was willing to tell all about his own activities but that he could not, because of moral scruples, bring himself to inform on other people.”\textsuperscript{15} In any event, Deutch made it clear that the subcommittee easily could get all of the requested information from Ross Richardson.

When the House of Representatives voted a contempt citation against Deutch, he was indicted for refusing to testify, a violation of 2 U.S.C. § 192.\textsuperscript{16} The bench trial took place in the District Court for the District of Columbia in 1956, before District Judge Alexander Holtzoff. In his brief opinion in support of the conviction as to four of the five counts, Judge Holtzoff not only gave a ringing endorsement to the power of Congress to investigate domestic Communism in all its details, but also summarily rejected Henry Sawyer's argument that the particular questions were not pertinent to the subcommittee's inquiry.

\textsuperscript{13} Id. at 5-6.

\textsuperscript{14} Id. at 8. The other questions and orders were:

2. Will you tell the committee, please, the source of that $100 contribution, if it was made?

3. Where were these meetings held?

4. Were you acquainted with Homer Owen?

5. The witness is directed to give the name of the person by whom he was approached.


\textsuperscript{15} Brief for Petitioner, supra note 12, at 7.

\textsuperscript{16} See 2 U.S.C § 192 (1997) (setting forth liability for refusing to testify or produce papers during a congressional investigation).
To this contention Judge Holtzoff wrote that a congressional committee "has a right to obtain cumulative testimony" and that it may question "several witnesses to the same matter in order to check the accuracy of the information that it is obtaining."\(^{17}\) Judge Holtzoff acquitted Deutch on Count Three, holding that someone "may not be punished for contempt of Congress merely for stating that he does not remember."\(^{18}\)

During its long pendency in the Court of Appeals for the District of Columbia Circuit, the case had become rather more difficult owing to the Supreme Court's intervening decision in *Barenblatt v. United States*.\(^{19}\) In that case the majority upheld, over First Amendment objections, the power of a subcommittee of the House Committee on Un-American Activities to inquire into a witness's past or present membership in the Communist Party. As the Court of Appeals interpreted *Barenblatt* in Deutch's case, "[T]he Government has proved beyond a reasonable doubt that the subject under inquiry and the pertinency of the questions were made to appear at the committee hearing with 'indisputable clarity.'"\(^{20}\) The Court of Appeals's affirmance thus could not have come as a surprise.

On July 13, 1960, when Henry Sawyer filed his petition for a writ of certiorari to the District of Columbia Circuit, he was facing what then had to have seemed like a hopeless enterprise. The Supreme Court only a year before had decided *Barenblatt*, with its approval of wide-ranging and intrusive power on the part of Communist-hunting congressional committees. Though *Barenblatt* was decided by a vote of five to four, the fifth vote was supplied by the Court's newest member, Justice Potter Stewart, who joined Justices Frankfurter, Harlan, Whittaker and Clark.\(^{21}\) Where Lloyd Barenblatt had cited the First Amendment as the basis for his refusal to answer, Bernhard Deutch only mentioned "moral scruples." True, Henry by this time had achieved some success for another "unfriendly" witness on the

---


\(^{18}\) Id.

\(^{19}\) 360 U.S. 109 (1959).

\(^{20}\) Deutch, 280 F.2d at 695. In addition to *Barenblatt*, the panel also cited *Watkins v. United States*, 354 U.S. 178 (1957), and *Flaxer v. United States*, 358 U.S. 147 (1958), for the quoted proposition—all cases decided after the bench trial before Judge Holtzoff.

\(^{21}\) Justice Black's dissent was unusually powerful, even for him, particularly in its quiet anger in referring to the majority's "conclusion that on balance the interest of the Government in stifling these freedoms [of speech, press, assembly and petition] is greater than the interest of the people in having them exercised." 360 U.S. at 143 (Black, J., dissenting) (internal quotations omitted).
grounds that a Senate subcommittee had failed to establish the subject matter of the inquiry. But this argument was not available to him in Deutch because, as seen above, the subcommittee counsel made the purpose of Deutch's appearance quite explicit at the beginning of the proceedings on April 12, 1954.

When, probably very much to his surprise, the Supreme Court granted his petition for a writ of certiorari, Henry Sawyer faced a daunting task. The judicial branch seemed oblivious to Senator McCarthy's death on May 2, 1957. Three years after that death, the Court of Appeals unanimously affirmed Judge Holtzoff, who was by then one of the most accomplished district judges in the federal system (the Court of Appeals panel included later-to-be Chief Justice Burger). The Barenblatt majority in 1959 included the intellectual weight of Justices Frankfurter and Harlan.

When it came time for the oral argument, however, Henry brought with him his ultimate weapon: his voice. He used that voice to tell quite a story. Henry stressed Ross Richardson's involvement in keeping Deutch in the Party. He pointed out that when Deutch wanted to resign his Party membership, Richardson retorted that Deutch was a "white chauvinist" who could not stand having a black in control. Deutch relented and remained in the Party until he left Cornell to attend the University of Pennsylvania for his graduate work.

After recounting these facts, I suspect from many years' experience with him that Henry Sawyer then allowed a long, pregnant pause to fill the courtroom with silence. He always would do this before his punch line, and what a powerful one he had that day for his audience of nine. Doubtless reaching down to the lowest registers of that grave

---

22 In Knowles v. United States, 280 F.2d 696 (D.C. Cir. 1960), Henry won a reversal on behalf of a woman discharged from her position as a librarian in Norwood, Massachusetts, who had been called three times to Washington to appear before the Subcommittee to Investigate the Administration of the Internal Security Act and Other Internal Security Laws, of the Committee on the Judiciary of the United States Senate. When she appeared on July 29, 1955, "no statement was made as to the Subcommittee's purpose in calling her, or as to the subject matter of the inquiry." Id. at 698. Her conviction therefore was reversed with instructions to dismiss the indictment. See id. at 700.

23 See supra text accompanying note 18.

voice of his that suitably could announce the Apocalypse, he then disclosed to the Court that, at all times, Ross Richardson worked as an agent for the Federal Bureau of Investigation.

On its face, Henry's shot at Ross Richardson was gratuitous and irrelevant to all but Bernhard Deutch and his advocate. Deutch was the one who refused to answer the subcommittee's questions, not Richardson. And Richardson's true role was utterly beside the point regarding the pertinence of those five questions to the subcommittee's inquiry.

But we know that this ornament to Henry's recitativ not only was relevant, but decisive for at least one other listener in that audience. As Henry many years later reported to me, "Justice Potter Stewart (hitherto on the government's side of this issue) leaned forward and said, 'Is that in the record, Mr. Sawyer?' I said indeed it was inasmuch as Richardson was the sole witness against Deutch at the trial." To this Justice Stewart leaned back and said to a colleague (probably Justice Brennan), "'Outrageous . . . outrageous!'" in a volume "more voce than sotto."

In writing his opinion for the five-Justice majority reversing Deutch's conviction, Justice Stewart, after recounting the basic facts, mentioned, at the very end of one of those pregnant and powerful footnotes the Supreme Court sometimes drops, that Ross Richardson had joined the Party "at the behest of the Federal Bureau of Investigation." He then went on to hold for the Court that the questions indeed were not "pertinent to the committee's inquiry."

Technically speaking, Justice Harlan's dissent in Deutch had much power, given the recent vintage of Barenblatt. But once the Court got the whole story of Deutch's case, in truth and justice there is no doubt that the conviction had to be reversed if the nation was to recover a modicum of the decency that Joseph Welch and many others had found so lacking in that poisonous time.

When Henry Sawyer's bass voice filled the silence of the Supreme Court's chamber with the words, "Ross Richardson was at all times an agent of the Federal Bureau of Investigation," can there be much

---

26 Id. at 3.
27 Deutch, 367 U.S. at 460 n.4.
28 These were the words Henry recited to me that he actually said. The Supreme Court Clerk's Office was unable to find a transcript of the oral argument in its retrievable records when I sought it in 1998 in writing the essay cited above in note 10.
doubt that these words, more than Joseph Welch’s oft-quoted comment to Senator McCarthy, truly began the end of this dark chapter in our jurisprudence and national life? To be sure, that voice remained jurisprudentially incorrect to some of the brightest and best judges for some years after McCarthy’s death—indeed, that voice so offended the judicial sensitivities of Justice Frankfurter in March of 1961 during Sawyer’s argument, that the learned Justice “turned his chair around, facing backward” as Henry continued speaking.29

Henry Sawyer’s voice really did have that much power to it. True, on a day-to-day basis it was often in the service of private interests, since Henry had to make a living and someone had to pay for all of that pro bono work. Indeed, Henry explicitly acknowledged to me that his work on behalf of significant economic interests served these larger ends.30 The truth of the matter is that I do not believe Henry Sawyer ever made a dime representing Communists and other political outcasts. It was simply of no interest to him.

But Henry Sawyer always would say that he got much more out of representing these clients who could not pay him. He unquestionably was not speaking of gaining notoriety. He was talking about sometimes actually achieving justice. And with that great voice of his, he

---

29 Sawyer, supra note 25, at 2 & n.4. To get another sample of the flavor of the judicial atmosphere long after the echo had died to Joseph Welch’s comment, the Supreme Court of Pennsylvania’s decision in Kaplan v. School District of Philadelphia, 388 Pa. 213 (1957) will do rather well. When that Court ruled against Henry Sawyer’s client, it began the opinion with the following two sentences:

On July 12, 1948, Samuel M. Kaplan, the plaintiff in this case, entered into a contract with the School District of [Philadelphia], the defendant, to teach English in the public schools of Philadelphia on a salary basis. So far as the record shows, he demonstrated himself to be a competent teacher in his field, but it was reported to the Superintendent of Philadelphia Schools that while Kaplan was instructing the children in his classes how to express themselves in English he was devoting time to an organization which, if successful in its plans, would eventually have those children or their children’s children speaking Russian in a Russian state.

Id. at 214 (Musmanno, J.).

30 To its credit, besides providing a significant financial subsidy, Drinker Biddle & Reath itself withstood a good deal of criticism when Henry was at the height of representing accused Communists in the McCarthy Era. He once told me a rather touching story in this respect. After Henry Drinker died, his secretary of many years reported for the first time to Henry Sawyer that a representative of one of the firm’s largest clients called to complain to Mr. Drinker about the young Sawyer representing some Communists in a protracted Smith Act prosecution. After Mr. Drinker assured himself that the client took no exception to Henry’s legal abilities or ethical performance—indeed, the essence of the client’s complaint was that too much talent was being put in service of such disreputable people—Mr. Drinker advised the client that if he was so upset he should take his business elsewhere.
did more than that. He was one of the pillars of his time in securing the liberties that benefit all Americans. Henry Sawyer’s voice therefore spoke for every one of us.
Practicing lawyers tend to get short shrift from the law reviews, especially when it comes to lapidary statements. Although not an everyday occurrence, it is not unusual for a law review to publish memorial tributes to a deceased law professor held in high esteem. And so too with an admired judge for whom the bell has tolled. But it is rare for a practitioner to be celebrated in this way. So it is a good thing that the editors of the University of Pennsylvania Law Review—recognizing that the practice of law is an established, here-to-stay, non-trivial ingredient of the profession—decided to dedicate an issue of the Law Review to the memory of a practitioner, Henry W. Sawyer, III. Henry was a singularly gifted advocate who achieved greatly in litigation of great consequence. He ennobled our profession. He deserves to be remembered in the pages of this venerable journal—the Law Review of which he was Managing Editor more than half a century ago.

The dominant dynamic elements, professional and personal, of this remarkable lawyer have been faithfully captured in the three tributes accompanying this one. Stewart Dalzell, in defining Henry Sawyer’s decisive impact on trial and appellate courtrooms, has, in a compellingly felicitous phrase, termed Henry a “Voice of Liberty.” William Coleman, building on Whitehead’s linkage of art and law, has crafted a lifelike “[p]ortrait” of Henry—a portrait of an artist who remained a young man almost to the end of his long life. And Arlin Adams has traced the unorthodox career of a lawyer willing to take on the “[u]npopular” cases.

A comprehensive survey of Henry’s artistry as an advocate would call for close scrutiny of the dozens and dozens of trials and appeals Henry handled in four decades of mastery of the courtroom. But the timetable of this tributary issue of the Law Review does not permit a research enterprise of that dimension. Perforce, it is necessary and

† United States District Judge for the Eastern District of Pennsylvania. For his description of the events that led to the filing of the suit that was to bear fruit as School District of Abington Township v. Schempp, 374 U.S. 203 (1963), the writer is indebted to his friend Bernard Wolfman—professor of law at Harvard since 1976, and Dean of the University of Pennsylvania Law School from 1970-75. See infra note 4 (describing Wolfman’s professional history).
proper to confine our focus to the Sawyer cases of greatest public consequence: the three cases Henry took to the Supreme Court. Confining the focus in this fashion foreshortens—and to that extent distorts—Sawyer's *corpus juris*, it leaves out of account chapters of the Sawyer story that matter a lot, such as the summer of 1965 which Henry spent in Mississippi together with scores of other lawyers (most of them younger) who signed on to protect the voting rights of black Americans. But the three Supreme Court cases were of greatest public consequence. In each Henry prevailed—to the immediate benefit of his clients and to the lasting benefit of his country.

The Dalzell essay centers on the first of the three cases—*Deutch v. United States*. In arguing *Deutch*, Henry wove into his detailed recital of the events giving rise to the prosecution a single electrifying fact—a fact of no readily demonstrable doctrinal authority but of overwhelming moral authority—which, as the Dalzell essay explains, appears to have been the ingredient which moved five justices to overturn Henry's client's conviction for contempt of the House Committee on Un-American Activities. I will discuss Henry's two other Supreme Court victories—*School District of Abington Township v. Schempp* and *Lemon v. Kurtzman*. In doing so, I will try to avoid another distortion endemic in reprises of leading Supreme Court cases—the tendency to look only at the argument and opinions in the Supreme Court, neglecting what transpired below. In both *Schempp* and *Lemon* it is clear that a masterly litigator planted in the trial court the seeds of the fruit that was to ripen on appeal.

**ABINGTON SCHOOL DISTRICT V. SCHEMPP**

The *Schempp* case had its inception in a letter written in 1957 to the Philadelphia chapter of the American Civil Liberties Union. The letter writer, Ellory Schempp, was a high school student, the oldest of three Schempp children enrolled in the public schools of Abington Township, a Philadelphia suburb. The letter described the discomfort that Ellory, whose family was Unitarian, felt during the Bible reading and recital of the Lord's Prayer that launched each school day. Ellory wondered whether the ACLU—which, so he understood, cared about issues of this sort—would regard this as a problem. Ellory had heard good things about the ACLU's endeavors and he enclosed ten dollars

---

2 374 U.S. 203 (1963) [hereinafter *Schempp*].
3 403 U.S. 602 (1971).
to further those endeavors.

On receipt of the letter, Spencer Cox, the executive director of the ACLU's Philadelphia chapter, consulted Bernard Wolfman, an ACLU member and young partner at Wolf, Block, Schorr & Solis-Cohen. Wolfman, who lived not far from Abington, agreed to call upon the Schempps and explore the matter. An interview with Ellory and his younger siblings, Donna and Roger—all three seriously troubled by the obligatory prayer exercises—and then with their parents, satisfied Wolfman that the Schempp family was prepared for the difficulties that litigation might entail, if the board of the ACLU chapter were to conclude that the chapter should take on the matter. Some initial research persuaded Wolfman that the morning exercises that disturbed the Schempps—the Bible reading portion of which was required by a Pennsylvania statute—posed substantial and unsettled constitutional questions.

Wolfman then reported his findings to the board of the ACLU chapter. All members of the Board were persuaded that the Schempps' religious freedom issues were proper ACLU issues. Several members, however, felt that taking on a litigation burden of such expectable magnitude was not a prudent allocation of limited resources, given their commitment to assisting those still being tarred by McCarthyism and its dismal legacy. After extended discussion, the Board voted—only to find itself equally divided. The deciding vote was that of the Chair, Clark Byse, a Penn law professor. Byse noted that, as a Catholic, he derived great comfort from the Bible; but, since all of his fellow board members saw in the Schempp children's predicament constitutional issues of gravity, he would vote with those who felt the chapter should agree to take the matter on.

The Board's decision to provide counsel for the Schempps, however, did not mean that Wolfman would be that counsel. Wolfman decided that for him, as a Jew, to represent the Schempps in a challenge to Bible reading and recitation of the Lord's Prayer merely would add unnecessary and probably detrimental baggage to what clearly would be a controversial and, in many quarters, an unpopular cause. Wolfman so advised Spencer Cox and recommended that his

---

4 Wolfman was a year behind Sawyer at the University of Pennsylvania Law School, and they served together as editors of the Law Review. In 1963, Wolfman left private practice to return to Penn Law School as a member of the law faculty, and he served as Dean from 1970-75. In 1976, Wolfman joined the Harvard Law School faculty as Fessenden Professor of Law.

5 Byse left Penn for Harvard in 1958, where he is Byrne Professor Emeritus.
friend Henry Sawyer—an ACLU member and young partner at Drinker Biddle & Reath whom Wolfman had known since they were fellow law students at Penn—be asked to assume the representation. Cox acquiesced, so Wolfman presented the proposal to Henry Sawyer, and Henry agreed to represent the Schempps.

In the District Court

In February 1958, Henry filed the Schempp complaint in the United States District Court for the Eastern District of Pennsylvania. The defendants were the Abington School District, its Superintendent, and other school officials. The object of the suit was to obtain a decree enjoining the enforcement of section 1516 of the Public School Act and to declare it unconstitutional. Section 1516 was a directive that "at least ten verses from the Holy Bible shall be read, or caused to be read, without comment, at the opening of each public school on each school day, by the teacher in charge." Although not required by the statute, the prescribed Bible reading was, in the Abington schools, routinely followed by a recitation in unison of the Lord's Prayer, which in turn usually was followed by the Pledge of Allegiance. Henry's theory of the case was that the prescribed Bible reading, whether or not followed by the Lord's Prayer, constituted both an establishment of religion and an infringement of the free exercise of religion in contravention of the First Amendment, as made applicable to the states by the Fourteenth Amendment.

Under the provisions of the federal Judicial Code (Title 28) then in force, a constitutional challenge to a state statute was required to be heard by a three-judge district court, at least one of whose members was a circuit judge. The Schempp district court consisted of Chief Circuit Judge Biggs and District Judges Kirkpatrick and Kraft. When the matter came on for trial, the district court heard testimony from the Schempps regarding the incompatibility of certain aspects of Biblical doctrine, especially portions of the King James Version, with the religious beliefs of the Schempp family. Further, both sides presented expert witnesses.

The expert witness presented by Henry Sawyer was Dr. Solomon

---

6 Public School Act, 24 PA. CONS. STAT. ANN. § 15-1516 (West 1949).

7 The statute further provided that any schoolteacher having responsibility for reading or causing the reading of the Bible verses who "shall fail or omit to do so . . . shall, upon charges preferred for such failure or omission, and proof of the same, before the board of directors of the school district, be discharged." Id. See also infra note 12.
Grayzel, an ordained rabbi who was editor of the Jewish Publication Society. In its opinion, the district court summarized Dr. Grayzel's testimony at some length:

Dr. Solomon Grayzel testified that there were marked differences between the Jewish Holy Scriptures and the Christian Holy Bible, the most obvious of which was the absence of the New Testament in the Jewish Holy Scriptures. Dr. Grayzel testified that portions of the New Testament were offensive to Jewish tradition and that, from the standpoint of Jewish faith, the concept of Jesus Christ as the Son of God was "practically blasphemous." He cited instances in the New Testament which, assertedly, were not only sectarian in nature but tended to bring the Jews into ridicule or scorn. Dr. Grayzel gave as his expert opinion that such material from the New Testament could be explained to Jewish children in such a way as to do no harm to them. But if portions of the New Testament were read without explanation, they could be, and in his specific experience with children Dr. Grayzel observed, had been, psychologically harmful to the child and had caused a divisive force within the social media of the school.

Dr. Grayzel also testified that there was significant difference in attitude with regard to the respective Books of the Jewish and Christian Religions in that Judaism attaches no special significance to the reading of the Bible *per se* and that the Jewish Holy Scriptures are source materials to be studied. But Dr. Grayzel did state that many portions of the New, as well as of the Old, Testament contained passages of great literary and moral value.8

The expert witness for the defense was Luther A. Weigle, an ordained Lutheran minister who was Dean Emeritus of the Yale Divinity School and Chairman of the Committee for the Preparation of the Revised Standard Version of the Bible. The district court also summarized Dr. Weigle's testimony at some length:

Dr. Luther A. Weigle, an expert witness for the defense, testified in some detail as to the reasons for and the methods employed in developing the King James and the Revised Standard Versions of the Bible. On direct examination, Dr. Weigle stated that the Bible was non-sectarian. He later [presumably on cross-examination by Henry Sawyer] stated that the phrase "non-sectarian" meant to him non-sectarian within the Christian faiths. Dr. Weigle stated that his definition of the Holy Bible would include the Jewish Holy Scriptures, but also stated that the "Holy Bible" would not be complete without the New Testament. He stated that the New Testament "conveyed the message of Christians." In his opinion, reading of the Holy Scriptures to the exclusion of the New Testament

---

would be a sectarian practice. Dr. Weigle stated that the Bible was of great moral, historical and literary value. This is conceded by all the parties and is also the view of the court.9

In September 1959, the district court, speaking through Chief Judge Biggs, issued its opinion. The decision relied on the Supreme Court's 1948 holding in *McCollum v. Board of Education*, which invalidated the Champaign, Illinois "released time" program.10 Under that plan, public school children were released from classes each week to attend religious education conducted on school premises during school hours. The *Schempp* district court concluded that Pennsylvania's obligatory Bible reading program constituted an establishment of religion. The court further held that the program inhibited the free exercise of religion not only of students but also of their parents.11

Although decided by the three-judge district court in 1959, and notwithstanding the defendants' prompt filing of a notice of appeal, *Schempp* was not to be addressed by the Supreme Court on the merits until 1963. In December 1959, three months after the district court judgment, the Pennsylvania General Assembly amended section 1516. The most significant change was the insertion—after the opening sentence requiring daily readings from the Bible—of the following provision: "Any child shall be excused from such Bible reading, or attending such Bible reading, upon the written request of his parent or guardian."12 Unsurprisingly, the Supreme Court vacated the district

---

9 *Id.* at 402.
11 *Schempp* III, 177 F. Supp. at 407 ("If the faith of a child is developed inconsistently with the faith of the parent and contrary to the wishes of the parent, interference with the familial right of the parent to inculcate in the child the religion the parent desires, is clear beyond doubt."). The court also found that teachers were unconstitutionally coerced, since failure to read, or arrange for the reading of, the ten Bible verses put a teacher at risk of discharge under the 1949 Bible reading statute as originally enacted. See *id.* at 406 ("The sanction imposed upon the school teachers is discharge from their offices if they fail to observe the requirement of the statute."). But see *infra* note 12 (describing the amendment to the 1949 Bible reading statute that eliminated the threat of discharge for non-cooperating teachers).
12 In amending section 1516, the General Assembly also deleted the provision that put a non-cooperating teacher at risk of discharge. See *supra* note 6. However, as the district court was to note in the subsequent opinion reexamining the case in the light of the 1959 amendments, a teacher "who refuses or fails to obey the mandate of the amended statute may have his contract of employment terminated pursuant to 24 P.S. § 11-1122 (Supp. 1960)." *Schempp* v. School Dist. of Abington Township, 201 F. Supp. 815, 817 (E.D. Pa. 1962) [hereinafter *Schempp* II]. Section 11-1122 provided that "[t]he only valid causes for termination of a contract heretofore or hereafter entered into with a professional employee [sic] shall be immorality, incompetency, intemperance, cruelty, persistent negligence, mental derangement, advocacy of or participat-
court's judgment and remanded the case "for such further proceedings as might be appropriate in light of [the amendments to the challenged statute]."

On remand, after intermediate activity of no consequence, the district court held an additional trial. Edward Schempp—the children's father—testified about his decision not to request that the two Schempp children still in school be excused from attending Bible reading pursuant to the amendment's opt-out provision: he was concerned that his children would be "labeled as 'odd-balls.'" Further, the district court learned that:

The procedure followed in the Abington Senior High School, following the amendment of Section 1516, did differ somewhat from that which was in effect prior to the amendment. We describe it briefly. The children attending the High School, Roger and Donna included, reported to their "homerooms" at 8:15 A.M. And a few minutes thereafter the Bible reading began with each pupil seated "at attention." The Bible reading consists of reading, without comment, over a loud speaker ten verses of the King James Version of the Bible. Then the children stood and repeated, with the public address system leading them, the Lord's Prayer. Next, still standing, the children gave the Flag Salute. They then sat down. Announcements were made and when the announcements were completed the students went to their classrooms for the first classes of the day.

In February 1962, the district court, again speaking through Chief Judge Biggs, issued an opinion supplementing and reaffirming the Establishment Clause aspect of its September 1959 opinion. Said Chief Judge Biggs:

The reading of the verses, even without comment, possesses a devotional and religious character and constitutes in effect a religious observance.

14 See 195 F. Supp. 518 (E.D. Pa. 1961) (allowing plaintiffs to amend their pleading in light of the statutory amendment). To complete the procedural picture, see also 184 F. Supp. 381 (E.D. Pa. 1959), which preceded the Supreme Court's vacate-and-remand order. There, the district court held that it did not have jurisdiction to entertain the defendant's motion for relief from the final decree in view of the fact that an appeal had been taken to the Supreme Court. Id. at 383-84.

15 Schempp, 364 U.S. at 298 (1960).
The devotional and religious nature of the morning exercises is made all the more apparent by the fact that the Bible reading is followed immediately by a recital in unison by the pupils of the Lord’s Prayer. The fact that some pupils, or theoretically all pupils, might be excused from attendance at the exercises does not mitigate the obligatory nature of the ceremony for the “new” Section 1516, as did the statute prior to its 1959 amendment, unequivocally requires the exercises to be held every school day in every school in the Commonwealth. The exercises are held in the school buildings and perforce are conducted by and under the authority of the local school authorities and during school sessions. Since the statute requires the reading of the “Holy Bible,” a Christian document, the practice, as we said in our first opinion, prefers the Christian religion. The record demonstrates that it was the intention of the General Assembly of the Commonwealth of Pennsylvania to introduce a religious ceremony into the public schools of the Commonwealth.17

On the basis of the foregoing findings, Chief Judge Biggs relied on McCollum v. Board of Education,18 the Champaign, Illinois “released time” case, as he had done in the 1959 district court decision.19 In addition to reinvoking McCollum, Chief Judge Biggs quoted the Supreme Court’s assurance that “[w]e follow the McCollum case,” articulated in Zorach v. Clauson, which upheld the constitutionality of New York City’s off-the-school-premises “released time” program.20 Notably, Chief Judge Biggs made no effort to show how the Supreme Court’s approval of the New York City program in Zorach could be squared with its disapproval of the Champaign program in McCollum—or, more to the point, just how, putting aside the rhetoric in Zorach, the holding in Zorach could be squared with the district court’s determination that Pennsylvania’s Bible-reading ceremony did not pass constitutional muster.21

Stating that “[t]he Commonwealth of Pennsylvania has seen fit to breach the wall between church and state,” Chief Judge Biggs went on to “hold the statute as amended unconstitutional on the ground that it violate[d] the ‘Establishment of Religion’ clause of the First Amendment made applicable to the Commonwealth . . . by the Four-

17 Id. at 819.
18 Id. (quoting McCollum v. Board of Educ., 333 U.S. 203 (1948)).
20 Schempp II, 201 F. Supp. at 815 (quoting Zorach v. Clauson, 343 U.S. 306, 315 (1952)).
21 Distinguishing Zorach would not, in fact, have been a difficult matter. Zorach involved religious exercises conducted by non-school authorities and not on school premises. To say that Zorach was a manifestly different case does not signify that Zorach was rightly decided.
teenth Amendment. The Chief Judge found it “unnecessary to pass upon any other contention made by the plaintiffs in respect to the unconstitutionality of the statute or the practices thereunder.” In short, the district court, on weighing the impact of the amendment permitting parents to withdraw their children from the Bible-reading ceremony, (1) reaffirmed its prior holding that the ceremony was religious and hence constituted a forbidden establishment of religion, and (2) sub silentio withdrew from, but did not repudiate, its prior holding that the ceremony worked a forbidden restraint on the free exercise of religion. This was the judgment that Henry Sawyer would have to defend before the Supreme Court.

**Between the District Court and the Supreme Court**

In April 1962, two months after the three-judge district court issued its final opinion in *Schempp*, the Maryland Court of Appeals decided *Murray v. Curlett*. The plaintiffs in *Murray* were William Murray, a student in the Baltimore public schools, and his mother, Madalyn. The Murrays were atheists and had gone to a Maryland state court to seek a writ of mandamus directing the Baltimore Board of School Commissioners to rescind, as unconstitutional, Article VI, section 6 of the Board’s Rules. That section provided that “[e]ach school, either collectively or in classes, shall be opened by the reading, without comment, of a chapter in the Holy Bible and/or the use of the Lord’s Prayer.” The rule further provided that “[a]ny child shall be excused from participating in the opening exercises or from attending the opening exercises upon written request of his parent or guardian.”

The trial court granted defendants’ motion to dismiss, and the Maryland Court of Appeals affirmed. The court’s decision analogized the opening prayer exercise to permissible prayer ceremonies in the state legislature, the Congress, and the state and federal courts. The court also found persuasive the fact that “the appellant-student in this case was not compelled to participate in or attend the program he claims is offensive to him.” Next, the decision referenced *Schempp* and noted that the Supreme Court:

---

22 *Schempp II*, 201 F. Supp. at 819.
23 Id.
25 Id. at 699 (emphasis omitted).
26 Id. at 702.
ordered *per curiam* that the judgment below be vacated and remanded the case to the district court for further proceedings, after it was learned that the Pennsylvania law had been so amended as to provide for the excusing of those students who objected to participating in a school opening ceremony quite similar to that in Baltimore City.

In the court's view, "the remand of [Schempp] at least indicated that the use of coercion or the lack of it may be the controlling factor in deciding whether or not a constitutional right has been denied." Speaking to the district's court's decision on remand, the *Murray* court noted that:

> [i]n reaching this conclusion we are not unmindful that the District Court for the Eastern District of Pennsylvania has, upon the remand, re-heard the case, and again held (in an opinion by John Biggs, Jr., Circuit Judge ...) that the Pennsylvania statute is not constitutional despite the fact that objecting students could have been excused on the request of their parents, but we do not find the decision on remand persuasive and decline to follow it....

Three members of the court, speaking through Chief Judge Brune, dissented, arguing that "[t]here seems to be no substantial room for dispute that the reading of passages from the Bible and the recital of the Lord's Prayer are Christian religious exercises... favor[ing] one religion and [doing] so against other religions and against non-believers in any religion." For this reason, the dissenters concluded that the prayer recitation "is directly contra to the prohibition against any 'law respecting an establishment of religion,' contained in the First Amendment, as that provision has been interpreted by the Supreme Court." The provision allowing parents to excuse their children from participating did not, in the view of the dissenters, "save the rule from collision with the 'establishment of religion' clause of the First Amendment, even if it could save it from collision with the 'free exercise of religion' clause." This was the case because the "coercive or compulsive power of the State is exercised at least to the extent of requiring pupils to attend school and it requires affirmative action to exempt them from participation in these religious exercises." Finally, the dissenters noted that their conclusion

---

27 *Id.*
28 *Id.*
29 *Id.*
30 *Id.* at 708.
31 *Id.*
32 *Id.* at 709.
33 *Id.*
was "in accord with the result reached by a special three-judge District Court in Pennsylvania in [Schempp]."  

In June 1962, on the last day of the 1961 Term, the Supreme Court, in *Engel v. Vitale*, held invalid New York's so-called "Regents' Prayer"—"Almighty God, we acknowledge our dependence upon Thee, and we beg Thy blessings upon us, our parents, our teachers, and our Country"—promulgated by the Regents of the State of New York for daily recitation, on a non-compulsory basis, in New York's public schools. Speaking through Justice Black, the Court held that "it is no part of the business of government to compose official prayers for any group of the American people to recite as a part of a religious program carried on by government." Justice Stewart dissented.

On October 8, 1962, the Supreme Court granted certiorari in *Murray*. The Court then set the case down for argument with *Schempp*. The arguments in *Schempp* and *Murray* took place in late February 1963, just over a year after Chief Judge Biggs filed the second merits opinion in *Schempp*. The Court's decisions in the two cases were announced in a single opinion on June 17, 1963, the last day of the 1962 Term; the Maryland Court of Appeals judgment was reversed, and the Eastern District of Pennsylvania judgment was affirmed. Justice Clark wrote the opinion of the Court. Justices Douglas and Goldberg (the latter joined by Justice Harlan) filed brief concurring opinions, and Justice Brennan filed a seventy-five page concurring opinion. All of the concurring Justices joined the opinion of the Court. Justice Stewart, the lone dissenter in *Engel v. Vitale* a year before, again dissented.

The care with which Henry Sawyer had built the record in the district court was fully vindicated in the Supreme Court. The Court's opinion—following the opening paragraph that stated the central constitutional question and how it was to be resolved—proceeded to set forth "The Facts in Each Case." In stating the facts in *Schempp*, the Court set forth verbatim the district court's summary of the testimony of Dr. Grayzel, Henry Sawyer's expert witness, and of Dr. Weigle, the defense expert whose testimony Henry turned to his own use.

---

54 Id.
56 Id. at 425.
59 See supra text accompanying notes 8-9 (describing the testimony of Dr. Grayzel and Dr. Weigle before the district court).
ther, the Court also set forth verbatim the district court's findings with respect to the "devotional and religious nature of the morning exercises." The crucial weight of this factual infrastructure becomes clear when one reads the Court's statement of the governing constitutional principles and the application of those principles to the two cases before the Court. In particular, the Court pointed to eight cases directly considering the Establishment Clause. Taken together, these cases articulated the following test of legislative power respecting religious belief:

>[What are the purpose and the primary effect of the enactment? If either is the advancement or inhibition of religion then the enactment exceeds the scope of legislative power as circumscribed by the Constitution. That is to say that to withstand the strictures of the Establishment Clause there must be a secular legislative purpose and a primary effect that neither advances nor inhibits religion . . . .

Applying the Establishment Clause principles to the cases at bar we find that the States are requiring the selection and reading at the opening of the school day of verses from the Holy Bible and the recitation of the Lord's Prayer by the students in unison. These exercises are prescribed as part of the curricular activities of students who are required by law to attend school. They are held in the school buildings under the supervision and with the participation of teachers employed in those schools. None of these factors, other than compulsory school attendance, was present in the program upheld in Zorach v. Clauson. The trial court in [Schempp] has found that such an opening exercise is a religious ceremony and was intended by the State to be so. We agree with the court's finding as to the religious character of the exercises. Given that finding, the exercises and the law requiring them are in violation of the Establishment Clause.

There is no such specific finding as to the religious character of the

---


41 In addition to McCollum, Zorach, and Engel, the five other cases referred to by the Court included: Torcaso v. Watkins, 367 U.S. 488 (1961) (invalidating a state constitutional provision requiring that public officials make "a declaration of belief in the existence of God"); McGowan v. Maryland, 366 U.S. 420 (1961) (sustaining "Sunday closing" law); Everson v. Board of Education, 330 U.S. 1 (1947) (sustaining government reimbursement of parents for the cost of transportation to parochial schools); Mundock v. Pennsylvania, 319 U.S. 105 (1943) (town ordinance requiring license and payment of fee to engage in solicitation invalid as applied to door-to-door solicitation by Jehovah's Witnesses); and Cantwell v. Connecticut, 310 U.S. 296 (1940) (invalidating a statute forbidding public solicitation on behalf of a religious cause without a public official's approval).
exercises in [Murray], and the State contends (as does the State in [Schempp]) that the program is an effort to extend its benefits to all public school children without regard to their religious belief. Included within its secular purposes, it says, are the promotion of moral values, the contradiction to the materialistic trends of our times, the perpetuation of our institutions and the teaching of literature. The case came up on demurrer, of course, to a petition which alleged that the uniform practice under the rule had been to read from the King James version of the Bible and that the exercise was sectarian. The short answer, therefore, is that the religious character of the exercise was admitted by the State.

The dual Establishment Clause principles announced by the Supreme Court in Schempp—"[T]o withstand the strictures of the Establishment Clause there must be a secular legislative purpose and a primary effect that neither advances nor inhibits religion"—were to be the doctrinal building blocks of the triad of Establishment Clause principles announced in Lemon v. Kurtzmann, Henry Sawyer's second religion case.43

**LEMON V. KURTZMAN**

*Lemon v. Kurtzman*, Henry Sawyer's second campaign to enforce the First Amendment guarantees of religious freedom, had as its target Pennsylvania's Nonpublic Elementary and Secondary Education Act—a statute enacted by the Pennsylvania general Assembly in 1968 with a view to alleviating the rapidly escalating financial burdens of Pennsylvania's nonpublic schools.44 The statute authorized Pennsylvania's Superintendent of Public Instruction to make a "contract" with a nonpublic school under which the school would agree to teach one or more "secular" subjects—mathematics, modern foreign languages, physical science and physical education. The state would reimburse the school for the costs of instruction, most particularly teacher's salaries, books, and instructional equipment. The books and equipment utilized in such a course were to be approved by the Superintendent of Public Instruction, and state reimbursement could not be authorized for "any subject matter expressing religious teaching, or the morals or forms of worship of any sect."45

42 Schempp, 374 U.S. at 222-24.
43 For the author's views of Schempp and Engel at the time of the Schempp decision, see Louis H. Pollak, Foreword: Public Prayers in Public Schools, 77 HARV. L. REV. 62 (1963).
45 Id. at 39-40.
When the statute went into effect Pennsylvania spent approximately $5,000,000 annually to reimburse nonpublic schools pursuant to these "contracts." The eleven hundred beneficiary schools had educational responsibility for upwards of 500,000 pupils—twenty percent of Pennsylvania's school population. More than ninety-six percent of these pupils were enrolled in church-related schools. Those who challenged the statute saw it as a device for funneling public money to nonpublic—and, particularly, church-related—schools by paying them for doing what they had traditionally done without depending on public largesse.

_In the District Court_

As he had done in _Schempp_, Henry Sawyer brought his suit in the District Court for the Eastern District of Pennsylvania. The suit—filed in 1969—sought to enjoin further expenditures of state funds for this allegedly unconstitutional enterprise.\(^46\) Once again, a three-judge district court was convened. The panel consisted of District Judges Luongo and Troutman and Chief Circuit Judge Hastie.

The defendants moved to dismiss. On November 28, 1969, the district court filed an opinion granting the motion.\(^47\) Judge Troutman wrote for the court, stating that the doctrinal principles governing the case were those announced by the Supreme Court in _Schempp_—i.e., that to withstand the strictures of the Establishment Clause, there must be "a secular legislative purpose and a primary effect that neither advances nor inhibits religion."\(^48\) Judge Troutman noted that "[t]he plaintiffs allege as a fact that the purpose and primary effect of the Education Act is to aid religion."\(^49\) However, Judge Troutman determined "that the allegation asserts not a fact but a conclusion of law and as such is not admitted for the purposes of testing the sufficiency of the complaint."\(^50\) He then held that "the purpose of the Education Act can be found clearly on its face."\(^51\) Further, "[t]he Legislature has

\(^{46}\) See id. The plaintiffs included individuals suing as taxpayers and also the following organizations: Americans United for Separation of Church and State; the Pennsylvania Council of Churches; the Pennsylvania Jewish Community Relations Conference; the Pennsylvania Conference of the NAACP; and the ACLU. The district court determined that the organizations did not have standing. _Id._ at 41.

\(^{47}\) _Id._ at 49.

\(^{48}\) _Id._ at 44 (quoting _Schempp_, 374 U.S. 203, 222 (1963)).

\(^{49}\) _Id._ at 43.

\(^{50}\) _Id._

\(^{51}\) _Id._ at 45.
declared that the purpose of the Education Act is 'to promote the welfare of the people of the Commonwealth of Pennsylvania' and 'to promote the secular education of children of the Commonwealth of Pennsylvania attending nonpublic schools.'

Chief Judge Hastie dissented, stating that:

the majority seems to view the question of the purpose and effect of the statute as foreclosed by declarations in the statute itself that the legislative purpose is "to promote the welfare of the people of the Commonwealth" and "to promote the secular education of children of the Commonwealth of Pennsylvania attending nonpublic schools." With this I cannot agree.

Then—after pointing out that the issue before the court was not "summary judgment where the factual posture of the case is established by affidavits and exhibits" but a motion to dismiss with respect to which "decision is controlled by the allegations of the complaint and our judgment as to the potentiality of proof thereunder"—Chief Judge Hastie continued:

But even if inquiry as to purpose and effect should be confined to examination of the language and scheme of the statute, I cannot avoid the conclusion that the primary purpose and effect of the enactment is to help the nonpublic schools by supplying them with needed financial aid, while whatever promotion of the public welfare is anticipated as a result of such public assistance is at best an incidental consequence claimed in justification of the state's action.

In the balance of his opinion, Chief Judge Hastie addressed the larger implications of the case:

It has already been pointed out that sectarian schools are only part of a complex of activities, many of them as 'secular' as the teaching of languages and physical science, which modern churches and religious institutions finance and conduct. Charities, hospitals, community centers and homes for the aged and infirm are familiar examples.... If the constitutional bar to state grants in all such cases should be removed, it is reasonable to anticipate continuing political controversy in every state and local community whether and to what extent public funds are to be granted to subsidize a large number and a broad range of activities of religious organizations....

So far we have escaped much of the divisiveness and antagonism of

---

52 Id.
53 Id. at 49-50.
54 Id. at 50 n.1.
55 Id. at 50.
political differences and controversies about religious matters because public financing of activities of religious organizations has been understood to be prohibited by our Constitution. Professor Paul Freund has perceptively pointed out that President Kennedy was able to avoid taking a political position upon issues of religious character by relying upon authoritative decisions on the constitutional separation of state and religion as controlling. [Paul A. Freund, Public Aid to Parochial Schools, 82 Harv. L. Rev. 1680, 1692 (1969)]. But if the present statute is held to be constitutional, I see no escape from the evils that attend a widespread and pervasive intermingling of politics and religion.56

Henry Sawyer filed a notice of appeal, and on April 20, 1970, the Supreme Court noted probable jurisdiction.57

On June 15, 1970, a three-judge district court in Rhode Island held invalid the Salary Supplement Act enacted by the Rhode Island Legislature in 1969.58 Under that statute Rhode Island undertook to pay to teachers of secular subjects in nonpublic elementary schools salary supplements of up to fifteen percent of a teacher's salary. The statute's stated purpose was "to assist non-public schools to provide salary scales which will enable them to retain and obtain teaching personnel who meet recognized standards of quality."59 The court conducted a trial, finding that the "evidence... fully corroborates the legislature's finding of a financial crisis in non-public education, but indicates that the crisis is largely confined to Rhode Island's Catholic schools" in which "[a]pproximately 95 per cent of the elementary school children attending non-public schools are enrolled..."60

The financial problems besetting the elementary parochial schools were found to have originated in the need to recruit hundreds of lay teachers to supplement the nuns who historically had constituted more than ninety percent of the instructional staff. The efforts of the parochial schools to recruit lay teachers, however, were made more difficult by virtue of the steadily rising salaries of public school teachers. The court determined that "[t]he Salary Supplement Act will not relieve the parishes or parents of their escalating burden, but will temporarily enable parochial schools to compete for qualified teachers."61 The court found that "[o]n the one hand, it aids the qual-

56 Id. at 51.
59 Id. at 114
60 Id. at 115.
61 Id.
ity of secular education; on the other, it provides support to a religious enterprise."\(^6\) In sum, the court saw "as the necessary effects of the kind of legislation involved here not only substantial support for a religious enterprise, but also the kind of reciprocal embroilments of government and religion which the First Amendment was meant to avoid."\(^6\) The opinion of the court was written by Circuit Judge Coffin and was joined by then-District Judge Bownes. Judge Pettine filed a separate concurring opinion. On November 9, 1970, the Supreme Court noted probable jurisdiction in the Rhode Island case,\(^6\) which was then set down for argument with *Lemon v. Kurtzman*.

**In the Supreme Court**

Argument took place on March 3, 1971, and the Court issued its decision on June 28, 1971—the last day of the 1970 Term.\(^6\) The Court reversed the judgment of the Pennsylvania three-judge district court and affirmed the judgment of the Rhode Island three-judge district court. As was true in *Schempp* and *Murray*, the disposition of the trial court that had dismissed the complaint for failure to state a cognizable cause of action was overturned, while the disposition of the trial court that had conducted a trial was sustained.

Chief Justice Burger wrote the opinion of the Court,\(^6\) and built upon the groundwork laid in *Schempp*.

Every analysis in this area must begin with consideration of the cumulative criteria developed by the Court over many years. Three such tests may be gleaned from our cases. First, the statute must have a secular leg-

---

\(^{62}\) Id. at 119.  
\(^{63}\) Id. at 122.  
\(^{66}\) Joining the opinion in its entirety were Justices Black, Douglas, Stewart and Blackmun. Justice Marshall joined the Rhode Island portion of the Chief Justice's opinion, but took no part in *Lemon*; Justice Marshall also filed a brief separate statement. Justice Douglas filed a concurring opinion in which Justice Black joined, and in which Justice Marshall substantially joined (except insofar as the opinion addressed *Lemon*). Justice Brennan filed a separate concurring opinion. Justice White filed an opinion concurring in *Lemon* and dissenting in the Rhode Island case. *Id.* at 604.

It seems a reasonable surmise that Justice Marshall recused himself in *Lemon* for the reason that the Pennsylvania Conference of the NAACP was one of the initial plaintiffs (albeit dismissed for lack of standing, along with the other organizational plaintiffs, by the district court). It is understood to have been Justice Marshall's settled practice not to participate in cases in which the NAACP was a party, in view of the Justice's long and triumphant association with the NAACP (and its affiliate, the NAACP Legal Defense and Educational fund) in his lawyering days.
islative purpose; second, its principal or primary effect must be one that neither advances nor inhibits religion, Board of Education v. Allen, 392 U.S. 236, 243 (1968) [quoting Schempp]; finally, the statute must not foster "an excessive government entanglement with religion." 67

Because the Rhode Island case had gone to trial, and hence had generated a significant factual record, the opinion of the Chief Justice had considerably more to say about Rhode Island's salary supplement program than about Pennsylvania's reimbursement program. The Chief Justice's discussion of the Pennsylvania program follows:

The Pennsylvania statute . . . provides state aid to church-related schools for teachers' salaries. The complaint describes an educational system that is very similar to the one existing in Rhode Island. According to the allegations, the church-related elementary and secondary schools are controlled by religious organizations, have the purpose of propagating and promoting a particular religion's faith, and conduct their operations to fulfill that purpose. Since this complaint was dismissed for failure to state a claim for relief, we must accept these allegations as true for purposes of our review.

As we noted earlier, the very restrictions and surveillance necessary to ensure that teachers play a strictly nonideological role give rise to entanglements between church and state. The Pennsylvania statute, like that of Rhode Island, fosters this kind of relationship. Reimbursement is not only limited to courses offered in the public schools and materials approved by state officials, but the statute excludes "any subject matter expressing religious teaching, or the morals or forms of worship of any sect." In addition, schools seeking reimbursement must maintain accounting procedures that require the State to establish the cost of the secular as distinguished from the religious instruction.

The Pennsylvania statute, moreover, has the further defect of providing state financial aid directly to the church-related school. This factor distinguishes both Everson and Allen, for in both those cases the Court was careful to point out that state aid was provided to the student and his parents—not to the church-related school. Board of Education v. Allen, supra at 243-244; Everson v. Board of Education, supra, at 18. In Walz v. Tax Commission, supra at 675, the Court warned of the dangers of direct payments to religious organizations:

Obviously a direct money subsidy would be a relationship pregnant with involvement and, as with most governmental grant programs, could encompass sustained and detailed administrative relationships for enforcement of statutory or administrative standards.

67 Id. at 612-13 (citation omitted).
The history of government grants of a continuing cash subsidy indicates that such programs have almost always been accompanied by varying measures of control and surveillance. The government cash grants before us now provide no basis for predicting that comprehensive measures of surveillance and controls will not follow. In particular the government's post-audit power to inspect and evaluate a church-related school's financial records and to determine which expenditures are religious and which are secular creates an intimate and continuing relationship between church and state.\(^{68}\)

It is clear that, although the Pennsylvania district court did not conduct a trial, the Chief Justice gave careful attention to the limited proceedings that did take place. Thus, the Chief Justice specifically adverted to the fact that Chief Judge Hastie had dissented.\(^{69}\) Moreover, in emphasizing the entanglement risks "presented by the divisive political potential of these state programs,"\(^{70}\) the Chief Justice cited (albeit without attribution to Chief Judge Hastie) the same page of the same article by Professor Freund that the Chief Judge had invoked. The Chief Justice put the matter succinctly:

Ordinarily political debate and division, however vigorous or even partisan, are normal and healthy manifestations of our democratic system of government, but political division along religious lines was one of the principal evils against which the First Amendment was intended to protect.\(^{71}\)

*Lemon v. Kurtzman* has been the law of the land for almost thirty years. The attacks on its three-fold “test” have been numerous. But the constitutional concerns animating the decision are no less valid today than when Chief Justice Burger (and Chief Judge Hastie) gave voice to them.

**CONCLUSION**

In 1988, having heard that Henry Sawyer was in the process of withdrawing from full-time active practice, a lawyer who had occasion over the years to observe Henry as an appellate advocate wrote him a letter. The letter-writing lawyer was William J. Brennan, Jr., the Senior Associate Justice, who had been a member of the Court since 1956—five years before Henry argued *Deutch*. The letter said: “Since I’ve

\(^{68}\) *Id.* at 620-22.

\(^{69}\) *Id.* at 611.

\(^{70}\) *Id.* at 622.

\(^{71}\) *Id.*
been here, few lawyers have equaled your advocacy."72 Praise from Sir Hubert is praise indeed.

72 Letter from William J. Brennan, Jr. to Henry Sawyer, III.
ARTICLES

THREATENING INEFFICIENT PERFORMANCE OF INJUNCTIONS AND CONTRACTS

IAN AYRES
KRISTIN MADISON†

INTRODUCTION ................................................................................... 46

I. A UNIFYING THEORY OF PERFORMANCE THREATS ...................... 58
   A. The Motive to Threaten Inefficient Performance ...................... 59
   B. When Will a Potential Defendant’s Performance Threat Be Credible? ........................................................................ 60
   C. When Will a Plaintiff’s Performance Threat Be Credible? .......... 64
   D. When Do Performance Threats Cause Payoffs to Diverge From Make-Whole Compensation? ........................................................... 68

II. HARNESSING PRIVATE INFORMATION TO DISCREDIT PERFORMANCE THREATS .......................................................... 71
   A. The Inalienable Injunction Option ........................................... 72
   B. The Defendant’s Damage Adjustment Option—Private Additur...... 79
   C. The Plaintiff’s Damage Adjustment Option—Private Remittitur .... 82

III. SHOULD THREATS OF INEFFICIENT PERFORMANCE BE DETERRED? 84
   A. Equity .................................................................................... 84
   B. Efficiency ............................................................................ 87

† Ian Ayres, William K. Townsend Professor, Yale Law School (ian.ayres@yale.edu); and Kristin Madison, J.D. Candidate, 2000, Yale Law School and Ph.D Candidate in Economics, 2001, Stanford University (kristin.madison@yale.edu). Barry Adler, Jonathan Baker, Lucian Bebchuk, Jennifer Brown, Adrienne Davis, Jill Fisch, Barbara Fried, Marcel Kahan, Candace Kovacic-Fleischer, Steve Parr, Ricki Revesz, Raphael Thomad- sen, Peter Siegelman, Steven Thel, Jon Rork, Akila Weerapan, and seminar partici- pants at American University, NYU, Stanford, and University of Tel Aviv law schools provided helpful comments. This material is based in part upon work supported un- der a National Science Foundation Graduate Fellowship received by Kristin Madison. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Sci- ence Foundation.

(45)
INTRODUCTION

Threats are often conditional promises to act inefficiently. The threatener in effect says: "I will do something that hurts you more than it helps me unless you pay me not to." Threatening inefficient action often in turn produces inefficiency because either the threatener follows through on her threat, resources are squandered in negotiating to avoid the threatened behavior, or the contracting parties take overly cautious steps to avoid being threatened. Contract scholars have long understood that this problem might arise when promisors threaten to breach. If contract damages are not sufficient to fully compensate a promisee for lack of performance, a promisor may threaten to breach in order to extract more favorable terms. For example, a seller may threaten to breach a supply agreement—even

---

1 Under this definition, threats are made to induce payment. See Robert Hale, Coercion and Distribution in a Supposedly Non-Coercive State, 38 Pol. Sci. Q. 470, 476 (1923) ("If I plan to do an act or to leave something undone for no other purpose than to induce payment, that might be conceded to be a 'threat.'"). In contrast, a promise to act efficiently is not a threat because the promisor does not seek payment and the promisee would not pay enough to stop the promisor from acting. Thus, the answer to the age-old tough-guy question: "Is that a threat or a promise?" may turn on whether the threatener/promisor seeks to change another person's behavior. Other characteristics of threatening behavior are explored in Ian Ayres & Barry J. Nalebuff, Common Knowledge as a Barrier to Negotiation, 44 UCLA L. Rev. 1631 (1997).

2 See, e.g., Saul Levmore, Strategic Delays and Fiduciary Duties, 74 Va. L. Rev. 863, 870 (1988) ("Perhaps the most convincing argument for discouraging delay when the defendant's behavior is not unambiguously wrongful is that such delay threatens to deter desirable behavior by potential defendants."); Mary Lou Serafine, Note, Repudiated Compromise After Breach, 100 Yale L.J. 2229, 2229 (1991) ("The repudiated compromise arises when one party to a contract threatens to or does actually breach some term of a contract and, rather than take the problem to court, the parties agree to a compromise."). Threatening inefficient breach to negotiate a more favorable price is vividly illustrated by Austin Instrument, Inc. v. Loral Corp., 272 N.E.2d 533, 534 (N.Y. 1971) which concerned a subcontractor's "threat" to stop deliveries unless prices were increased.
when it is clear that performance is efficient—solely to renegotiate a higher price. Because such renegotiations are often thought to be presumptively inefficient, the rules invalidating bad faith or opportunistic renegotiation attempt to deter promisors from making the initial threat.

A parallel problem has gone virtually unnoticed: threatening to perform. In this article, we will present two broad contexts where parties threaten inefficient performance of contractual promises or other legal duties solely to gain bargaining power in a subsequent negotiation:

1. A potential plaintiff who is owed a duty may, at times, seek inefficient injunctive relief instead of damages merely to induce a defendant (the person owing the duty) to pay an amount higher than expected court-awarded damages.

2. And, more perversely, a potential defendant who owes a duty to another may, at times, threaten to perform an inefficient duty merely to induce the plaintiff (the person owed the duty) to accept an amount less than expected court-awarded damages.

When a performance of some duty becomes inefficient (in the straightforward sense that the cost of performance is greater than the benefit), we will show that one side often will desire to threaten performance merely to gain bargaining power. The impulse to threaten inefficient performance does not connote, however, an ability to make credible threats. We will discuss conditions under which such threats are credible, giving rise not only to substantial negotiation costs but also to bargaining outcomes that diverge substantially from make-whole damages.

In the contractual context, promisees at times will inefficiently seek specific performance not because they value actual performance more than damages, but because they want to sell their court-ordered right to performance back to the promisor. These promisees represent to the court that monetary damages would be insufficient to make them whole and then—before the ink dries on the injunction—offer a price to relieve the promisor of the court-ordered duty. Judge Posner foresaw just this possibility in declining to award an injunction to a coal seller:
With continued production uneconomical, it is unlikely that an order of specific performance, if made, would ever actually be implemented. . . . [B]y offering [the seller] more than contract damages . . . [the buyer] could induce [the seller] to discharge the contract and release [the buyer] to buy cheaper coal . . . . Probably, therefore, [the seller] is seeking specific performance in order to have bargaining leverage with [the buyer], and we can think of no reason why the law should give it such leverage.3

When promisors seek to breach what have become inefficient promises, promisees may seek specific performance merely to induce the promisors to pay more than expectation (make-whole) damages.

To understand the incentives promisees have to seek inefficient specific performance, consider a stylized variation on the facts of *Peeryhouse*.4 A miner has promised to return the topsoil on a farmer's strip-mined land to its original position. Imagine that the cost of moving the topsoil turns out to be $30,000, but that the court is expected to award only diminution-in-value damages of $10,000 if the miner fails to perform. If we also assume that the farmer's actual benefit from performance (moving the soil) is $8000, the farmer has a strategic rationale for seeking specific performance of the contract.5 If the court awards specific performance, then the Nash bargaining solution6 is for the miner to pay the farmer $19,000 to avoid moving the soil.7 Even though $10,000 in damages provides more compensation than $8000 in make-whole relief, the farmer has an incentive to seek

---

5 Judith Maute has suggested that the plaintiff's attorney may have elected not to seek specific performance in order to increase the size of his contingent fee. See Judith L. Maute, *Peeryhouse* v. Garland Coal & Mining Co. *Revisited: The Ballad of Willie and Lucille*, 89 Nw. U. L. Rev. 1341, 1449-50 (1995).
6 The bargaining solution is the amount at which the bargaining parties ultimately settle. The settlement amount could conceivably be anywhere within the parties' bargaining range. However, in this Article we have chosen to use the Nash bargaining solution, which maximizes the product of the parties' bargaining gains. See, e.g., FRANK STÄHLER, ECONOMIC GAMES AND STRATEGIC BEHAVIOUR: THEORY AND APPLICATION 40-41 (1998) (describing the role of John Nash in the development of economic theory and the basic principles of his bargaining solution).
7 \((8000 + 30,000)/2\). An agreement to avoid performing the injunction creates a total surplus of $22,000. The Nash solution splits this surplus between the two parties. The defendant pays the plaintiff $11,000 less than she would have spent had she performed—the difference between the $30,000 cost of performance and the $19,000 payment to the plaintiff under the Nash solution. The plaintiff receives $11,000 more than she would have received in the event of performance (the difference between the $19,000 payment and the $8000 benefit from performance).
an inefficient injunction in order to increase his bargaining power. Scholars have mistakenly argued that "an injured party would not choose specific performance unless damages undercompensated the party." This simple example, however, shows that plaintiff/promisees may choose specific performance even when damages would overcompensate them. Although the equal division of the bargaining surplus implied by the Nash bargaining solution (which implicitly assumes equal bargaining power) may not apply to particular contexts, under a variety of alternative bargaining-power assumptions, the farmer/promisee will threaten inefficient performance as a bargaining chip.

Plaintiffs also seek inefficient injunctions outside of contractual settings. Indeed, the incentive to seek inefficient injunctions solely for settlement value is a possibility whenever the law gives aggrieved parties the option to seek an injunction instead of monetary damages. For example, consider the classic 1895 encroachment case of...

---


9 See Ian Ayres & Peter Siegelman, (Econ)stitutional Law: Standing, Harm, and Revealed Preference (Aug. 29, 1999) (unpublished manuscript, on file with the University of Pennsylvania Law Review) for examples of several threats to seek injunctions in contexts outside of contracting. As the authors point out, a company may bring an antitrust action seeking an injunction to block a merger as being anticompetitive. The company seeks not to redress its own injury, but instead to capture a share of the gains the merging parties will achieve by selling back the injunction. See id. at 35 n.71. See also Joseph F. Brodley, Antitrust Standing in Private Merger Cases: Reconciling Private Incentives and Public Enforcement Goals, 94 MICH. L. REV. 1, 23 (1995), which states:

Frank Easterbrook and Daniel Fischel have taken the position that merger injunction actions are inherently inferior to damage actions. They argue that injunction actions create acute holdup problems because each plaintiff can threaten to block the acquisition unless paid the merger's full transactional value, a sum likely to exceed any threatened injury to the plaintiff.

Id. (citing Frank H. Easterbrook & Daniel R. Fischel, Antitrust Suits by Targets of Tender Offers, 80 MICH. L. REV. 1155, 1169 (1982)).

Even constitutional claims may be used to extract a settlement. In designing an auction of Personal Communications Services frequencies, the Federal Communications Commission ("FCC") initially proposed rules giving bidding preferences to firms owned by women and/or minorities and to small businesses in the auctions for certain licenses. Telephone Electronics Corporation ("TEC"), whose $200 million in annual revenues made it too large to qualify for bidding preferences as a small business, challenged the constitutionality of race and gender preferences. TEC's president was clear about why he had sought the injunction against the auctions: once it blocked the auctions, TEC could offer to drop its suit in exchange for an agreement with the FCC to grant it the bidding preference for which it had been ineligible. The FCC was about to grant TEC an exemption when TEC formed a joint venture with several larger firms and dropped its suit. See Ayres & Siegelman, supra, at 36-38.
Pile v. Pedrick. After being misinformed by a surveyor, Pedrick built a factory wall with a foundation that extended $1 \frac{3}{8}$ inches onto Pile's land (below the surface of the land). The court offered Pile a choice of either damages for the permanent trespass or a court order to remove the wall. Pile insisted upon the latter.

Imagine that Pedrick's cost of removing the wall was $10,000, but the court's estimate of permanent trespass damages was only $500. If we also assume that Pile's actual benefit from performance (removing the wall) was $0, we can see Pile had a strategic rationale for seeking an injunction—even if Pile knew that tearing down the wall was inefficient. The Nash bargaining solution in the shadow of an injunction was for Pedrick to pay Pile $5000.11 Even though performance was inefficient and damages provided more than make-whole relief, Pile had an incentive to seek an injunction in order to increase his bargaining power. As we will show more formally below, the incentive to seek inefficient injunctions is particularly strong when the likely court-awarded damages are substantially lower than the cost of performance.

Lest our gentle reader think that the incentive of plaintiffs to seek inefficient injunctions is merely another perverse, but other-worldly, implication of game theory, consider the two common law chestnuts of Edwards v. Allouez Mining Co.12 and Rievman v. Burlington Northern Railroad Co.13

In Edwards, the defendant, in 1874, "at a cost of some sixty thousand dollars erected a stamp mill on the banks of Hill creek."14 The operation of the mill necessitated depositing large quantities of sand on the bottom lands below. As Justice Cooley summarized:

The year following the erection of defendant's mill, complainant purchased a piece of land through which the creek runs a short distance below the mill, and upon which the mill as operated was depositing sand. The land was not purchased for use or occupation, but as a matter of speculation, and apparently under an expectation of being able to force defendant to buy it at a large advance on the purchase price. It was offered to defendant soon after the purchase, and though no price was named, the valuation which has been put upon it by complainant and his witnesses is from three to five times what it cost him, and this perhaps gives some in-

---

10 31 A. 646 (Pa. 1895).
11 $(10,000 + 0)/2$.
12 38 Mich. 46 (1878).
dication what his expectations were.\textsuperscript{15}

As Edward Yorio has noted: "[T]he peculiar facts of Edwards dramatize how equitable remedies may be used to extort overcompensatory settlements."\textsuperscript{16} Edwards represents a strategic "coming to the nuisance" in order to extort a supercompensatory payment.

For a more contemporary example, consider the facts of Riemann. In this case, bonds issued in 1896 were secured with realty that by 1985 was worth billions of dollars more than the outstanding principal of the bonds. The terms of the bond mortgages, however, severely inhibited the sale and development of the realty. A class of bondholders brought suit "to enjoin the [defendant] Railroad from substituting other collateral for [the realty] by which the bond mortgages [were] secured."\textsuperscript{17} The court expressly endorsed the bondholders' right to "hold up" the defendant for an immediate "premium" payment of $35.5 million (in addition to providing substitute collateral that virtually eliminated any chance of default) by threatening specific enforcement of the collateral provisions.\textsuperscript{18} Examples abound in which the plaintiffs seek inefficient injunctive relief in order to extract a premium above the value of actual performance.\textsuperscript{19}

\textsuperscript{15} Id. (emphasis added).
\textsuperscript{16} \textsc{Edward Yorio, Contract Enforcement: Specific Performance and Injunctions} 85 (1989).
\textsuperscript{17} Riemann, 118 F.R.D. at 30.
\textsuperscript{18} See id. at 33 ("The bondholders' lien . . . permits [bondholders] to insist on receiving the 'hold-up' premium to which that lien has given rise."). The court, however, also refused to allow a minority of bondholders to extract even more money from the defendant by objecting to the $35 million premium. \textit{Id}
\textsuperscript{19} See, e.g., Northern Ind. Pub. Serv. Co. v. Carbon County Coal Co., 799 F.2d 265, 279 (7th Cir. 1986) ("Probably, therefore, [the seller] is seeking specific performance in order to have bargaining leverage with [the buyer] . . . ."); Foster v. American Mach. & Foundry Co., 492 F.2d 1317, 1324 (2d Cir. 1974) (stating that an injunction prohibiting infringement "is not intended as a club to be wielded by a patentee to enhance his negotiating stance"); Bracwell v. Appleby, [1975] Ch. 408, 416 (refusing to grant an injunction that would prohibit the defendant from occupying his own house because an in part injunction would place plaintiffs in "an unassailable bargaining position"); \textit{See also} \textsc{Yorio, supra} note 16, at 83 ("The availability of specific performance or injunctive relief gives the plaintiff considerable leverage in negotiations between the parties.").

Defendants against whom injunctions issue at times pay an amount in settlement instead of performing. For example, in \textit{Tulk v. Moxhay}, Tulk sued to prevent Moxhay from building on Leicester Square garden. 41 Eng. Rep. 1143 (Ch. 1848). Since Tulk's deed had a covenant requiring that the Square be "uncovered with any buildings," the court granted Tulk the injunction. \textit{See id.} at 1143. The Tulk family ultimately traded its injunctive right in exchange for a valuable option. \textit{See} \textsc{Jesse Dukeminier \& James E. Krier, Property} 863 (3d ed. 1993) (explaining the subsequent history of \textit{Tulk v. Moxhay}).
While these previous examples concern how people *who are owed* a duty can have an incentive to threaten inefficient performance of injunctions, it is also possible that people *who owe* a duty will threaten inefficient performance merely to increase their bargaining power. People who owe performance of a duty are likely to threaten inefficient performance whenever expected damages equal or exceed the cost of performance. This can be seen in a contractual setting when a court is expected to award cost of performance damages.

Whenever expected court-awarded damages exceed the promisee’s benefit from performance, a promisor’s threat to perform may induce the promisee to settle a case for less than the court award. Returning again to a stylized variation of *Peevyhouse*, imagine now a jurisdiction that would award $30,000 as cost of performance damages if the miner breached its promise (to restore the topsoil of the strip-mined land), even though the farmer’s value of actual performance is only $8000. Before performance is due, the promisor might try bargaining her way out of performing by offering to pay the promisee some amount between $8000 and $30,000, say by splitting the difference at $19,000—again, the Nash bargaining solution.

The naive promisee at this point might respond: “Why should I accept $19,000 when my contract damages will be $30,000?” A savvy miner, however, will answer: “You’re mistakenly assuming that I will breach if we don’t reach agreement. If we don’t renegotiate this contract, you should know that I am going to perform and you will end up with only an $8000 benefit.”

The promisor gains bargaining power by threatening to perform. Even though the promisor knows performance is inefficient, threatening performance can be an individually rational strategy because it

In a recent, excellent article, Ward Farnsworth found that in twenty recent nuisance cases in which injunctions were issued, the parties failed to negotiate whether a payment would be made in lieu of performance. Ward Farnsworth, *Do Parties to Nuisance Cases Bargain After Judgment? A Glimpse Inside the Cathedral*, 66 U. Chi. L. Rev. 373, 381-83 (1999). Farnsworth argues that acrimony and distaste for bargaining may have deterred such negotiations. *See id.* at 384. If the Farnsworth result were generally true, then one might reasonably doubt whether plaintiffs ever seek injunctions in order to extract supercompensatory payments from the defendant. Besides the aforementioned counterexamples such as *Edwards*, there are strong reasons to question whether one can make a generalization from these twenty observations. Farnsworth only examines appellate decisions. As he acknowledges, the litigants that fail to settle by this point in the litigation might be strongly predisposed toward acrimony and distaste for bargaining. *See id.* And even to the extent that one can generalize the result, our proposed reforms, by more explicitly stating when an injunction is alienable, may encourage negotiations among litigants who may have otherwise failed to consider negotiation.
disproportionately hurts the promisee. Moreover, under the assumed facts, the threat is credible because in the absence of agreement actual performance would cost the promisor no more than he would have to pay in damages—indeed, actual performance potentially saves attorney fees.

Inefficient performance threats can produce both ex ante and ex post types of inefficiency. Ex post, a threat of inefficient performance might entail the cost of the negotiations themselves, and the costs resulting when the failure to reach a bargain results in inefficient performance. As Richard Epstein has noted: "Injunctive relief thus poses two major risks: first, that parties will waste enormous resources in bargaining over the surplus and, second, that they will not be able to reach any agreement at all given the tendency to bluff and bluster." More subtly, the prospect of being threatened ex post may distort ex ante behavior. Thus, in the contract setting, the prospect of both inefficient threats and payoffs that substantially diverge from make-whole damages may adversely affect the parties' original willingness to contract or to rely on the contract.

Inefficient threats are commonplace in bargaining. Jones may resist selling her car to Smith, even though she knows that Smith values it more, merely to induce a higher purchase price. The inefficiency produced by such quotidian threats, however, is greatly reduced by competition. Smith can look for other sellers. The performance threats analyzed in this article are distinguishable because they arise under conditions of bilateral monopoly. Once a court orders Pedrick to remove the encroaching wall, Pile is the only person to whom Pedrick can look to avoid the performance inefficiency. Threatening inefficient behavior to induce more favorable contract terms is more worrisome when the parties do not have outside, competitive options. The common law at times, such as in the case of salvage, constrains the parties' ability in bilateral monopoly contexts to make inefficient threats as a way of limiting the potential inefficiencies and inequities of such negotiations.

---


22 Pedrick is the only person Pile can hold up.

23 See infra notes 24-25 (suggesting legal rules which would reduce the parties' use of inefficient threats).
Under conditions of bilateral monopoly, it is presumptively inefficient for either promisor or promisee to threaten inefficient performance merely to gain renegotiation bargaining power. The law might usefully be structured to deter such threats—even though it does not intervene to regulate negotiations where the bargainers have outside options. Courts, however, will often have difficulty distinguishing efficient from inefficient threats. For example, promisees might seek specific performance because they place a high subjective value on performance rather than a desire to sell their injunction back to the promisor. Our solution is to suggest legal rules that economize on the parties’ private information, particularly on the threatened party’s knowledge.

We will argue that the law can give the threatened party options to make inefficient threats less attractive. In particular, we assess the efficiency effects of two remedial reforms that undermine the credibility of inefficient threats:

**Inalienable Injunction Option:** Before asking plaintiffs to elect monetary or injunctive relief, courts could routinely give defendants an option to make any injunctive relief inalienable. Threatened parties would have the option of forgoing their ability to buy their way out of an injunction. By exercising this option, a threatened promisor in effect would be telling the promisee: “Force my performance if you really want my performance, but don’t seek an injunction merely to hold me up for money.” Under the facts of *Pile*, making the injunction inalienable might have deterred the plaintiff from seeking the inefficient injunction to have the encroaching wall removed. The plaintiff might have preferred monetary damages (of $500) to an inalienable injunction that would have provided the plaintiff only a negligible benefit.

**Private Additur and Remittitur:** Courts could routinely give the

---

24 Louis Kaplow and Steven Shavell have shown how liability rules can economize on private information. See Louis Kaplow & Steven Shavell, *Property Rules Versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 726-27 (1996) (“[U]nder the liability rule, the state is able to make implicit use of injurers’ information about prevention costs, because injurers know their actual prevention cost, which they compare to average harm.”); see also Ian Ayres & J.M. Balkin, *Legal Entitlements as Auctions: Property Rules, Liability Rules, and Beyond*, 106 YALE L.J. 703, 749 (1996) (discussing the advantages of higher order liability rules, including the fact that they harness information efficiently). Our theory suggests that when one side has the option of choosing a property rule negotiation, the other side should be given an offsetting option.
threatened party the option to commit to "less favorable" monetary damages. Before performance is due, promisees could be given the option to decrease the potential legal damages they would receive in the event of breach. Likewise, before promisees commit to injunctive relief, promisors could be given the option to increase the potential legal damages they would pay in the event of breach. This would amount to a system of private additur and remittitur, in which plaintiffs could choose before trial to reduce the damages they would potentially collect and defendants could choose to increase the damages they would potentially pay.  

An extension of the earlier cost of performance strip-mining example shows how these options might work. When a promisor, bargaining in the shadow of $30,000 cost of performance damages, inefficiently threatens to perform, the farmer/promissee should have the option of reducing his legal damages to, for example, $25,000. While choosing lower legal damages seems superficially "less favorable" for a plaintiff/promissee, doing so can actually benefit the promisee. As long as the promisor can credibly threaten performance, it may be able to buy back its promissory duty for a relatively small amount (say $19,000). By lowering the background damages, the promisee can make the promisor's threat non-credible. The promisor might still claim that he will perform in the absence of renegotiation, but the promisee (having reduced the damages) could now respond, "I don't believe you; when push comes to shove, you will breach and pay me $25,000 instead of performing at a cost of $30,000." Surprisingly, the promisee can increase its expected payoff by reducing its potential legal damages.

While plaintiff remittitur and defendant additur, combined with injunction inalienability, can both deter threats of inefficient performance, we will argue that there is a stronger case in equity and efficiency for structuring the law to deter plaintiffs' threats of inefficient injunctions. Plaintiffs' threats are likely to drive plaintiffs' payoffs substantially above make-whole damages while defendants' threats are likely to drive plaintiffs' payoffs down closer toward make-whole dam-

25 A provision in an initial agreement that fixed liquidated damages as the exclusive remedy could obviate the need for the additur and remittitur options. But parties may be reluctant to fix liquidated damages when they cannot anticipate what the amounts should be at the time they enter the contract. As discussed below in Part IV.A, our proposed reforms only constitute default rules that would govern in the absence of a contrary agreement.
ages. We suggest that courts give defendants the options of injunctive inalienability and additur to deter plaintiffs' efforts to increase payments above expectation damages, but that courts not try to deter defendants' efforts to use performance threats to reduce damages toward the expectation amount.

While it is nigh-on impossible to construct a single damage rule that will induce efficient behavior along all possible dimensions, giving defendants the options of inalienability and additur leads toward more efficient plaintiff precaution and, under certain conditions, more efficient defendant reliance as well as a possibility of more efficient ex post negotiations. The strongest rationale for injunctive inalienability (and additur), however, is based not on efficiency but on equity. Some injunctions are granted in order to give the plaintiff rights-holder an entitlement to bargain for whatever exchange value the plaintiff can negotiate. Injunctions against patent infringement, for example, are often aimed at giving the patentee the ability to negotiate a high licensing fee. But judges award many, if not most, injunctions merely to provide the plaintiff the "use value" of actual performance. "Use value" injunctions are granted in part because of courts' concerns that monetary damages may not fully compensate the plaintiff. Giving plaintiffs the option to seek alienable injunctions reduces the chance of undercompensation at too high an equitable price by creating the possibility of substantial overcompensation. Giving plaintiffs instead the choice of either an inalienable injunction or damages (possibly enhanced by the defendant) retains the prime benefit of an alienable injunction, the elimination of the threat of undercompensation, by ensuring that plaintiffs can receive in kind the actual performance to which they are entitled. But an inalienable injunction sharply reduces the inequitable risk of overcompensation—that is, the risk that plaintiffs will seek an injunction solely for the pur-

26 See Steven Shavell, Damage Measures for Breach of Contract, 11 BELL J. ECON. 466, 488-89 (1980) (concluding that although the use of damage measures is in the mutual interest of both parties to a contract and leads to efficient behavior, factors such as the respective risk aversion of the parties and limited information of the enforcing court can limit the effectiveness of such measures). In a 1985 article, Robert Cooter emphasizes the difficulty of formulating efficient legal rules when efficiency requires bilateral precaution. See Robert Cooter, Unity in Tort, Contract, and Property: The Model of Precaution, 73 CAL. L. REV. 1, 4 (1985) (arguing that, because assigning full responsibility for the injury to one party or parceling it out between the parties cannot fully internalize costs for both of them, there is no level of compensation that achieves double responsibility at the margin).

27 The distinction between "use value" and "exchange value" is discussed below at notes 114-15 and accompanying text.
pose of receiving a payment far greater than the amount that they actually value performance.

Common law courts already respond to the problem of plaintiffs threatening inefficient performance by denying injunctions that impose disproportionate hardship on the defendant in comparison to the benefit that the plaintiff would derive from performance of the injunction.\(^\text{28}\) We support these decisions. Often, however, courts with imperfect information or courts worried about the possibility of plaintiff undercompensation grant an injunction where the defendant's burden outweighs the plaintiff's benefit. Courts should accordingly go beyond the decisions judiciously denying injunctions; they should consider regulating plaintiffs' ability to hold up defendants for super-compensatory amounts.

Our defendant additur and inalienability options rely on the game-theoretic prediction that such defendant options can deter plaintiffs from initially seeking injunctions. A problem with such reforms is that if they do not succeed in deterring plaintiffs, they can lead toward even more inefficient performance. As an alternative, we suggest that judges consider subjecting all injunctive settlements to the same type of remittitur analysis to which a jury award would be subjected. A remittitur review of injunctive settlements—that is settlements whereby defendants agree to pay plaintiffs in lieu of performance—would amount to an ex post judicial cap on how much plaintiffs could gain from an injunction. Such a review might achieve some of the deterrence effects of inalienability, without seeming like such a radical departure from current procedure and without creating as great a risk of inefficient performance. Any of our proposed reforms—judicial remittitur, inalienability, or defendant additur—should be thought of as default rules that the parties could disclaim in an initial contract or that defendants could disclaim at an early stage in the litigation before the plaintiff elects monetary or injunctive relief.

This Article is divided into four sections. The first section explores the conditions under which people who are owed duties and people who owe duties will threaten inefficient performance. The second section shows how the inalienable injunction and private remittitur and additur options will deter such threats by undermining their credibility. The third section assesses the ex ante and ex post efficiency effects of threats to perform and of our proposals to stop

\(^{28}\) See infra notes 45-46 and accompanying text (addressing the common law undue burden rules that courts use to deny injunctive relief in certain situations).
them. The final section discusses more specific means by which courts may limit threats to perform.

I. A Unifying Theory of Performance Threats

This section presents a simple model to show (1) when someone who owes or is owed a duty will desire to threaten inefficient performance and (2) when such a threat will be credible. In this model, imagine that a potential defendant owes some duty of performance to a potential plaintiff. The duty might be contractual in nature (as in the Peeryhouse duty to return topsoil) or non-contractual in nature (as in the Pile duty to remove an encroaching wall). Assume that:

- \( C \) = the defendant’s cost of performance; and
- \( B \) = the plaintiff’s benefit from performance.

Because we are interested in exploring threats when performance is inefficient, we also assume that at the time performance is due \( C > B \). If either side credibly threatens performance (as formally defined below), the parties will negotiate to avoid the inefficiency—specifically, the defendant will offer to pay the plaintiff to release the defendant from her performance duty. When the parties have asymmetric information about \( B \) and \( C \), this negotiation can produce inefficient results. For now, we assume \( B \) and \( C \) are common knowledge between the players and negotiation costs are nil so that the players expect negotiations to succeed. We also assume the expected negotiated amount the defendant pays for the plaintiff’s release in the shadow of a credible performance threat is:

\[
N = \alpha C + (1 - \alpha) B, \text{ where } \alpha \sim [0,1] \text{ is a measure of plaintiff’s bargaining power.}
\]

\( C \) and \( B \) are the players’ “threat points” or “BATNAs” (best alternative to negotiated agreement) once performance has been credibly threatened. The negotiated amount, \( N \), falls somewhere between these threat points depending on the parties’ relative bargaining

\[\text{By flip of a coin, we have decided to refer to the plaintiffs and the defendants by male and female pronouns respectively.}\]
\[\text{Even though contractors would not enter into a contract expecting } C \text{ to be greater than } B, \text{ after a contract is formed, } C \text{ may turn out to be higher than expected (or } B \text{ may turn out to be lower than expected).}\]
power, which for convenience, we have reduced to the scalar, \( \alpha \). 31

Finally, we assume that if the defendant fails to perform, the plaintiff can seek either specific performance or monetary damages equal to:

\[
D = \text{expected monetary damages if the defendant fails to perform (and if plaintiff does not seek specific performance).}
\]

We will consider the players' incentives under a range of different damage levels—including subcompensatory \( (D < B) \), compensatory \( (D = B) \), supercompensatory \( (D > B) \), and a special subcategory of supercompensatory, cost of performance \( (D = C) \).

Two conditions must be present before performance threats will generate bargaining. First, one of the parties must desire to threaten performance (motive), and second, that party must be able to make a credible threat (opportunity).

A. The Motive to Threaten Inefficient Performance

The relationship between the level of damages \( (D) \) and the expected payoff from negotiation \( (N) \) will determine which side has an incentive to make a threat. The motive to threaten inefficient performance is the motive to supplant the damage award with a more favorable negotiated amount. More specifically:

- If \( D < N \), the plaintiff will want to threaten inefficient performance to increase his expected payment; and

- If \( D > N \), the defendant will want to threaten inefficient performance to decrease her expected payment.

These inequalities suggest that one side or the other will often have a desire to threaten inefficient performance in order to supplant the expected award of damages. Damages are usually set to approximate \( B \) in order to make the plaintiff whole, or alternatively set to approx-

---

31 Bargaining power might turn on a host of bargaining primitives, including procedure (take-it-or-leave-it vs. alternating offers, etc.), the players' relative impatience, the cost of bargaining, and even the hopes of the parties involved. See Jennifer Gerarda Brown, The Role of Hope in Negotiation, 44 UCLA L. REV. 1661, 1669 (1997) (arguing that hope acts as an independent, primitive variable upon which negotiation behavior depends).
proximate C (as with cost of performance or restitutionary awards). When the parties bargain in the shadow of a performance threat, however, the negotiated amount, N, will systematically be between B and C. The larger the inefficiency (i.e., the more C exceeds B), the larger the likelihood that one side or the other will have a strong incentive to change the expected payment from D to N. Thus, in our earlier encroachment example, when legal damages were set to approximate B, the plaintiff had a strong incentive to seek an injunction to threaten inefficient performance and the likely negotiation it in turn would produce. In our cost of performance example, where damages were set to approximate C, the promisor had a strong incentive to threaten inefficient performance to reduce her ultimate payment. In a richer model, the costs of renegotiation and the costs of failed negotiation could dampen the parties' desire to make such threats, but perversely, the more inefficient performance becomes, the larger the chance that at least one party will have a desire to threaten inefficient performance.

B. When Will a Potential Defendant's Performance Threat Be Credible?

Motive, however, does not connote opportunity. Only credible threats of inefficient performance will spur negotiation. If the threatened party has no reason to believe the threatener will carry out the threat to perform, he has no reason to bargain. This credibility problem particularly restricts the ability of potential defendants to threaten inefficient performance. The opportunity for defendants to threaten performance arises when, before performance is due, potential de-

---

32 Cost of performance damages can approximate expectation damages under the assumption that the plaintiff will use the damage amount to purchase performance and thereby put himself in the same position as if the defendant had performed. See Steven J. Burton, More on Good Faith Performance of a Contract: A Reply to Professor Summers, 69 IOWA L. REV. 497, 506 (1984). When changed circumstances cause the defendant's cost of performance to radically exceed the plaintiff's benefit, however, then the plaintiff is unlikely to use the cost of performance damages amount received to buy substitute performance—so that cost of performance puts the plaintiff in a better circumstance than actual performance.

33 After performance is due, a promisor cannot credibly threaten to perform (unless the promisor has a right to cure, see U.C.C. § 2-508 (1989)). More generally, a person owing some non-contractual duty to another cannot threaten inefficient performance if there is not sufficient opportunity to contract for payment instead of performance. Imagine, for example, a stylized version of Hewlett in which a court imposed a tort duty of care on barge operators to carry a particular radio in case of an
fendants threaten to actually perform their promise or noncontractual duty—insead of failing to perform and being held liable for damages (D). The potential defendant implicitly threatens the plaintiff: "Let me pay you N instead of D or else I will perform and you'll only have a benefit of B." This threat is only credible, however, if the plaintiff believes that if he refuses the offer, the defendant will actually perform. In the simple model, the defendant's threat will be credible only if:

\[ D > C. \]

When expected damages equal or exceed the defendant's expected cost of performance, the defendant will be able to credibly threaten performance in the absence of renegotiation because carrying through on the threat costs the defendant the same as or less than paying damages.

While legal damages often focus on making the plaintiff whole, in a variety of contexts courts instead invoke a "disgorgement" principle that attempts to put the defendant in the same position as if she had performed her duty.\(^4\) One particular form of disgorgement remedy, cost of performance damages, can often give potential defendants the ability to make a credible performance threat. If cost of performance damages equal the defendant's actual cost of performance, the defendant loses nothing by carrying out her threatened performance in the event that negotiations fail.

Commentators have correctly noted that cost of performance damages and other extraordinary damages that place the defendant

---

\(^4\) See E. Allan Farnsworth, *Your Loss or My Gain? The Dilemma of the Disgorgement Principle in Breach of Contract*, 94 *Yale L.J.* 1339, 1354-69 (1985) (discussing judicial use of the disgorgement principle in cases "involving fiduciaries, sellers of goods who would be liable in conversion, and sellers of land"). Courts often struggle with the issue of whether to give cost of performance or cost of repair damages when such damages would greatly exceed a "diminution in value" measure. See, e.g., Heninger v. Dunn, 162 Cal. Rptr. 104, 106-09 (1980) (explaining the use of the diminution of value and restoration damages where replacement costs are unreasonable); Jacob & Youngs, Inc. v. Kent, 129 N.E. 889, 891 (N.Y. 1921) (Cardozo, J.) (arguing the merits of diminution of value versus cost of completion remedies in a breach of contract case where "the cost of completion is grossly and unfairly out of proportion to the good to be attained").
in the position of performance are (like specific performance) a species of what Calabresi and Melamed called "property rules," which tend to deter defendants from non-consensual takings.\textsuperscript{35} Avery Katz has called such damages "liquidated specific performance," arguing that such remedies give the plaintiff all of the gains from efficient breach.\textsuperscript{36} But the foregoing shows that cost of performance and other disgorgement damages may not result in giving the plaintiff all of the gains from performance. By allowing the defendant to credibly threaten inefficient performance, disgorgement remedies may allow the defendant to bargain to pay substantially less.

What happens when damages are less than the defendant's cost of performance (D < C)? It would at first seem that the promisor's threat to perform would lose its credibility. The validity of this conclusion, however, hinges on the divisibility of the defendant's performance. In a 1996 article, Lucian Bebchuk showed that plaintiffs who expect to spend more litigating than they would gain from winning a suit might still have a credible threat to sue if the litigation costs are incurred in stages over time.\textsuperscript{37} Bebchuk's insight can be applied in the bargaining context where the anticipated damage award is less than the cost of performance. If the defendant's performance is due in discrete stages, it may be credible for the defendant to threaten performance even though the total cost of performance is greater than the expected damages. For example, assume that the total cost of performance is 30, but that it is broken up into two discrete stages each costing 15. Assume also that the plaintiff's benefit (B) is 8, that expected damages (D) are 27, and that the parties have equal bargaining power and an opportunity to bargain before each stage of

\textsuperscript{35} See generally David D. Haddock et al., An Ordinary Economic Rationale for Extraordinary Legal Sanctions, 78 CAL. L. REV. 1 (1990) (developing a property analysis of the role of extraordinary legal sanctions in an efficient legal system and extending the model to explain various seemingly illogical and disjointed tort and contract damage rules).


\textsuperscript{37} See Lucian Arye Bebchuk, A New Theory Concerning the Credibility and Success of Threats to Sue, 25 J. LEGAL STUD. 1 (1996) (utilizing backward induction to explain why defendants can expect to settle negative value suits due to the divisibility of litigation costs). When litigation costs are divisible, the costs incurred in the last stage may be less than the damage award, making a threat to sue credible at that stage. If the parties reach this stage, they will expect to settle. The plaintiff will of course take this anticipated last-stage settlement into consideration when he decides whether to embark on the next-to-last stage. If his payoff from this last-stage settlement is greater than his next-to-last stage litigation costs, then he has reason to initiate a suit. In other words, his litigation threat is credible.
performance. Then it is possible to show—à la Bebchuk—that the defendant can credibly threaten performance. Even though the total cost of performing (30) is more than the defendants expect to pay in legal damages (27), the division of performance over time makes the threat credible. It can be shown that if the parties have equal bargaining power in equilibrium, the defendant should only expect to pay the plaintiff 17.25 to release herself from her duty, an amount substantially less than either damages or cost of performance. The defendant’s threat to perform will be credible for any fixed damage amount greater than 26.5. This example shows then that performing over time can increase the defendant’s ability to make credible threats of inefficient performance.

The size of D is crucial to determining whether the defendant’s threat to perform is credible, but if the performance threat is credible the damages have no independent impact on the expected settlement. A core result of non-cooperative bargaining theory is that the negotiated amount will be a function of the parties’ threat points. If

---

38 Breaking performance into two stages makes the defendant’s performance threat credible. In the last stage, the defendant will actually prefer to perform (since she will spend 15 rather than pay damages of 27). Since her performance threat at this last stage is credible, the parties would agree to settle at this stage (if they ever reached this stage) for 11.5 (the midpoint solution between the plaintiff’s benefit of 8 and the seller’s non-sunk prospective costs of 15). Anticipating this negotiation at the last stage, the defendant prefers to expend 15 at the first stage, since 15 + 11.5 = 26.5 is less than damages at 27. But given that the staged costs make full performance credible from the beginning, both parties prefer to negotiate immediately, settling on a payment midway between the parties’ threat points: a (8 + .5 (26.5 - 8)) = 17.25 payment by the defendant to the plaintiff.

39 There is, however, a caveat concerning this analogy to Bebchuk’s litigation cost model. In the litigation cost model, expenditures on suit preparation are assumed not to be correlated with the amount of damages that would be awarded in the suit. In other words, a non-credible threat becomes credible since the division into stages makes the litigation costs the party has yet to incur drop, while the suit damages remain constant. In contrast, damages awarded in a breach of contract suit may be a function of the performance that has occurred. For example, a painter paints the bottom floor of a house, but announces that he will breach the contract before completing the second floor of the house. The cost of painting the first floor will likely be deducted from the damages that would have been awarded had the contract been breached before any performance. If the level of damages and the cost of performance decrease by exactly the same amount, the threat will remain non-credible. For this reason, if partial performance reduces the expected damage award (D), then the plaintiff’s performance threat will only be credible where the adjusted damage award is proportionally larger than D. This might happen if the court believes it is difficult for a second contractor to pick up where the first contractor left off. Contracts involving some sort of learning curve would be one example. In this case, the cost of completion damage award would necessarily be greater than the original promisor’s remaining cost of performance.
the defendant can credibly threaten performance, then the relevant payoffs in the absence of agreement become B and C. The level of damages should play no role in determining the size of the negotiated amount (N) other than ensuring credibility because, in the shadow of a credible performance threat, there are no circumstances under which the defendant will ever have to pay court-awarded damages.

Focusing on the defendant's ability to make credible threats also illuminates an interesting distinction between cases of "impossibility" and "impracticability." Defendant threats are not credible when performance itself is literally impossible. A defendant facing cost of performance damages, however, can credibly threaten performance when performance is possible but merely impracticable (i.e., extremely costly). The law, at times, responds to evidence of either impracticability or unconscionability by reducing the defendant's expected damages. This is primarily accomplished by using the diminution in value damage measure instead of a cost of performance measure40 (and more rarely by voiding the defendant's duty to perform altogether, effectively reducing the defendant's damages for nonperformance to $0).41 If lawmakers wish to deter defendant threats,42 there is a stronger case for reducing the defendant's damages when the defendant's performance is merely impracticable not impossible. Reducing damages substantially below the cost of performance when performance is impracticable can render the defendant's threats non-credible. When the defendant's performance is impossible, however, it is not necessary to reduce the defendant's liability to deter performance threats. The defendant can not credibly threaten to do the impossible. Thus, while we normally think that evidence of impossibility provides a stronger rationale for relief than mere impracticability, from the standpoint of deterring the defendant's threats of performance, the opposite is true.

C. When Will a Plaintiff's Performance Threat Be Credible?

Plaintiffs threaten to cause defendants' performance by seeking injunctive orders of performance. The major obstacles in making this threat are equitable rules limiting the award of this "extraordinary"
remedy. The primary limitation, as traditionally formulated, is that injunctions will only be granted to prevent an irreparable injury (or as equivalently formulated, if there is no "adequate legal remedy"). But Douglas Laycock has shown as a matter of positive U.S. law that "[t]he irreparable injury rule almost never bars specific relief" and an emerging consensus is that courts routinely give plaintiffs an option for specific relief in a far wider range of contexts than previously thought. Under a variety of doctrines, however, courts will still deny injunctive relief that imposes undue hardship on a defendant (in comparison to the benefit that the plaintiff will derive from the injunction). These undue burden limitations make good economic sense because they limit the ability of plaintiffs to threaten inefficient performance when their motive to do so is the strongest. The essence of the "undue burden" rules is to identify circumstances where the defendant's cost of performing some duty (C) far exceeds the plaintiff's benefit of performance (B). It is in just these circumstances that the amount expected from negotiating away an injunction (N) is likely to far exceed what the plaintiff might receive in court-awarded damages, and that plaintiffs will therefore have an incentive to seek the injunction as a bargaining chip. The refusal of courts to issue injunctions that

---


45 See RESTATEMENT (SECOND) OF TORTS § 826(a) (1977) (intentional nuisance should be enjoined as unreasonable only if "gravity of the harm outweighs the utility of the actor's conduct"); YORIO, supra note 16, at 41 ("For specific performance to be proper, however, the marginal benefit to the promisee must be sufficiently great that it outweighs the marginal costs imposed on the promisor and on the legal system."); see also DOUGLAS LAYCOCK, MODERN AMERICAN REMEDIES: CASES AND MATERIALS 219 (1985) ("One of the grounds for vacating a final judgment of injunction is that the injunction is causing undue hardship to defendant...”).

46 See RESTATEMENT (SECOND) OF CONTRACTS § 364(1)(b) (1979) (discussing when injunctive relief is inappropriate due to unreasonable hardship to the party in breach); YORIO, supra note 16, at 110 ("[T]he issue of equitable relief is viewed often in terms of a balancing test, with specific performance denied if the burden on the defendant from enforcement of the contract would exceed the benefit to the plaintiff"); see, e.g., Ben Simon's, Inc. v. Lincoln Joint-Venture, 535 N.W.2d 712, 715 (Neb. 1995) (refusing to issue an injunction to tear down restaurant built in violation of lease because benefit to plaintiff was greatly exceeded by burden on defendant).

47 The classic illustration of this is Boomer v. Atlantic Cement Co., Inc., 257 N.E.2d 870, 872 (N.Y. 1970) ("The total damage to plaintiff's properties is, however, relatively small in comparison with the value of defendant's operation and with the consequences of the injunction which plaintiffs seek."). As parsed recently by Judge Posner: The defendant's factory was emitting cement dust that caused the plaintiffs harm monetized at less than $200,000, and the only way to abate the harm would have been to close down the factory, which had cost $45 million to
would impose an undue hardship on defendants has "the effect of preventing the plaintiff from using an equitable remedy to extort an overcompensatory settlement." The possibility of such equitable defenses accordingly undermines the credibility of plaintiffs' performance threats. A court's refusal to grant an injunction on such equitable grounds, however, also increases the chance plaintiffs will be undercompensated by monetary damages.

Because of this risk of undercompensation, courts at times do grant injunctions when defendant's injunctive burden (C) is substantially greater than plaintiff's benefit (B). Courts often have difficulty assessing the true costs and benefits of injunctive performance and thus may be unable to determine whether performance would be inefficient. Even when courts believe that performing an injunction would be inefficient, their worry that damages would be inadequate (i.e., would undercompensate a plaintiff) may cause them to issue an injunction as an equitable matter. Courts might be especially concerned that, because juries decide the awards, the risk of undercompensation is unavoidable.

When courts do give plaintiffs the choice of monetary or injunctive relief, plaintiffs can credibly threaten to seek injunctive performance. It might seem at first that plaintiffs could not credibly threaten to seek an injunction when they would benefit less from injunctive performance than from expected damages (B < D). After all, defen-
defendants have an analogous trouble threatening performance when the cost of performance is greater than the cost of damages \( C > D \). This analogy fails because, unlike defendants, plaintiffs have the ability to commit to performance in the absence of a successful negotiation. The structure of civil litigation gives plaintiffs the ability to make such a pre-negotiation commitment. There is almost always a period of time between the plaintiff's election of an injunctive remedy and the defendant's actual performance. Since the defendant will be willing to pay \( N \) once the plaintiff has sought an injunction, it becomes credible for a plaintiff to seek the injunction—even if \( B < D \).

Figure 1 is a simple game-tree example showing why a plaintiff can credibly seek an inefficient injunction, even when the plaintiff's benefit from receiving actual performance is less than his expected damages. In this example, we assume that the defendant's performance of some duty is inefficient \( B = 0 < C = 30 \). Expected legal damages are supercompensatory \( D = 1 > B = 0 \), but less than the expected negotiated payment that would result if the plaintiff could credibly threaten to cause performance \( D = 1 < N = 15 \). We also assume the parties negotiate initially, and that the plaintiff then sues for either an injunction or monetary damages. If the plaintiff elects an injunction, the parties have an opportunity to negotiate again. The circles in the figure identify the decision makers at each stage of the game, and the brackets show the monetary payoffs (for the plaintiff and defendant, respectively) for each potential sequence of decisions.

![Figure 1: Credibility of Seeking Injunction When B < D](image-url)
Solving backward, it is easy to see that if the parties reach the final negotiation node, both will prefer to settle (i.e., to agree to dissolve the injunction against the defendant in exchange for payment of 15 to the plaintiff). Foreseeing this outcome, the plaintiff will strongly prefer an injunction (with a 15 payoff) to monetary damages (with a 1 payoff). Finally, because both parties can foresee that the plaintiff will elect injunctive relief, they are likely to settle immediately for \( N = 15 \) (especially if litigation costs are positive).

D. When Do Performance Threats Cause Payoffs to Diverge From Make-Whole Compensation?

Restricting our attention to the situation in which performance is inefficient (so that \( B < N < C \)), damages can fall into one of four relevant damage ranges as depicted in Figure 2:

![Figure 2: Four Possible Expected Damage Ranges When Performance Is Inefficient](image)
Table 1 summarizes much of the foregoing discussion in terms of these four damage ranges.

<table>
<thead>
<tr>
<th>Damage Level</th>
<th>Whose Motive To Threaten</th>
<th>Is the Threat Credible?</th>
<th>Expected Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. D &lt; B &lt; N &lt; C</td>
<td>Plaintiff</td>
<td>Yes</td>
<td>N</td>
</tr>
<tr>
<td>2. B &lt; D &lt; N &lt; C</td>
<td>Plaintiff</td>
<td>Yes</td>
<td>N</td>
</tr>
<tr>
<td>3. B &lt; N &lt; D &lt; C</td>
<td>Defendant</td>
<td>No$^{50}$</td>
<td>D</td>
</tr>
<tr>
<td>4. B &lt; N &lt; C &lt; D</td>
<td>Defendant</td>
<td>Yes</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 1: Credible Threats in Four Potential Damage Ranges

When expected damages are in Ranges 1 or 2, the plaintiff will want to seek injunctive relief to increase his expected payoff from D to N.$^{51}$ As long as the court is willing to issue an injunction in this type of case (for example, where there is no finding of an undue burden on the defendant), the plaintiff's threat to seek an injunction will be credible and the parties will try to negotiate a payment of N to avoid the inefficiency.

The possibility that court-awarded damages will fail to make the plaintiff whole (Range 1) is one of the primary justifications for granting injunctions.$^{52}$ Judge Richard Posner in particular has criticized the inaccuracy of court-awarded damages relative to negotiated outcomes in the shadow of an impending injunction: "A battle of experts is a less reliable method of determining the actual cost to [plaintiff] of [not receiving defendant's performance] than negotiations between [plaintiff] and [defendant] over the price at which [plaintiff] would feel adequately compensated for [not receiving performance]."$^{53}$ Once we appreciate the possibility of threatening inefficient performance, however, we see that the payment negotiated in an injunctive

---

$^{50}$ The threat may be credible, however, if costs are incurred in stages as in the Bebchuk model. In this case, the expected payment would be N as well. See Bebchuk, supra note 37, at 29 (stating that when costs are incurred over time, a plaintiff's otherwise noncredible threat to sue may become credible).

$^{51}$ Of course, if the defendant's performance creates no benefit for the plaintiff, then Range 1 will be a null set—i.e., there is no possibility that damages will be less than the plaintiff's benefit.

$^{52}$ See Miller v. LeSea Broad., Inc., 87 F.3d 224, 230 (7th Cir. 1996) (noting that damages may be "an inadequate remedy... because of the defendant's lack of solvency or because of the difficulty of quantifying the injury to the victim of the breach").

$^{53}$ Walgreen Co. v. Sara Creek Property Co., 966 F.2d 273, 276 (7th Cir. 1992).
shadow can sharply diverge from make-whole damages. Injunctions not only generate negotiation costs but also introduce inaccuracies of their own. Damages may over- or undercompensate, but injunctions create the risk of substantial overcompensation, and this risk grows when the bargaining range grows. When the inefficiency of performance creates a huge bargaining range, damages may better approximate true make-whole relief than negotiations in the shadow of a credible performance threat.

In Range 2, the threat of inefficient performance unambiguously moves the plaintiff's payoff further away from make-whole relief. Expected damages are already above the make-whole level and threatening inefficient performance only exacerbates the problem (B < D < N). In Range 1, by contrast, inefficient performance threats substitute a supercompensatory payment (N > B) for subcompensatory damages (D < B). In both ranges, as the difference between the defendant's cost and the plaintiff's benefit grows, it becomes likely that the negotiated payment (which must fall between these two amounts) will deviate more from make-whole relief than even speculative damages. Giving plaintiffs the option of injunctive relief thus can give plaintiffs the opportunity to raise damages substantially above the break-even level.

In damage ranges 3 and 4, it is the defendant's turn to threaten performance, in the hope of reducing her expected payment from D to N. This threat clearly will be credible in Range 4 because the cost of following through on the performance threat is less than the cost of paying damages. However, in Range 3, the performance threat will not be credible unless divisible performance extends a potential de-

---

54 Judge Posner sees the primary costs of damage remedies to be the costs of court determination and the costs of inaccuracy, while he sees the primary costs of injunctive relief to be the private "bilateral monopoly" cost of determining the payment:

[When an inefficient injunction creates a large bargaining range,] both parties will have an incentive to devote substantial resources of time and money to the negotiation process. The process may even break down, if one or both parties wants to create for future use a reputation as a hard bargainer; and if it does break down, the injunction will have brought about an inefficient result. All these are in one form or another costs of the injunctive process that can be avoided by substituting damages.

*Id.* at 276. Both injunctive and damage remedies can produce determination costs and inaccuracies. It is important to recognize that negotiating in the shadow of performance threats does not necessarily produce outcomes that more accurately approximate make-whole damages than court-determined awards.

55 For ease of exposition, Table 1 ignores the case when expected damages equal B or N or C. D = B will generate the same result as Range 1; D = N will make each side indifferent to threatening inefficient performance; D = C will generate the same result as Range 4.
fendant's ability to threaten along the lines of the Bebchukian argument discussed above.\textsuperscript{56}

Performance threats by defendants, unlike performance threats by plaintiffs, move the plaintiff’s expected payoff unambiguously toward make-whole relief. Potential defendants threaten performance when expected court-awarded damages exceed the expected settlement with the plaintiff, which must itself exceed the plaintiff’s benefit of performance ($D > N > B$). While plaintiff threats tend to cause the plaintiff’s payoffs to exceed make-whole relief, defendant threats (if successful in producing a negotiated settlement) cause the plaintiff’s payoff to approximate more closely make-whole relief. This distinction will play an important role in Part III when we analyze its implications for choosing efficient legal rules.

II. HARNESSING PRIVATE INFORMATION TO DISCREDIT PERFORMANCE THREATS

This section describes how the law might be changed to undermine the credibility of performance threats. The basic approach is to give the threatened party a countervailing option which makes actually carrying through on the threat unprofitable for the threatener. We begin by showing how allowing defendants to make injunctions inalienable can deter some plaintiff threats (in Range 1). By inalienable, we mean only that the defendant cannot pay the plaintiff to release the defendant from her duty to perform the injunction.\textsuperscript{57} Inalienability simply prohibits defendants from paying plaintiffs in lieu of

\textsuperscript{56} See supra text accompanying notes 37-39.

\textsuperscript{57} There is currently a strong consensus that defendants can freely negotiate to discharge their injunctive duties. See DUKEMINIER & KRIER, supra note 19, at 863 (“It is said that ‘an injunction is for sale,’ meaning the person who holds it may sell it to the enjoined party if the price is right.”). Some courts have even held that the Federal Rules of Civil Procedure do not give them power to restrict the ability of disputants to settle, instead of perform, an injunction. See Smith v. Phillips, 881 F.2d 902, 904 (10th Cir. 1989) (holding that the parties have a “right to unconditional dismissal under Rule 41(a) (1)(ii)’’); Wheeler v. American Home Prods. Corp., 582 F.2d 891, 896 (5th Cir. 1977) (“The only provision [in the Federal Rules of Civil Procedure] for approval of a settlement is that for dismissal or compromise of a class action . . . .”). For a further discussion of inalienable injunctions, see Ayres & Siegelman, supra note 9. Still, there is no compelling equitable or efficiency reason why judges, in tailoring the scope of an equitable remedy, should not be able to restrict the alienability of injunctions. At least one case seems to imply that a permanent injunction, once issued, would be inalienable. See Rievman v. Burlington N. R.R. Co., 118 F.R.D. 29, 33 (S.D.N.Y. 1987) (“Should defendants be permanently enjoined from prematurely releasing the . . . lien . . . that lien would cease to confer any ‘hold-up’ value on the bonds.”).
performing an injunction. The plaintiff’s injunctive right would be freely alienable to anyone except the defendant. Plaintiffs worried that legal remedies would be inadequate would still be free to seek injunctions, but the inalienability of the injunction would tend to assure that it was sought for the performance itself and not for the extra bargaining power conferred by the threat of performance.

We then show that allowing the threatened party to adjust prospective damages seemingly against its own interest, what we call private additur and remittitur, can further undermine the credibility of the other side’s threats. For example, if a defendant says that A dollars should be added onto any court-determined damage award (D), then a plaintiff must choose between the benefits of a possibly inalienable injunction and inflated expected damages (D + A). Perversely, because of the possibility of inefficient threats, defendants can actually make themselves better off by petitioning the court to award larger damages and plaintiffs can make themselves better off by petitioning the court to award smaller damage.

A. The Inalienable Injunction Option

Figure 3 illustrates the potential of an inalienability option to deter threats of inefficient injunctive performance. For concreteness, assume an encroachment dispute in which the defendant’s cost of removing the encroachment is 30, the plaintiff’s benefit of unencroached land is de minimis (= 0), expected damages are 1, and the expected negotiated payment, if an alienable injunction issues, is 15.58 These were the same numbers used above in Figure 1, where we saw that the plaintiff, under existing law, has an incentive to threaten inefficient injunctive performance—even though damages are expected to be supercompensatory.

---

58 Although we have chosen an encroachment setting, the simple economic structure might also capture many contractual disputes where the defendant’s cost of performance ends up being far greater than the plaintiff’s benefit from performance.
The game tree in Figure 3 is modified, however, to give the defendant an option to make any injunction awarded inalienable.\(^5\) If the underlying values (B, D, C, and N) are common knowledge,\(^6\) the defendant foresees that if she chooses inalienability, the plaintiff will have an incentive to opt for monetary damages. The plaintiff prefers the damages option since actual performance of an inalienable injunction gives the plaintiff no benefit (B = 0), while damages are expected to be 1. For this reason, if the defendant elects “inalienabil-

---

\(^5\) The order of play in the game is as follows: in the first stage, the parties negotiate; in the second stage, the defendant chooses whether an injunction would be inalienable or not; in the third stage, the plaintiff elects injunctive or damage relief; and finally, if the defendant has chosen alienability in the second stage and the plaintiff has chosen injunctive relief in the third stage, the parties then have an opportunity to bargain to discharge the injunction.

\(^6\) See Ayres & Nalebuff, supra note 1, at 1692 n.2 ("'Something is common knowledge if it is known to each player, and, in addition, each player knows that the other player has this knowledge; knows that the other person knows the player knows it; and so forth.'") (quoting DOUGLAS G. BAIRD ET AL., GAME THEORY AND THE LAW 304 (1994)).
ity," she will pay damages of only 1. In contrast, choosing “alienable” would recreate the plaintiff’s incentive to threaten inefficient performance and hence would lead to a much larger payment (N = 15). In equilibrium, the parties are likely to settle initially for 1 (since they know that this will be the outcome if they continue to progress through the stages of the game).\(^6\)

In order for inalienable injunctions to deter inefficient injunctive threats effectively, it is important that: (1) the defendant be able to move first, i.e., be able to commit to inalienability before the plaintiff commits to injunctive relief; and (2) the parties not be allowed to “settle” a case for money once the defendant commits to inalienable injunctive relief. If the court first asks the plaintiff whether he wants injunctive relief and only then asks the defendant whether the

---

\(^6\) This simple game excludes negotiation and litigation costs. Adding these costs does not, however, qualitatively change the central result that giving defendants an inalienability option can deter plaintiffs’ incentives to seek inefficient injunctions.
injunction should be inalienable, the plaintiff will still threaten an injunction. Under this scenario (as illustrated in Figure 4), once the plaintiff commits to an injunction, the defendant will choose to have the injunction be alienable so that she can avoid paying 30. The inalienability option is only effective in deterring inefficient injunction threats if the defendant can commit to inalienability before the plaintiff commits to an injunction.

Moreover, inalienability options will only be effective if the court can enforce their inalienability, i.e., if they can prohibit monetary settlements of lawsuits after a plaintiff has opted for an inalienable injunction. In many litigation settings, one can imagine a delay of several weeks between the time that a plaintiff opts for injunctive relief and the time that the court actually issues the injunction. If plaintiffs know that they will have an opportunity to bargain to settle a lawsuit for monetary damages before an injunction is actually issued, then notwithstanding the defendant's choice of "inalienability," plaintiffs will seek an injunction in order to gain bargaining power to settle the case before the injunction actually issues. As shown in Figure 5 below, the ability to settle for money after the defendant and plaintiff opt for "inalienable" and "injunction" respectively fails to deter inefficient injunctive threats. The plaintiff forces a negotiation with the same expected payoff as we saw under the current injunctive scheme (in Figure 1) without an inalienability option for the defendant.

Indeed, even if the court prohibits monetary settlements (in the interim between the plaintiff's election of injunctive relief and the actual issuance of the injunction), the court also needs to worry that the parties will negotiate a monetary settlement after an injunction issues, notwithstanding the nominal inalienability of the court's order. After the fact, the parties will have a joint interest in negotiating away the inefficiency of the injunction and may be willing to collude to circumvent the inalienability restriction. To create the beneficial ex ante effects of the inalienability restriction, courts therefore will need to deter ex post settlements.

This analysis suggests that by adhering to the traditional election-of-remedies doctrine (which required plaintiffs to initially choose, and thereby be wedded to, a prayer for either monetary or injunctive relief), courts may unwittingly facilitate plaintiffs in threatening inefficient injunctive performance. See Yorio, supra note 16, at 213 ("[I]n the traditional election-of-remedies doctrine... the plaintiff was required to choose—and be wedded to—a single remedial theory.").
Courts could accomplish this in many settings by placing judicial liens on defendants' property that would only be lifted on proof to the court that the defendant had actually performed the injunction. Courts would find verifying actual performance easiest in contexts involving positive, durable injunctions—that is, orders that defendants do a particular act one time. For example, in *Pile* the court could have refused to lift a judicial lien until it was shown photographic proof that the fence was removed. In contrast, it would be much harder for courts to deter ex post settlements concerning negative, non-durable
precautions—that is, orders that defendants should not do a particular act over time. For example, if a court ordered a defendant not to trespass on the plaintiff's adjoining driveway, the court would have trouble verifying whether the parties covertly cut a deal to settle (i.e., nullify) an inalienable order. Even here the court might offer bounties to third parties who provide evidence of such a settlement. The court could threaten contempt sanctions, including jail for defendants who fail to perform inalienable injunctions and for plaintiffs who accept ex post payments.

The take home lesson of this section is that a properly implemented inalienable injunction may, in at least some contexts, deter plaintiffs from threatening inefficient injunctions. It is important to emphasize, however, that the examples thus far have concerned only supercompensatory damages ($D = 1 > B = 0$), what we earlier called Range 2. Inalienable injunctions deter plaintiffs from making inefficient performance threats when damages are supercompensatory (Range 2), but inalienable injunctions might actually induce inefficient performance when damages are subcompensatory (Range 1). When damages fail to make the plaintiff whole, the plaintiff will continue to prefer injunctive relief—even if the injunction is inalienable. For example, if we change the foregoing example by raising $B$ from 0

---

63 For the inalienability commitment to be effective, the court must not only have knowledge of whether the injunction is actually performed, but must also have the political will to follow through and require inefficient performance over the ex post protests of both the plaintiff and defendant. In certain contexts, the judicial lien may be sufficient to apprise the court of performance, but the difficulty of assuring the court's political will may be another reason to favor the judicial remittitur alternative discussed below in the text accompanying notes 119-20. The civil contempt power is discussed extensively in Bagwell v. International Union, 423 S.E.2d 349, 356 (Va. 1992), rev'd, 512 U.S. 821 (1994).

64 Defendants may, as a theoretical matter, already be able to privately opt for inalienability—by entering into a Schelling-like commitment contract with a third party. See generally THOMAS C. SCHELLING, THE STRATEGY OF CONFlict 22-28 (1960) (describing the benefits of introducing a third party into two-party negotiations). Under such a contract, the defendant would promise to pay a third party some outrageously large sum if the defendant were ever found to have negotiated her way out of a duty to perform. Such contracts would not completely destroy the incentive to negotiate one's way out of performance, rather the incentive would now exist for the defendant to pay off the third party as well as the plaintiff to avoid performance. If effective third-party commitment contracts are feasible, one might expect a commitment race whereby plaintiffs sought to commit to seeking an injunction before the defendant committed to inalienability.

65 See supra discussion accompanying Table 1 (introducing the concept of ranges of super- and subcompensatory damages). Ranges 3 and 4 also concern supercompensatory damages, but as discussed above, since $D > N$ in these ranges, the plaintiff would not desire to replace court-ordered damages with injunctive relief.
to 2 (holding $D = 1$, $C = 30$, and $N = 15$ constant), then the plaintiff would opt for an injunction over damages, even if the defendant had previously declared that any injunction would be inalienable. The plaintiff prefers an inalienable injunction under this fact pattern because it produces a payoff of 2 while court-ordered damages produce a payoff of only 1. Even though performance of the injunction is massively inefficient ($C = 30 > B = 2$), the plaintiff still prefers actual performance to damages. When damages are subcompensatory, both parties may be worse off with an inalienable injunction than with an alienable injunction.

The possibility of subcompensatory damages thus provides a strong reason for not asking courts to routinely make injunctions inalienable. Inalienability is an effective deterrent in Range 2, but courts will often have difficulty determining whether damages are sub- or supercompensatory, i.e., whether the parties are interacting in a Range 1 or a Range 2 context. The magnitude and direction of D's deviation from B may depend on the plaintiff's subjective valuation of B, a valuation that may be difficult to verify. Instead of asking courts to impose inalienability only when the court is confident that damages are not subcompensatory, we prefer giving defendants the inalienability choice. This preference is motivated by our belief that the defendants are likely to be better informed about the underlying relationships among the values of B, C, N, and D. We cannot see why a court should impose an inalienability restriction when the defendant objects. If the defendant thinks that inalienability would fail to deter a plaintiff from seeking an injunction, we think courts should generally defer to the defendant's judgement.

The inalienability restriction is a natural outgrowth of courts' concern about issuing injunctions that impose undue burdens. We envision that courts at times should continue to deny injunctions when they are confident that a plaintiff is seeking the injunction as a bargaining chip to supplement otherwise adequate damages. In other settings, when the court is suspicious that the plaintiff is merely trying to threaten inefficient injunctive performance, the court might propose inalienability and ask if the defendant objects or give the defendant the option of making any injunction inalienable.

---

66 Assuming that $N$ is unchanged (when $B$ increases) effectively increases the defendant's bargaining power. Alternatively, we could assume that bargaining power ($\alpha$) is unchanged and increase $N$ to 16.

67 Anticipating the possibility of inefficient performance, parties in a "Range 1" situation are likely to settle for $N$ before the defendant chooses "inalienability."
Granting defendants an inalienability option is effective in deter-
rming performance threats where the court is likely to award super-
compensatory damages in Range 2, but making the injunction inal-
ienable will not deter plaintiffs from seeking injunctions where the
court is likely to award subcompensatory damages (Range 1). Indeed,
even if the court is expected to issue precise make-whole damages (D = B), the plaintiff will have a credible threat to seek an inefficient in-
junction—because an inalienable injunction will make the plaintiff
just as well off as monetary damages. To deter plaintiffs from seeking
inefficient injunctions when damages are subcompensatory or exactly
compensatory (D < B), something more than inalienability is needed.

To solve this Range 1 problem, we recommend giving defendants
a "private additur" option in addition to the "inalienability" option de-
scribed above. Before a plaintiff is allowed to commit to injunctive re-
lief, the defendant would be able to commit to paying an "additur"
amount (A) in addition to any damages awarded by a jury. If the
plaintiff chooses monetary instead of injunctive relief, the defendant
would be required to pay A in addition to any amount awarded at
trial. The ultimate trier of fact would remain uninformed about the
additur arrangement so that D would be unaffected. The private
additur option would accordingly have the effect of raising the plain-
tiff’s expected payoff of monetary damages from D to D + A.68

If we analyze a Range 1 example, we can see that the additur op-
tion in combination with the inalienability option can deter a larger
range of inefficient performance threats than the inalienability option
alone. If D = 1, B = 2, N = 15, and C = 30, a defendant would have an
incentive to choose inalienability and to choose A = 1 + ε. The plain-
tiff would now face the choice of receiving the benefits of an inalien-
able injunction (B = 2) or the benefits of monetary relief (D + A = 2 + ε).
The plaintiff in pretrial negotiations could threaten to seek an in-
junction unless the defendant pays 15, but this threat would no longer
be credible. In the absence of agreement, the plaintiff would seek
monetary relief instead of an inefficient injunction.

68 The court would need to require a litigation bond or establish a damage escrow
to ensure that the defendant would be able to pay D + A.
69 The additur regime might instead be structured to allow the additur amount to
be a function of the damages amount—so that the defendant could guarantee the
plaintiff a minimum level of damages.
Perversely, the defendant is made better off by asking the court to increase the potential damages it must pay. While this result seems nonintuitive, Eric Rasmusen has noted: "One of game theory's most profound lessons . . . is that a player can benefit from new rules which reduce his payoffs on out-of-equilibrium paths."\(^{70}\) By invoking the private additur option, the defendant lowers her expected payment from 15 to \(2 + \varepsilon\); without a damage increase, the defendant would expect the plaintiff to seek an injunction, and therefore would be willing to pay 15 to avoid incurring the cost of performance of 30. Private additur changes the plaintiff's threat point in any pretrial negotiation. Without private additur, the plaintiff in the absence of agreement faces a choice between monetary and injunctive relief. Since by assumption in Range 1 monetary relief is inferior to the benefits of actual performance, the plaintiff can credibly threaten to seek an injunction. But private additur changes the plaintiff's best alternative. With private additur, monetary relief (yielding \(2 + \varepsilon\)) now dominates the benefit of seeking actual performance (2).

In simple models, the defendant would need to make the net monetary damages only infinitesimally larger than B in order to deter the plaintiff from seeking an injunction. In more complicated models (for example where the defendant is uncertain about the size of B) and in the real world, however, we believe that defendants would set additur amounts to make net damages substantially exceed their estimates of B. In the foregoing example, a defendant might set net damages at 8 or 10, as well as opt for inalienability, to increase the chance that the plaintiff would be sacrificing not a trivial amount to follow through on her threat, but a more substantial shortfall (8 or 10 as opposed to 2).

It will strike many readers that defendants already have an option of private additur because a defendant in pretrial negotiations can directly offer to settle the case for \(D + A\). There is a crucial contractual distinction, however, between settlement offers and private additur offers. A defendant's settlement offer is not "firm" and hence its effectiveness can be destroyed merely by a plaintiff's rejection. The traditional common law rule is that an offeree's rejection or counteroffer destroys an initial offer.\(^{71}\) Accordingly, a defendant's settlement offer

\(^{70}\) Eric Rasmusen, Book Review, 33 J. ECON. LITERATURE 1979, 1980 (1995) (reviewing KEN BINMORE, 1 GAME THEORY AND THE SOCIAL CONTRACT: PLAYING FAIR (1994)). Here a player benefits from reducing her payoff on what was an out-of-equilibrium path—in order to make it (monetary relief) an equilibrium path.

\(^{71}\) See, e.g., Beverly Way Assocs. v. Barham, 276 Cal. Rptr. 240, 244 (Ct. App. 1990)
is not effective in changing the plaintiff's BATNA. In Range 1, if the defendant offers to pay the plaintiff $3 to settle the suit, the plaintiff by merely responding "no" destroys the defendant's offer and again makes credible the threat of seeking an injunction. Since the private additur amount is an irrevocable offer to settle, the plaintiff's resistance cannot change the fact that the plaintiff, in electing remedies, must choose between an inalienable injunction yielding $B$ and money damages of $D + A$.72

By making a firm commitment to pay higher monetary damages (and by opting for inalienability of any injunction), defendants can deter plaintiffs' inefficient performance threats and thereby decrease their ultimate expected payment. Defendants will only exercise their private additur option when they believe court-awarded damages will be undercompensatory, in order to discourage plaintiffs from seeking inefficient injunctions. The additur option harnesses defendants' private information in an attempt to transform subcompensatory awards into awards that better approximate make-whole compensation. The additur option allows a defendant to convert a Range 1 undercompensatory threat into a Range 2 supercompensatory threat, which then can be deterred by making the injunction inalienable. If defendants find it in their self-interest to commit to paying higher damages (in order to deter plaintiffs from seeking inefficient injunctions), courts should acquiesce in their requests.73

("It is hornbook that an unequivocal rejection by an offeree, communicated to the offeror, terminates the offer . . ."); RESTATEMENT (SECOND) OF CONTRACTS § 36 (1979) (stating that a rejection or counter-offer terminates the offeree's power of acceptance). The U.C.C. already allows merchants to make limited firm offers regarding the sale of goods. See U.C.C. § 2-205 (1962) ("An offer by a merchant to buy or sell goods in a signed writing which by its terms gives assurance that it will be held open is not revokable. . ."). The U.C.C. provisions, however, would not apply to most settlement negotiations and the firmness created by the U.C.C. does not nearly match the firmness the court could demand with performance bonds and escrows and even the possibility of contempt proceedings if the defendant failed to live up to her additur offer.

72 The private additur offer is not structurally the same as a take-it-or-leave-it offer because we would allow either the plaintiff or the defendant to make additional non-firm offers. Still, the irrevocability of the private additur offer in this model has the same effect as giving the defendant the right to make a take-it-or-leave-it offer. The defendant would never have an incentive to make other offers or to accept other offers from the plaintiff. The private additur offer effectively limits the plaintiff to one of two choices: accepting or rejecting monetary damages.

73 A caveat to this proposal and others in this paper is that in the contractual setting we would propose that inalienability and damages adjustment options be just default rules. See infra Part IV.A.
C. The Plaintiff's Damage Adjustment Option—Private Remittitur

An analogue to private addituir can similarly be used to deter potential defendants from threatening inefficient performance when damages are in Range 4. As discussed earlier, if potential damages equal or exceed the defendant's cost of performance, the defendant will credibly threaten to perform unless the plaintiff accepts a payment substantially less than damages. Thus, if \( B = 0, N = 15, C = 30, \) and \( D = 32, \) before breaching the defendant will threaten to perform if the plaintiff does not accept 15 in lieu of performance. The defendant's threat is credible because in the absence of agreement, the defendant would rather perform her duty at a cost of 30 than pay 32 in damages.

It is again possible to give the threatened party options to undermine the credibility of the other side's threats. Just as we gave defendants addituir options to counter the possibility of plaintiff performance threats, it is possible to give plaintiffs a "remittitur option" to counter the possibility of defendant performance threats. A remittitur option would allow the plaintiff to announce a remittitur amount \( (R) \) by which any court award would be reduced. For example, if the jurisdiction is likely to impose cost of performance damages \( (D) \), the plaintiff could announce that the defendant would only have to pay \( D - R \). In the foregoing Range 4 example \( (B = 0, N = 15, C = 30, \) and \( D = 32) \), the plaintiff would have an incentive to set \( R \) equal to \( 2 + \epsilon \). Doing so would reduce the defendant's net expected damages to \( 30 - \epsilon \) and thereby undermine a threat to perform (at cost 30) if negotiations failed.\(^7\)

Giving plaintiffs an option to reduce the amount they would receive in damages increases their expected payment. In the foregoing example, if the plaintiff fails to reduce expected damages (setting \( R = 0 \)), the defendant will be able to credibly threaten performance and thereby negotiate a payment of only 15; but if the plaintiff reduces his expected monetary damages by a little more than 2, the defendant will

\(^7\) Our discussion of Bebchuk suggests that when performance costs are incurred over time, a defendant may under some circumstances credibly be able to threaten performance even when the total costs of performance are greater than the expected damages the defendant would have to pay for breaching. See supra notes 37-39 and accompanying text. Even when the Bebchuk result holds, however, this simply means that there will be some critical level \( C^* < C \) that will determine when a defendant threat will be credible. Under the Bebchuk variation, the defendant would set the remittitur amount so that \( D - R < C^* \) to render the defendant's performance threat non-credible.
simply breach the contract and pay the plaintiff an amount much closer to 30.

Just as defendant additur could change the plaintiff's BATNA when the plaintiff was threatening an inefficient injunction, plaintiff remittitur can change the defendant's threat point. In Range 4, the defendant's best alternative to a negotiated agreement is to perform her duty, but the remittitur makes the defendant's best alternative failing to perform her duty (and paying damages) if the parties cannot agree to an alternative arrangement. Plaintiffs will only take the extreme step of committing to lower damages if they believe doing so will counteract the defendant's threat to behave inefficiently. As with defendant additur, the option is self-regulating. Plaintiffs will have no incentive to set the remittitur amount too high.

An important problem with implementation, however, concerns the procedural setting of defendant threats. Since potential defendants use the threat of performance before performance is due, potential plaintiffs under current law would not have standing to bring suit, much less announce a remittitur amount. For example, in the contractual setting, a promisor who uses a performance threat to negotiate a settlement price that is less than the expected damages would not give the promisee a basis for filing a lawsuit. The threat not only occurs before performance is due, but the threat to perform is the antithesis of anticipatory repudiation. Perversely, a promisor's threat is more akin to what the U.C.C. calls giving adequate assurance that, notwithstanding higher costs, the promisor still intends to perform when performance comes due. Yet while it is necessary for a promisee to be able to commit before a breach to seek R less than a jury would normally award, it is not necessary that the commitment be filed with a court. It is only necessary that the law allow potential plaintiffs to make irrevocable offers to reduce potential damages at the time the potential defendant threatens breach. In the end, these problems of implementing plaintiff remittitur are not of great concern—because the next section finds that even a perfectly implemented regime often does not further net equity and efficiency. Thus, while we favor defendant inalienability and additur options, we disfavor the use of plaintiff remittitur to deter defendant performance

---

75 As emphasized above, a remittitur offer, unlike a mere settlement offer, cannot be destroyed by the other side's rejection of the offer.

76 See U.C.C. § 2-609(1) (1989) ("When reasonable grounds for insecurity arise with respect to the performance of either party the other may... demand adequate assurance of due performance...").
threats.

Table 1 identified credible performance threats supported by three different damage ranges. In this section we have shown three ways to modify current remedial doctrines to discourage such threats. Giving defendants an inalienability option was sufficient to deter Range 2 plaintiff threats; this inalienability option and defendant additur together were sufficient to deter Range 1 plaintiff threats; and finally, plaintiff remittitur was sufficient to deter Range 4 defendant threats. In each of these settings, by giving the threatened party countervailing options, courts can harness private information to deter threats of inefficient performance. Rendering such threats non-credible is likely to reduce the costs of negotiation because the parties have little, if anything, about which to negotiate. Such efficiency effects are properly the topic of the next section.

III. SHOULD THREATS OF INEFFICIENT PERFORMANCE BE DETERRED?

Just because the law can be structured to deter threats of inefficient performance does not mean that the law should be so structured. Indeed, this section argues on the basis of equity and efficiency that the law should presumptively use injunctive inalienability and defendant additur to deter plaintiff performance threats, but should not try to deter defendant performance threats.

A. Equity

The strongest rationale for deterring plaintiff performance threats is the equitable impulse to deter plaintiff overcompensation. Plaintiff threats increase the risk of plaintiff overcompensation, while defendant threats reduce the risk of plaintiff overcompensation. As shown above in Table 1, plaintiffs will at times seek injunctions instead of supercompensatory damages because they expect to extract an even larger amount from defendants if injunctive performance is threatened.\(^7\) By contrast, defendant threats reduce the risk that defendants will

---

\(^7\) Table 1 also showed, however, that plaintiffs may also threaten injunctions when damages are subcompensatory (Range 1). In such circumstances, the plaintiff's threat substitutes an inequitable supercompensatory payoff for an inequitable subcompensatory payoff. However, as performance becomes increasingly inefficient (\(C > B\)), it becomes increasingly likely that plaintiff performance threats will cause a larger absolute deviation from make-whole damages (B)—that is, \((B - D) < (N - B)\).
have to overcompensate plaintiffs. Credible defendant threats of inefficient performance occur when expected damages are supercompensatory \((D > C > B)\) and have the effect of reducing the plaintiff's pay-off toward make-whole damages. Moreover, defendant threats reduce the risk of plaintiff overcompensation without increasing the risk of plaintiff undercompensation. Defendant threats are useful only in inducing plaintiffs to accept some amount between a minimum defined by the plaintiff's benefit from performance and a maximum defined by the defendant's cost of performance. There are no conditions under which a plaintiff can be induced to accept a settlement amount less than its valuation of performance, for the simple reason that a plaintiff faced with such an offer would tell the defendant to go ahead and perform.

Inalienable injunctions combined with defendant additur share the equitable traits of defendant threats: they reduce the risk that defendants will be forced to overcompensate plaintiffs without increasing the risk that plaintiffs will be undercompensated. Courts at times issue what they have every reason to believe are inefficient injunctions \((B < C)\) because they have a strong aversion to the possibility that a comparatively innocent plaintiff will be undercompensated. Giving plaintiffs the remedial choice of injunctive or monetary relief helps assure that plaintiffs will not be undercompensated, because plaintiffs can always choose the performance that is due them. The current practice of giving plaintiffs the choice of alienable injunctions, however, leads to the obverse risk—that plaintiffs may be massively overcompensated by using injunctions to bargain for settlement amounts that may be as much a function of the defendant's cost of performance as they are the plaintiff's valuation of performance. Instead of giving plaintiffs the weapon of alienability, we prefer a regime that preserves plaintiffs' unfettered rights to actual performance of the duty they are owed but which undermines their ability to bargain for a payoff in excess of the amount by which they value performance. Making injunctions inalienable does just this. Inalienability does not increase the risk that plaintiffs will be undercompensated, because they can still opt for specific performance, but it does reduce the risk that defendants will be forced to overcompensate plaintiffs by limiting plaintiffs' opportunities to use injunctive remedies merely as a strategic tool for extracting high settlements.

The foregoing discussion of over- and undercompensation implicitly assumes, however, that a plaintiff is only entitled to the "use value" of performance—that is, the benefits that flow to the plaintiff from
the defendant actually performing (or the monetary equivalent of this value). In some settings, however, it is possible to think that a plaintiff should have a right to either use or sell a particular right at stake—and that by granting an alienable injunction a court merely maintains the plaintiff's "exchange value" of the right in question. For example, if a landowner (Pedrick), after initially attempting to buy a small strip of land from his next door neighbor (Pile) then "willfully" encroaches, a court in equity may want to issue an alienable injunction to recreate the bargaining setting that the defendant wrongfully avoided. In such settings, a payment negotiated in the shadow of an alienable injunction (N) may be more equitable than a payment which more closely approximates the plaintiff's use value (B).78 Accordingly, courts might usefully distinguish between mistaken and willful encroachment—granting inalienable injunctions only for the former. More generally, "exchange value" injunctions may be appropriate whenever a plaintiff has wrongfully converted a defendant's exchange right.

 Courts should consider whether in issuing an injunction they seek to protect merely the plaintiff's "use value" or the plaintiff's "exchange value." Even though property rights are said to traditionally include the right to exclude,79 and hence the right of a property owner to resist selling to enhance her exchange value, most contracts intend to give the promisee only the use value of consideration and not the right to hold up the promisor for sums disconnected from the promisee's "expectation." In the end, distinguishing between "use value" and "exchange value" entitlements restricts the application of our analysis but still leaves a vibrant class of cases in which defendant inalienability and additur could further equity.

 Finally, there is also an equitable sense in which inalienable injunctions help deter fraud on the issuing courts. Plaintiffs seeking injunctions may represent that monetary relief could not make them whole and then figuratively turn around before the ink is dry on the injunction and start bargaining for monetary relief. The attempt to sell injunctive rights is not necessarily fraudulent; a finding of fraud would turn on the often implicit representations made during litiga-

---

78 As discussed infra at note 114, however, an injunction does not necessarily recreate the bargaining setting the parties initially faced. After the wall is built, the bargaining situation has fundamentally changed, and the resulting bargain (which now depends on the cost of wall removal) may be drastically different than it would have been earlier.

tion. Still, one can easily imagine that many courts fail to consider that injunctions they grant may be bargained away. At a minimum, it might be useful for courts in deciding whether monetary damages are inadequate to have an explicit dialogue with plaintiffs about whether they intend to sell their injunctive right and if so, for how much.

B. Efficiency

The efficiency effects of moving to a system of inalienable injunctions and defendant additur are more muddled. While our proposed reform has the salutary effect of moving the plaintiff’s expected payoff toward expectation damages, law and economics scholars in the last decade have found it increasingly difficult to conjure a single damage measure that induces both sides to behave efficiently on a variety of dimensions both before (ex ante) and after (ex post) a potential breach. This section shows that our reform proposal—giving defendants the choice of inalienability and additur—is likely to induce more efficient defendant precaution and risk allocation. We also show limited conditions under which our proposal could enhance the efficiency of plaintiff reliance and ex post bargaining.

While much of our analysis derives from the literature on efficient contract damages, the analysis can be extended to noncontractual settings as well. For example, in our earlier encroachment example, defendant precautions (in reducing C) could alternatively be modeled as precautions to avoid negligent encroachment and plaintiff reliance (in increasing B) could be modeled as actions that either mitigate or fail to mitigate the plaintiff’s damages from such encroachment. In some non-contractual contexts, however, the ex ante effects of inalienability and additur are likely to be much more attenuated. In such settings, the utility of our reform proposal turns instead on equitable and ex post bargaining considerations.

80 Judge Richard Posner is a notable exception. See supra text accompanying note 3.

81 Some scholars have shown, however, certain contexts in which parties can induce efficient investment by agreeing to alienable injunctions (ordering specific performance of the original contract if parties fail to renegotiate). See, e.g., Philippe Aghion et al., Renegotiation Design with Unverifiable Information, 62 ECONOMETRICA 257, 268 (1994) (arguing that when one party is given “the adequate choice of the default option,” and the other is endowed with all the bargaining power, efficient investments can be achieved).
1. Ex Ante Effects on Defendant Precaution, Plaintiff Reliance, and Risk

Contract remedies can affect not only who will bear the risk of inefficient performance (the risk that C will end up being greater than B) but also the size of this risk by affecting plaintiffs' and defendants' incentives to make investments that change the expected benefits and costs of performance. Damages can affect the defendant's precautionary investment in reducing C and the plaintiff's reliance investment in increasing B. This section analyzes the effect of giving defendants inalienability and additur options on the ex ante efficiency of defendant precaution, plaintiff reliance, and the parties' risk allocation.

The willingness of defendants to invest in reducing C and the willingness of plaintiffs to invest in increasing B will turn, in part, on their expectations about how their respective investments will affect the size of the payment that the defendant will make to the plaintiff if performance does not occur. If the expected payment is tied exclusively to the plaintiff's benefit and is therefore disconnected from the defendant's cost, then the defendant will have an incentive to invest efficiently in precaution, and the plaintiff will have an incentive to invest excessively (from an efficiency standpoint) in reliance. This is a slight generalization of the standard result that expectation damages lead defendant/promisors to make efficient investments to protect against unexpected increases in the cost of performance. The defendant/promisor acts efficiently because, otherwise, she will have to pay the plaintiff/promisee for the benefits the latter has forgone because of the breach. Expectation damages thus measure the real social cost of failure to perform. Tying the level of damages to the plaintiff's benefit makes the defendant a residual claimant with regard to her

---

82 Recent work has also analyzed the effect of expected damages on "cooperative" investments—i.e., the possibility that plaintiffs could invest to reduce C or that defendants could invest to increase B. See Yeon-Koo Che & Tai-Yeong Chung, Contract Damages and Cooperative Investments, 30 RAND J. ECON. 84, 85 (1999) (defining a hypothetical "cooperative" investment by the seller as one that "increases the buyer's benefit (stochastically) without lowering the seller's cost of performance").

83 See Shavell, supra note 26, at 487 ("[U]nder the expectation measure, our results indicate that to the extent that each party . . . believes he himself will default, he will engage appropriately in reliance . . ."); see also Robert Cooter & Melvin Aron Eisenberg, Damages for Breach of Contract, 73 CAL. L. REV. 1432, 1464 (1985) ("Expectation damages . . . therefore cause the promisor to internalize the cost of his failure to take adequate precaution . . . and create incentives for efficient precaution against breach.").
decision of how much to invest in reductions in C. As a residual claimant, the defendant bears the full costs and receives the full benefits of her decisions, so (to the extent she maximizes profits) she will make the socially optimal decision.

In contrast, a plaintiff anticipating expectation damages may not make the socially optimal decision. Tying the level of damages to the plaintiff's benefit externalizes the effects of the plaintiff's investment decisions. Under such a regime, in deciding whether to invest in increasing B, the plaintiff does not take into account the possibility that an increase in costs may cause performance to become inefficient. The plaintiff will receive B in kind or in money regardless of whether the defendant performs—so the plaintiff has an incentive to rely excessively.84

A converse story holds for damages that are tied exclusively to the defendant's cost and are therefore disconnected from the plaintiff's benefit. If we put aside for a moment the possibility of defendant performance threats, then setting damages equal to the defendant's cost of performance makes the plaintiff a residual claimant with regard to his investments in increasing B. The plaintiff's reliance therefore will be limited since, in the event of breach, he may find it in his best interest to take C in cash (rendering worthless any investments that were to have paid off in the event of performance). At the same time, setting damages at cost of performance means that the defendant will no longer take into account B in determining precaution, only C itself. As a result, such a regime will cause the plaintiff to invest efficiently in reliance but will cause the defendant to invest excessively in precaution. In a world without defendant performance threats, defendants realize that they will be liable for paying the cost of performance (in kind or its monetary equivalent) whether or not performance turns out to be efficient, and thus they invest too highly in precautionary measures designed to limit their ultimate cost of performance.

The important lessons here are that (1) as the size of the expected defendant payment becomes less connected to the realized cost of performance, the defendant's incentives to invest in reducing C will become more efficient; and (2) as the size of the expected defendant

84 A plaintiff excessively relies when the plaintiff makes an investment a profit-maximizing actor would not make if it had to absorb its own costs in the event of non-performance. See Shavell, supra note 26, at 487 ("[T]o the extent that each party believes he will be the victim of a breach, he will engage in excessive reliance . . . ."); see also Cooter & Eisenberg, supra note 83, at 1465-68 (discussing a party's "surplus-enhancing reliance").
payment becomes less connected to the realized plaintiff benefit from performance, the plaintiff's incentives to invest in increasing B will become more efficient. We were careful in the previous sentence to use the term "expected defendant payment" instead of "expected damages," because initial investments should turn not on nominal damages that the court might award, but, rather, on the actual payments that the parties expect in the shadow of potentially credible threats of inefficient performance. Thus, in a jurisdiction that awards cost of performance damages, the plaintiff and the defendant should expect a payment of N (to be negotiated when performance becomes inefficient) instead of a payment of C.

Understanding how the ex ante investments of the plaintiff and the defendant will be tied to their expectations about the sensitivity of ultimate defendant payments to realized levels of the costs and benefits of performance allows us to assess the efficiency of deterring plaintiff and defendant threats. Table 2 shows, for the four damage ranges initially discussed in connection with Table 1, how our proposals to deter performance threats would affect the expected defendant payment in the event that performance were to become inefficient. The table also summarizes the effects on different dimensions of ex ante efficiency. Pluses indicate that the proposed legal regime in question produces superior precaution, reliance, or risk allocation relative to the conventional regime. Minuses indicate that the alternative legal regime's expected outcomes are worse than those of the conventional regime.

---

85 It is possible to have damages that are disconnected from marginal changes in both the defendant's cost and the plaintiff's benefit. Indeed, Cooter and Eisenberg have shown that a number of common law rules—for example, denying damages for excessive reliance—as well as liquidated damage clauses may have the effect of inducing efficient ex ante investment behavior of both plaintiffs and defendants. See Cooter & Eisenberg, supra note 83, at 1467 (noting that, under existing law, "the expectation and reliance measures undoubtedly contemplate that only reasonable reliance will be compensated").

86 See supra table in text accompanying note 50 (summarizing the effects of inefficient threats in four defined damage ranges).
Table 2: Efficiency Effects of Deterring Performance Threats

<table>
<thead>
<tr>
<th>Possible Damage Ranges When Performance is Inefficient</th>
<th>(1) $D&lt;B&lt;N&lt;C$</th>
<th>(2) $B&lt;D&lt;N&lt;C$</th>
<th>(3) $B&lt;N&lt;D&lt;C$</th>
<th>(4) $B&lt;N&lt;C&lt;D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whose Threats Are Credible</td>
<td>Plaintiff Threats (potentially deterred by inalienability and additurb)</td>
<td>No Threats</td>
<td>Defendant Threats (potentially deterred by remittitur)</td>
<td></td>
</tr>
<tr>
<td>Ultimate Damage Payment Under Usual Regime</td>
<td>$N$</td>
<td>$N$</td>
<td>$D$</td>
<td>$N$</td>
</tr>
<tr>
<td>Damage Payment Under Alternative Threat Deterring Regime</td>
<td>$B + \varepsilon$</td>
<td>$D$</td>
<td>$D$</td>
<td>$C - \varepsilon$</td>
</tr>
<tr>
<td>Defendant Investment in Precaution (to reduce $C$)</td>
<td>$+$</td>
<td>$+$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plaintiff Investment in Reliance (to increase $B$)</td>
<td>$-$</td>
<td>$+?$</td>
<td></td>
<td>$+$</td>
</tr>
<tr>
<td>Risk Allocation (when at least one party is risk averse)</td>
<td>$+$</td>
<td>$+$</td>
<td></td>
<td>$-$</td>
</tr>
<tr>
<td>Ex Post Bargaining Efficiency</td>
<td>$?$</td>
<td>$+$</td>
<td></td>
<td>$?$</td>
</tr>
</tbody>
</table>
When damages are below the amount that might be negotiated in the shadow of a performance threat (Ranges 1 and 2), a plaintiff, given a traditional choice of remedies, will seek an alienable injunction to increase his expected payment to N. If, however, defendants know B, C, and D and are given the inalienability and additur choice, then they will deter the injunction threats. In Range 1, where damages are subcompensatory, the defendant will opt for inalienability and will choose an additur amount that causes the defendant's expected payment to slightly exceed the plaintiff's benefit of performance (B + ε). In Range 2, where damages are supercompensatory, the defendant will choose inalienability which, in turn, will cause the plaintiff to opt for damages of D.

Giving defendants the inalienability and additur options increases the efficiency of their ex ante investments in reducing C because defendants' expected payoffs are less correlated with the ultimate size of C. If injunctions are alienable, the defendant expects to pay N, which is explicitly a function of C; when the parties have equal bargaining power, the negotiated payment N will equal (B + C)/2. This means that the defendant's precautions reduce not only the expected cost of performance, but also the expected payment when performance does not occur. Alienable injunctions thus will lead to excessive defendant investment. Inalienability and additur regimes, in contrast, create expected payments that are unconnected to the ultimate size of the defendant's cost of performance. In Range 1, the expected payment is slightly above the plaintiff's benefit from performance, and in Range 2, the supercompensatory damage amount is, by assumption, more of an estimate (albeit overly generous) of the plaintiff's benefit than of the defendant's cost. Defendants' investments lower their costs when performance occurs but not when performance does not occur, so defendants have better incentives to take the efficient level of precaution.

Table 2 also shows why giving plaintiffs remittitur options can exacerbate defendants' incentives to invest excessively in reducing their costs of performance. When defendants can credibly threaten performance in Range 4, they expect to pay an amount (N) that is only partially responsive to reductions in C. If the Nash bargaining outcome applies, every dollar reduction in C will lead to a fifty cent reduction in N. If plaintiff remittitur is allowed, however, the defendant will expect to pay just slightly less than her cost of performance (C - ε); if the defendant reduces cost of performance (C) by one dollar, she simultaneously reduces her damage payment for failure to perform (C - ε) by one dollar. Since investments in reductions in C greatly reduce the amounts that the defendant will pay when performance does not occur, a plaintiff remittitur regime will exacerbate defendants' incentives to invest excessively in precaution.
While giving defendants inalienability and additur options unambiguously increases the efficiency of defendants' ex ante investments (in reducing C), the effect of deterring performance threats on plaintiffs' ex ante investments (in increasing B) is much more ambiguous. When court-awarded damages are expected to be less than the plaintiff's B (i.e., when damages are in Range 1), the inalienability and additur options undermine the plaintiff's incentives to invest efficiently by making ultimate damage payments more sensitive to B. In the shadow of traditional alienable injunction threats, plaintiffs expect that an investment intended to increase the value of B will be only partially recovered if performance does not occur. Under the Nash bargaining outcome, for example, plaintiffs would expect every dollar increase in B to result in only a fifty cent increase in N, the payoff if the defendant fails to perform. Under the deterrence regime, however, a defendant will have an ex post incentive to announce an additur amount that raises its expected payment slightly above the plaintiff's realized benefit of performance (B + s). Accordingly, when defendant additur is allowed, plaintiffs will expect that investing to increase B by one dollar will increase their expected payoff by one dollar whether or not performance turns out to be efficient. For this reason, deterring plaintiff threats in Range 1 exacerbates the plaintiff's incentive to over-rely.

In Range 2, however, a threat deterrence strategy may induce more efficient plaintiff investments in increasing B. The crucial question is whether the supercompensatory damages (D) are more or less sensitive to changes in B than the amount expected to be negotiated in the shadow of a performance threat (N). This will turn in part on the relative bargaining power of the parties, but it is also partially dependent on the methodology by which the court calculates damages. Damage calculations that are generous but relatively insensitive to marginal investments to increase B will tend to channel performance threats into Range 2 and give the plaintiff as well as the defendant incentives to make more efficient ex ante investments. Cooter and Eisenberg have suggested that several common law principles may already be in place to reduce the sensitivity of damages to excessive reli-

---

88 While we do not endorse giving plaintiffs remittitur options to deter defendant performance threats, our analysis suggests that the remittitur option will tend to increase the efficiency of the plaintiff's investments. The remittitur option makes the defendant's expected payments less sensitive to B than bargaining in the shadow of a credible performance threat. (This can be seen in Table 2, as (C - e) is less sensitive to changes in B than N.) As a result, plaintiffs risk losing more of their investment in B if performance becomes inefficient, and they therefore have less incentive to over-rely.
Table 2 also shows that deterring plaintiff performance threats can lead to a more efficient allocation of risk when at least one of the parties is risk averse and the inefficiency of performance is due to an unexpected increase in cost. Mitch Polinsky has shown that breach remedies allocate among the contractual parties the risk stemming from changes in circumstances. For example, Polinsky has shown that when production costs fluctuate, the optimal damage payment in terms of risk allocation will be somewhere between the plaintiff’s ex ante expected benefit from performance (B) and the defendant’s ex ante expected cost of performance (C_{ex ante}). Setting damages equal to B is one way that the plaintiff can avoid risk—because the plaintiff will receive a (monetary or in kind) benefit equaling B whether or not performance occurs, leaving the defendant to bear all the risk. If the defendant is risk averse and the plaintiff risk neutral, however, then the plaintiff should bear all the risk. Reducing damages from B to C_{ex ante} has the effect of shifting risk from the defendant to the plaintiff. If the ultimate performance cost equals C_{ex ante}, the (party who would otherwise be a) defendant will perform and get her initial expected profit; if the performance cost rises above C_{ex ante}, the defendant will breach the contract and pay damages of C_{ex ante}, once again getting her initial expected profit. If both the plaintiff and the defendant are risk averse, then the ideal damage level from a risk allocation perspective would be somewhere between C_{ex ante} and B.

Inefficient threats can produce damage payments outside of this range. A plaintiff who threatens inefficient performance, for example, will expect to receive a payment of N, which exceeds the optimal risk-allocation damage range ceiling of B. Having this option to seek

---

89 See Cooter & Eisenberg, supra note 83, at 1467 (stating that the existing practice of limiting damage awards to reasonable reliance costs makes damages somewhat insensitive to actual reliance).
90 A similar analysis applies if performance becomes inefficient due to an unexpected decrease in the benefit of performance.
91 See A. Mitchell Polinsky, Risk Sharing Through Breach of Contract Remedies, 12 J. LEGAL STUD. 427 (1983), for a discussion of optimal risk allocation under various circumstances, including different reasons for breach and different risk preferences among buyers and sellers.
92 See id. at 442 (discussing optimal damage payments when production costs fluctuate).
93 See id. at 443 (explaining that if the buyer is risk averse and the seller is risk neutral, the optimal damage payment equals the buyer’s benefit).
94 See id. at 442 (explaining that if both parties are risk averse, the optimal damage payment is between the seller’s normal production cost and the buyer’s benefit).
high damages effectively forces the plaintiff to buy and forces the defendant to sell a lottery ticket. The plaintiff bears the risk of receiving, and the defendant bears the risk of paying, a supercompensatory amount in the event performance does not occur. Deterring plaintiff performance threats with inalienability and additur options reduces the lottery component of nonperformance by reducing damages from \( N \) down to \( D \) or \( (B + \varepsilon) \) and, thus, provides a more efficient risk allocation if at least one of the parties is risk averse.

2. Ex Post Effects

Threatening inefficient performance may occasion needless costs of bargaining and, if bargaining breaks down, result in inefficient performance. These ex post bargaining inefficiencies are largely a function of the parties' potentially imperfect information. While plaintiffs are likely to know the value of receiving defendants' performance (B) and defendants are likely to know the cost of performing (C) as performance becomes due, the parties may not know each other's valuation. When plaintiffs are imperfectly informed about C and defendants are imperfectly informed about B, ex post bargaining may be inefficient.

When damages fall in Range 2 (\( B < D < N < C \)), giving defendants inalienability and additur options to deter plaintiff threats enhances bargaining efficiency. As summarized in Table 2, the defendant will opt for inalienability (but not additur), and the plaintiff will respond by electing monetary damages (D). There will be no chance of the inefficient performance actually occurring, as can happen if bargaining fails in the shadow of a credible performance threat. Additionally, the plaintiff and the defendant at most would need to bargain over any uncertainty in the ultimate size of court-awarded damages (D) and over how to split the total costs avoided by not having an actual trial. This latter negotiation is likely to be more efficient because the parties will have better information about the much smaller issues at stake.\(^95\)

When plaintiff performance threats arise in Range 1 (or when the defendant is uncertain whether B is greater or less than D), however, it is ambiguous whether giving defendants inalienability and additur options would improve ex post bargaining efficiency. For example,

\(^95\) We might, however, imagine situations in which the parties have better information about C and B than they do about what a jury might award as D.
imagine the following Range 1 scenario:

Both the plaintiff and the defendant know that the defendant’s C has unexpectedly risen to $150. The defendant is not sure of the exact value of the plaintiff’s B but she believes that it is equally likely to be any value above $0 but below $100. The parties also both know that a court would only award nominal damages which for simplicity we assume to be $0.

Under these assumptions, the plaintiff will threaten an inefficient injunction, and the defendant knows that damages are in Range 1 (D = 0 < B < 100 < C = 150). In the simpler model with full information discussed above, the defendant knew the exact size of B and so could choose an additur amount (A) such that supplemented damages would exceed the plaintiff’s valuation (D + A > B). In this scenario, the defendant can ensure an efficient result by making the injunction inalienable and choosing an additur amount of slightly more than $100. The plaintiff would then seek monetary damages, thus avoiding the possibility of inefficient performance.

The defendant, however, will often not have an incentive to announce such a high additur amount. Giving the defendant inalienability and additur options in a sense allows the defendant to make the plaintiff a take-it-or-leave-it offer. Defendants with the power to make take-it-or-leave-it offers—but who are imperfectly informed about the plaintiff’s benefit of performance—will often find it to their advantage to reduce the additur amount so that there is some chance that plaintiffs will still opt for an inefficient (and now inalienable) in-

---

96 In other words, the defendant knows only that B is uniformly distributed between $0 and $100.

97 The bargaining inefficiencies produced in this model are similar to a stylized version of the *Peevyhouse* case discussed in DOUGLAS G. BAIRD ET AL., GAME THEORY AND THE LAW 224-32 (1994). The authors examine a situation in which there are two “types” of plaintiffs: a high-valuing type who places a subjective value of $800,000 on land restoration, and a low-valuing type who values such reclamation at only $200,000. The cost of restoration is commonly known in their model to be $1,000,000. *Id.* at 224-26. As in this example, it is common knowledge that performance is inefficient, but plaintiffs with relatively high valuations nevertheless seek injunctions in equilibrium. For further discussion of this example, see Ian Ayres & Eric Talley, *Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade*, 104 YALE L.J. 1027, 1061 n.105 (1995) (discussing “exit options” in the context of Baird, Gertner, and Picker’s *Peevyhouse* example).

98 After the defendant chooses inalienability and announces an additur amount (A), the plaintiff must either accept the monetary offer (of D + A), or reject it (and thereby accept performance). Once the plaintiff takes or leaves the offer, the inalienability of the injunction ensures that there are no further negotiations.
juncture.\textsuperscript{99} Under the foregoing assumptions, for example, the defendant would minimize her expected payments by choosing inalienability and announcing an additur amount of only $75—meaning that twenty-five percent of the time the plaintiff would still choose an injunction that would result in inefficient performance.\textsuperscript{100}

When a defendant is poorly informed about the plaintiff's benefit of performance, then a defendant additur option can lead to less efficient ex post outcomes. When the defendant knows relatively more about B than the plaintiff knows about C, however, inalienability and additur options will tend to produce more efficient bargaining even in Range 1.\textsuperscript{101} In the extreme case, if the defendant knows B precisely (but the plaintiff does not know C precisely), then defendant inalienability and additur options will eliminate the chance of inefficient performance, since the defendant will name an additur amount which slightly exceeds the plaintiff's benefit of performance and thus will deter the plaintiff from opting for an inalienable injunction.\textsuperscript{102} Allowing

\textsuperscript{99} Strategic inefficiencies often arise when one side to a negotiation has bargaining power (such as having a monopoly on the right to make offers) and the other side has private information. See Ian Ayres & Robert Gertner, \textit{Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules}, 101 YALE L.J. 729, 737-42 (1992) (providing an example of a commercial carrier with market power negotiating with shippers who have private valuation information).

\textsuperscript{100} If the defendant chooses an additur A of $100, she will pay $100 in damages 100% of the time; the plaintiff's benefit from performance B is always less than $100, so he will always opt for the $100 in damages rather than the injunction. The defendant's expected payment if A is $100 will therefore be $100 = $100. In contrast, if the defendant chooses an A of $25, there is a 25% chance that the plaintiff's B will be less than A, so there is a 25% chance that the plaintiff will opt for damages at a cost to the defendant of $25. At the same time, there will be a 75% chance that the plaintiff will refuse $25 and instead demand performance, which costs the defendant $150. If the defendant offers an additur amount of $25, she can therefore expect to spend (on average) (.25 * 25) + (.75 * 150) = $118.75. In other words, under the assumptions of the above scenario, the defendant who chooses a total damage payment A between 0 and 100 will have an A% chance of paying A, and a (100 - A) % chance of paying $150. She will choose A to minimize her expected payment, which can be represented by the equation (.01A)A + (1 - .01A) 150. The A that minimizes the payment is $75. If the defendant offers the plaintiff $75 in damages, she can expect to pay out .75 * 75 + .25 * 150 = $93.75.

\textsuperscript{101} Because the effect of giving defendants inalienability and additur options in Range 1 turns on whether defendants are relatively informed about the plaintiffs' benefit of performance (B), Table 2 characterizes the effect on ex post bargaining efficiency as ambiguous.

\textsuperscript{102} Conversely, when the plaintiff knows more about the defendant's C than the defendant knows about B, giving the plaintiff a remittitur option will tend to produce more efficient ex post bargaining. When given a remittitur option, a knowledgeable plaintiff can set a remittitur amount at a level that would eliminate bargaining by ensuring that performance threats are not credible.
the plaintiff to obtain an alienable injunction, in contrast, might lead to inefficiency because a plaintiff who does not know the defendant's true cost of performance might demand such a high payment that settlement negotiations drag on or break down entirely.

In sum, granting defendant inalienability and additur options has attractive equity and efficiency characteristics. As an equitable matter, inalienability and additur options decrease the chance of overcompensation without increasing the chance of undercompensation. As a matter of efficiency, Table 2 suggests that deterring plaintiff threats will induce more efficient (1) ex ante defendant precautions and (2) joint risk-bearing. Moreover, if courts commit to damages which are generous but relatively insensitive to plaintiff reliance, deterring plaintiff threats with inalienability and additur options will also induce more efficient (3) ex ante plaintiff reliance and (4) ex post bargaining. We emphasize, however, that alternative and richer assumptions about ex ante decision-making might reverse some of these results. While comprehensive efficiency analysis is becoming increasingly difficult as lawyer/economists develop increasingly complex models, the appealing equity and efficiency attributes of the deterrence strategy make it a plausible reform candidate.

IV. MEANS OF ADDRESSING THREATS OF INEFFICIENT PERFORMANCE

A. Default Choice of Legal Regime

As we have discussed, one means to deter performance threats is to implement a legal regime allowing the defendant inalienability and additur options. Since many performance threats can arise in contractual settings, however, it is necessary to analyze default choice. For example, should the default interpretation of contracts be that injunctions were meant to be inalienable (unless the parties opt out of inalienability) or should the default interpretation be that injunctions were meant to be alienable (unless the parties opt into inalienability)? This section argues that inalienability should be a default rule—meaning that all court-ordered injunctions should be presumptively inalienable unless (1) the parties explicitly contracted for alienability of injunctions ex ante or (2) the defendant unilaterally petitioned the

103 For example, we do not consider the potential effects of inalienability and additur options on one party's decision to make investments that benefit the other party. See Che & Chung, supra note 82, at 54, for a discussion of the role of damage rules in inducing such “cooperative investments.”
court to allow any injunction to be alienable at some point in the proceeding before the plaintiff irrevocably elected injunctive or monetary relief.\(^{104}\) We do not argue that courts should change their practice about when to issue injunctions.\(^{105}\) Instead, we suggest that whenever a court decides that an injunction is appropriate, it should presumptively prohibit the defendant from discharging the injunction by paying the plaintiff.\(^{106}\)

We must first justify why our proposal should be merely a default or presumptive rule instead of a mandatory feature of all injunctions. We see no reason based on either externalities or parentalism why plaintiffs and defendants should not be free to contract ex ante for alienable injunctions.\(^ {107}\) Indeed, Table 2 shows that inalienability and additur may cause less efficient behavior on certain dimensions. For example, if damages are likely to be either undercompensatory or perfectly compensatory, then the parties may want to contract to give the plaintiff the option of seeking alienable injunctions in order to induce more efficient plaintiff reliance.\(^ {108}\) When a plaintiff’s ex ante reliance investments in increasing B are particularly salient, the plain-

\(^{104}\) As a legal matter, we would also enforce “firm” contractual inalienability provisions that waived defendants’ options subsequently to waive inalienability. As we discuss above in the text accompanying note 98, however, we predict that defendants will seldom have an incentive to waive inalienability ex post, so that such a provision would serve little purpose.

\(^{105}\) As discussed above, we are attracted both to the “undue burden” rule which tends to deny injunctions when the defendant’s cost (C) is much greater than the plaintiff’s benefit (B), see supra note 46-48 and accompanying text, and to the equitable impulse to issue inefficient (C > B) injunctions nonetheless when jury-awarded damages (D) may be undercompensatory (D < B), see supra notes 49 and accompanying text.

\(^{106}\) To the extent possible, we have also suggested that when giving plaintiffs the choice between injunctive and monetary relief, courts should strive to assure that damages will be generous. Doing so can reduce the defendant’s expected payment (compared to a regime with alienable injunctions) and has attractive equity and efficiency consequences. Specifically, the more uncertain the court’s assessment of the plaintiff’s B, the more generous the court’s assessment of damage should be.

\(^{107}\) In contrast, however, we do see externality reasons for restricting a plaintiff’s ability to unilaterally opt for alienability after a dispute has arisen. A plaintiff does not bear all the costs of bargaining or failing to reach an agreement in the shadow of an alienable performance threat, and thus may choose alienability even when doing so is socially inefficient. In fact, the plaintiff is likely to bear a much smaller proportion of the costs, since the defendant must bear the cost of the inefficient performance.

\(^{108}\) See supra text following note 87 (noting that where court-awarded damages are less than the plaintiff’s benefit, inalienability and additur options undermine the plaintiff’s incentives to invest efficiently); see also William P. Rogerson, Efficient Reliance and Damage Measures for Breach of Contract, 15 RAND J. ECON. 39, 50-51 (1984) (noting that ex post bargaining in the shadow of alienable injunctive relief may produce more efficient ex ante plaintiff reliance).
tiff and the defendant might jointly decide to contract ex ante for alienable injunctions.109

Our proposal would also allow the defendant to unilaterally opt for alienability at any stage before the plaintiff made an irrevocable election between monetary and injunctive relief.110 This unilateral option would apply not only to contractual settings, but also to any non-contractual settings—such as the encroachment example—where a court might give the plaintiff the choice of injunctive or monetary relief. Our model suggests that defendants armed with an additur option would rarely petition the court to make any prospective injunction alienable. Without the additur option, defendants in Range 1 might opt for alienability, fearing that plaintiffs would prefer an inefficient, alienable injunction to subcompensatory damages. Defendants who also have the additur option, however, can eliminate the risk of subcompensatory damages and hence their primary rationale for preferring alienability. Inalienability and additur in effect give defendants the right to make a take-it-or-leave-it offer. Game theory suggests that once a dispute arises, defendants would rarely have a rationale for forgoing the right to maintain an offer monopoly—which would have the same effect as opting for alienability. Concluding that defendants would rarely have an ex post incentive to choose alienability, however, is not a reason for limiting their ability to do so. We see neither externality nor parentalism rationales for prohibiting defendants from opting for alienability at an early stage of a dispute, and, accordingly, we would allow it. Parties already have the ability to contract ex ante to make damages the exclusive remedy in case of breach, thus prohibiting injunctions. It is arguably less of an intrusion on judicial power to give parties the limited option of making injunctions inalienable rather than the more expansive intrusion of stripping judges’ injunctive power altogether.

Finally, it is useful to justify explicitly why the inalienability default that we propose is superior to an alienability default.111 As a descrip-

109 For example, a factory contracting to buy an expensive but standardized machine may understand that there is relatively small risk that C will fluctuate but a substantial risk that B will fluctuate if the buyer is not given adequate incentives to rely.

110 As discussed above in notes 60-63 and accompanying text, a defendant must be able to commit to inalienability before the plaintiff elects a remedy in order to undermine the credibility of the plaintiff’s performance threat.

111 We also believe that the default additur amount should be zero dollars. When damages are compensatory or supercompensatory, the defendant would opt for a zero dollar default, and even when damages are likely to be undercompensatory, the court is ill-equipped to choose an additur amount.
tive matter, the alienability of current injunctions is probably a default characteristic; parties under prevailing law can contract ex ante to waive the right to settle for money in lieu of performing an injunction.\(^{112}\) This means that alienability is probably the current default, one that contractual parties could avoid by expressly contracting ex ante that any injunction that might issue would not be alienable.

We believe that an inalienability default is both more efficient and more equitable than the current alienability default. Our previous analysis of ex ante and ex post efficiency suggests that a majority of transacting parties are likely to substantively prefer inalienable to alienable injunctions. Thus, an inalienability default is likely to be the majoritarian default that minimizes the costs of contracting around. Moreover, in a world in which contracting parties rarely address whether or not injunctions will issue, much less whether such injunctions should be alienable, we predict that few parties would contract around either default. To the extent that there are some parties that would substantively prefer alienable injunctions and others that would substantively prefer inalienable injunctions, the choice of default will largely determine which type of failures to contract we see. Because we believe inalienability to produce more equitable payoffs, we also believe that failures to contract around an inalienability default promote equity, while failures to contract around alienability defaults conduce toward inequity (in the form of an enhanced probability that defendants will have to overcompensate plaintiffs).

B. Considerations in the Decision to Grant Injunctive Relief

This Article highlights several factors that should be considered in shaping injunctive relief. First, in deciding whether to grant plaintiffs the option of injunctive relief, courts should weigh the possible undue burdens that injunction performance might impose on defendants against the potential that monetary damages will undercompensate plaintiffs.\(^{113}\) Judges should continue to deny injunctions that impose undue burdens on a defendant as compared to the benefit to the

\(^{112}\) Some decisions suggest, however, that the alienability of injunctions is an immutable characteristic. See supra note 57.

\(^{113}\) This is analogous to Judge Posner's balancing approach in American Hospital Supply Corp. v. Hospital Products Ltd., 780 F.2d 589, 593 (7th Cir. 1986) (stating that a preliminary injunction should be granted if \(P \times H_p > (1 - P) \times H_d\), where \(P\) equals the probability that the denial of the preliminary injunction would be an error, \(H_p\) equals the harm the plaintiff would suffer if the injunction were denied, and \(H_d\) equals the harm the defendant would suffer if the injunction were granted).
plaintiff. Second, when issuing injunctions, courts should consider whether the injunction is meant merely to protect a plaintiff's right in its "use value" of performance or whether the injunction is meant to go further and protect the plaintiff's right to the "exchange value." The strongest justification for alienable injunctions exists when a defendant is found to have wrongfully acted to circumvent a plaintiff's right to bargain for a higher price. Thus, an injunction against a "willful" encroacher who unilaterally takes, instead of negotiating for the additional land, is more likely to be created to recreate the plaintiff's exchange value. Alternatively, an injunction against patent infringement may have not only the goal of protecting the patentee's use value (the value of practicing the invention), but also the goal of protecting the patentee's exchange value (the value of being able to negotiate for a high licensing fee by threatening not to license).

When the court is merely granting an injunction to protect a plaintiff's right to the "use value" of performance, however, there is a stronger rationale for making the injunction inalienable. Courts ordering specific performance of contractual promises, for example, are most likely trying to give the plaintiff the benefit of actual performance and not a weapon to hold up the defendant for a payment that would put the plaintiff in a position dramatically superior to that resulting from actual performance. Contractors would rarely agree that when changed conditions make the defendant's performance inefficient that plaintiffs should be able to hold up the defendant for part of the inefficiency.

---

114 A problem with "exchange value" injunctions (that is, injunctions that aim to protect the plaintiff's right to block all non-consensual entitlement transfers) is that they can also overcompensate the plaintiff once the defendant has relied to her detriment on the taking. Thus, in the willful encroachment context, the amount that the plaintiff might have bargained for ex ante (before the encroaching wall was constructed) might be much less than the amount that the plaintiff might be able to bargain for in the shadow of an injunction ordering the wall to be removed (at substantial defendant cost).

116 Injunctive orders concerning child custody are probably intended to protect parents' use value and not to create a bargaining chip that has an exchange value for the beneficiary.

118 This is especially true when the inefficiency is caused by bad news (i.e., by an increase in a seller's expected costs). When the inefficiency is caused by good news (i.e., by an increase in a seller's opportunity cost of performance), however, there is a chance that the parties would agree to share in the good news. Thus, if a seller agrees to sell a widget for $100 to an initial buyer and then, before performance, is due another buyer offers $2,000 (an amount substantially above the initial buyer's valuation), the parties might negotiate for the initial buyer to participate in the good news. However, even here, if the initial buyer is relatively risk averse, the parties would tend to place the risk of a subsequent buyer showing up on the seller. See Polinsky, supra note
Third, when giving plaintiffs an election between injunctive and monetary relief, courts should, to the extent possible, strive to make the monetary damages generous—that is, to assure that the damages are supercompensatory to the extent that they fall into Range 2. We have shown that generous damages combined with inalienable injunctions can actually reduce defendants’ expected payment by inducing plaintiffs to opt for monetary damages. Supercompensatory damages can be both more equitable (moving the defendants’ expected payment closer to make-whole compensation) and more efficient (producing more efficient ex ante incentives and ex post negotiations) than subcompensatory damages. This “generosity principle” also militates in favor of giving defendants an option to increase damages paid to plaintiffs through an additur.

C. Judicial Review of Injunctive Settlements

This Article’s analysis also suggests that, as an alternative to the inalienability and additur options, courts should review the substance of injunctive settlements and reject those settlements where defendants pay an amount above any reasonable estimate of the plaintiff’s possible benefit. Such a review might be similar to a court’s evaluation of whether a liquidated damage amount is reasonably commensurate with a plaintiff’s actual damages. Couched differently, a court might subject a proposed settlement payment (in lieu of injunctive performance) to the same remittitur analysis to which it would subject a jury award. Traditional judicial remittitur is an amount announced after

91, at 432 (discussing optimal risk allocation according to buyers and sellers’ risk aversions). If the seller agrees to sell a widget for $100 to an initial buyer and then the cost of performance rises unexpectedly to $10,000, it is less likely that the buyer and seller would agree to allow the buyer to profit from this bad news. A legal regime permitting alienable injunctions would result in a higher variance of returns for both the buyer and the seller compared to a regime that only permitted expectation damages.

177 To be consistent with our earlier analysis, courts might, before asking plaintiffs to elect a remedy, ask defendants whether any injunction settlement should be subject to remittitur review.

178 See Eric L. Talley, Note, Contract Renegotiation, Mechanism Design, and the Liquidated Damages Rule, 46 STAN. L. REV. 1195, 1200-05 (1994) (“While courts potentially can invalidate underliquidated provisions, such maneuvers are much less frequent than the invalidation of overliquidated damages.”).

179 This analysis might require that judges listen to evidence about the level of potential damages since such evidence might not have been introduced in an injunction proceeding. Because the defendant ex post may have lower incentives to contest the plaintiff’s evidence of professed benefits, it might even be appropriate for courts to assess the plaintiff’s potential benefit of performance before issuing an injunction. This evidence might already have been introduced to satisfy the equitable prerequisite
a jury has made its award, but it would be possible for courts in issuing an injunction to move first by simultaneously announcing a "settlement cap"—the maximum amount that a defendant might pay to discharge its injunctive duty. To the extent that this maximum is below the cost of performance threat point, it may mitigate the potential damage of a threat to perform inefficiently. Such a settlement cap would compress the bargaining range and, therefore, would likely produce a settlement amount that would more closely approximate make-whole compensation. Moreover, a price cap on alienable injunctions would make negotiations more efficient. For example, if it will cost a defendant $10,000 to remove an encroaching wall, but actually removing the wall will only produce a benefit of $500 for the plaintiff, then we can imagine how the parties would react to an alienable injunction with a $1000 settlement cap. The plaintiff would likely elect injunctive relief, and the parties would quickly settle for $500. Because in the absence of a settlement the injunction will be performed, the judicial settlement cap does not change the parties' threat points. Rather, it limits the maximum amount that the plaintiff can hope to take away from such a negotiation. The likely result is that the parties will settle more quickly and with fewer bargaining breakdowns since there is no risk that the plaintiff will hold out seeking too much money.

Richard Epstein has explicitly argued for analogous price constrained bargaining games in other bilateral monopoly settings. In discussing the classic necessity context of dock usage during the exigencies of a storm, Epstein considers a legal regime that would give a dock owner "the absolute right to exclude, but if he chooses to permit the use of the dock then his compensation will be limited to the dock's rental value plus the damage caused by the boat).
mit, then it can only be on condition that he accept a compensation package limited to the rental value of the dock, plus the property damages caused by the owner.\textsuperscript{123} Epstein sees that constrained price bargaining may be more efficient:

The use of this two-point distribution in effect rules out all intermediate solutions and thus makes it impossible to haggle over the price within some large range. In particular, the dockowner cannot insist on capturing the net worth of the shipowner, so that the bargaining problem is therefore effectively obviated.

A regime that legally constrains the price over which parties might bargain can thus avoid the ex post inefficiencies of bilateral monopoly haggling. When applied to the injunctive setting, capping the maximum amount that the defendant could pay in lieu of injunctive performance can not only streamline ex post negotiations, but also produce more efficient ex ante incentives and an arguably more equitable result.

Unlike our more extreme inalienability proposal, price regulation (of either the ex ante price cap or the ex post judicial remittitur kind) merely restricts the free alienability of injunctions. Restricting the price at which injunctions can be sold may also produce a more robust way to avoid inefficient injunctive performance. The defendant inalienability and additur options are crucially premised on game-theoretic predictions that plaintiffs will not seek injunctions if the benefits from actual performance can be made less than expected trial damages. The model predicts that few plaintiffs would seek injunctions once a defendant has opted for inalienability. If this deterrent effect does not come to pass, however, inalienable injunctions may lead to more inefficiency than the current alienability regime. Capping the amount that defendants can settle for may not deter as many plaintiffs from seeking inefficient injunctions, but it may lead to more efficient (and equitable) settlements in lieu of actual performance. Plaintiffs who represent to a court that they really want performance because monetary damages are inadequate have little reason to complain if the court allows the defendant to commit to a maximum price above which she cannot pay to buy back the injunction.

\textsuperscript{123} Epstein, supra note 121, at 57.

\textsuperscript{124} Id.
CONCLUSION

In this Article, we have shown that people who are owed duties (potential plaintiffs) and people who owe duties (potential defendants) may threaten inefficient performance of those duties solely to improve their individual payoffs. Potential plaintiffs will at times threaten to force defendant performance by seeking injunctions instead of accepting monetary damages, while potential defendants will threaten to perform prior to breach instead of paying monetary damages.

Naive Holmesianism would suggest that when changed circumstances make some performance inefficient, a promisor merely has the choice between performing and paying damages. In many contexts, however, the promisor has not two, but three choices to consider: performing, paying damages, or negotiating in the shadow of a credible performance threat. We have shown that the amounts paid in the shadow of performance threats by plaintiffs or defendants can differ substantially from those amounts that would be assessed as legal damages. Moreover, appreciating the third possibility of negotiating in the shadow of performance threats leads to different predictions about the parties' preferences and behavior. Once we allow for the possibility of performance threats, we see that plaintiffs at times may perversely prefer lower expected damages—because lower damages may render defendant threats of inefficient performance incredible. And defendants at times may prefer higher expected damages—because higher damages may induce plaintiffs to opt for legal instead of equitable relief.

In the contractual context, threats to breach inefficiently are thought to be presumptively opportunistic. If fishermen in the middle of an Alaskan fishing season threaten to breach their promise to fish, not because they face unexpectedly high performance costs but solely to renegotiate a higher wage, the employees' actions are roundly condemned by the common law and commentators alike as acting in bad faith.125

Yet a similar consensus has not emerged with regard to threats to perform inefficiently. If a plaintiff seeks an injunction for perform-

125 See Alaska Packers' Ass'n v. Domenico, 117 F. 99, 102 (9th Cir. 1902) ("No astute reasoning can change the plain fact that the party who refuses to perform, and thereby coerces a promise from the other party to the contract to pay him an increased compensation for doing that which he is legally bound to do, takes an unjustifiable advantage of the necessities of the other party.").
THREATENING INEFFECTIVE PERFORMANCE

ance that he knows to be inefficient when expected damages are more than compensatory, he intentionally subjects himself to both the risk of prolonged negotiations and the risk that negotiations will fail and inefficient performance will ensue. If either risk were to materialize, costs could be high for all parties involved. In considering the effects of granting Walgreen an injunction to prevent another pharmacy (Phar-Mor) from renting space at a particular mall, Judge Posner acknowledged the inefficiencies that could result from negotiations:

Suppose the cost to Walgreen of facing the competition of Phar-Mor at the Southgate Mall would be $1 million, and the benefit... of leasing to Phar-Mor would be $2 million. Then at any price between those figures for a waiver of Walgreen’s injunctive right both parties would be better off... But each of the parties would like to engross as much of the bargaining range as possible—Walgreen to press the price toward $2 million, [Phar-Mor] to depress it toward $1 million. With so much at stake, both parties will have an incentive to devote substantial resources of time and money to the negotiation process. The process may even break down, if one or both parties want to create for future use a reputation as a hard bargainer; and if it does break down, the injunction will have brought about an inefficient result.1

Judge Posner acutely understood that the costs of bargaining are likely to be positively correlated with the size of the bargaining range. When an inefficient injunction creates “a huge bargaining range[,] ... the costs of negotiating to a point within it might... [be] immense.”127 As we have seen, bargaining in the shadow of either plaintiff or defendant performance threats can create just such large bargaining ranges. Bargaining in the shadow of payments based on expectation damages (as would occur under the inalienability and additum regime in Ranges 1 or 2) can be much more efficient because the bargaining range is likely to be smaller; the money to be saved by avoiding going to court (largely attorneys’ fees) is much smaller and more easily estimated than the money to be saved from avoiding inefficient performance.

Courts award inefficient injunctions in part to reduce the possibility that relatively innocent plaintiffs will be undercompensated by monetary damages. By granting injunctions, however, courts may unwittingly increase the chance of overcompensation. Making injunctions presumptively inalienable and giving defendants the option of private additum reduces the risk of this overcompensation without in-

---

126 Walgreen Co. v. Sara Creek Property Co., 966 F.2d 273, 276 (7th Cir. 1992).
127 Id. at 278.
creasing the chance that plaintiffs will be undercompensated. Less radically, courts might simply treat proposed defendant settlements (in lieu of injunctive performance) as jury awards and subject them to similar scrutiny for potential remittitur. We have shown that while efficiency is not unambiguously improved along every dimension, on the whole presumptive inalienability and defendant additur arguably enhance both efficiency and equity. Plaintiffs who represent to a court that they really want performance because monetary damages are inadequate have little reason to complain if the court allows the defendant to commit to either inalienability or a maximum price above which she cannot pay to buy back the injunction.\textsuperscript{128} Deterring plaintiff threats of inefficient performance can potentially reduce the costs of ex post negotiations and move the ex post payments closer to make-whole compensation.

\textsuperscript{128} This Article suggests that when plaintiffs threaten inefficient performance (via injunctions) defendants should have offsetting options. Ayres & Balkin, supra note 24, at 745, analyzed the converse situation where a defendant threatened inefficient breach (via anticipatory repudiation) and suggested that plaintiffs should be given the offsetting option of being able to purchase specific performance.
SUPPLEMENTAL JURISDICTION AND SECTION 1367: 
THE CASE FOR A SYMPATHETIC TEXTUALISM

JAMES E. PFANDER

INTRODUCTION

Something appears to be going badly wrong with the interpretation of 28 U.S.C. § 1367 ("section 1367"). In the nine short years since Congress enacted it as one of several "noncontroversial" provisions of the Judicial Improvements Act of 1990, the statute that defines the supplemental jurisdiction of the district courts of the United

† Professor of Law, University of Illinois College of Law. Thanks to Steve Burbank, Ellen Deason, Rick Marcus, Tom Mengler, John Oakley, Tom Rowe, David Shapiro, Joan Steinman, and Jay Tidmarsh for helpful comments on earlier drafts. In suggesting the need for a sympathetic reading, I follow the lead of the drafters of section 1367 although my reading of the text departs from theirs in a number of particulars.


2 Before section 1367 codified them under the common label of "supplemental jurisdiction," the doctrines of "pendent" and "ancillary" jurisdiction had developed along two separate lines in the decisional law of the Supreme Court. See Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 370 (1978) (describing pendent and ancillary jurisdiction as "two species of the same generic problem"). For nice accounts of the separate but related development of the doctrines, see Richard D. Freer, A Principled Statutory Approach to Supplemental Jurisdiction, 1987 DUKE L.J. 34, and Richard Matasar, A Pendent and Ancillary Jurisdiction Primer: The Scope and Limits of Supplemental Jurisdiction, 17 U.C. DAVIS L. REV. 103 (1983). As both works suggest in their titles, the doctrines of pendent and ancillary jurisdiction often marched under the supplemental banner

(109)
States has attracted an enormous body of scholarly commentary, much of it critical of Congress's handiwork. With this body of critical commentary have come the predictable calls for reform and revision. Although defenders of the statute initially argued that the federal courts could work around the problems through flexible interpretation, the prospects for such creative solutions have diminished in re-

before they were codified in such terms in section 1367.

Professor Richard Freer was among the first and sharpest critics of the statute. See Richard D. Freer, Compounding Confusion and Hampering Diversity: Life after Finley and the Supplemental Jurisdiction Statute, 40 EMORY L.J. 445, 471 (1991) [hereinafter Freer, Life After Finley] (attacking the supplemental jurisdiction statute for “maiming efficient packaging of diversity cases . . . precluding supplemental jurisdiction in alien-age cases and confusing areas that had been relatively clear”). Professor Freer's critique drew a defense from the drafters—Tom Rowe, Steve Burbank, and Tom Mengler—and the exchange grew increasingly heated. See Thomas D. Rowe, Jr. et al., Compounding Confusion or Creating Confusion About Supplemental Jurisdiction? A Reply to Professor Freer, 40 EMORY L.J. 943, 943-44 (1991) [hereinafter Rowe et al., A Reply] (“Professor Richard Freer purports to separate the wheat from the chaff and then proceeds to torch the farm, exuberantly and extensively telling the federal courts how to get it all wrong.”); Thomas C. Arthur & Richard D. Freer, Grasping at Burnt Straws: The Disaster of the Supplemental Jurisdiction Statute, 40 EMORY L.J. 963, 963 (1991) [hereinafter Arthur & Freer, The Disaster] (“If Professor Freer in fact torched the entire farm, it is because there was so much dry straw lying around after the three drafters finished tilting with the strawmen they created in their response to Professor Freer’s article.”); Thomas D. Rowe, Jr. et al, A Coda on Supplemental Jurisdiction, 40 EMORY L.J. 993, 1006 (1991) [hereinafter Rowe et al., A Coda] (“[W]e trust that cooler heads than those of Professors Arthur and Freer will join us in resisting their call to gut or scrap a needed statute that is already proving its value.”); Thomas C. Arthur & Richard D. Freer, Close Enough for Government Work: What Happens When Congress Doesn’t Do Its Job, 40 EMORY L.J. 1007, 1007 (1991) [hereinafter Arthur & Freer, Close Enough] (“[I]f only someone had spent as much time writing the statute as the trio [of drafters] has spent writing about the statute.”). This more pointed rhetoric may have persuaded other scholars to steer clear of the controversy, at least initially. Some of the contributors to an Emory Symposium on the subject one year later appear to have worried about becoming entangled in a similar exchange. See Erwin Chemerinsky, Rationalizing Jurisdiction, 41 EMORY L.J. 3, 4 (1992) (describing both the drafting Trio and the critical Duo as right in certain respects and studiously declining to offer any opinion on “each point in the . . . debate”); Rochelle Cooper Dreyfuss, The Debate Over § 1367: Defining the Power to Define Federal Judicial Power, 41 EMORY L.J. 13, 13 (1992) (noting the “intensity” of the debate but describing it as a mistake to “join issue”).

Professors Freer and Arthur initially called for statutory revision as part of their criticism of section 1367. See Arthur & Freer, The Disaster, supra note 3, at 985 (describing a congressional fix as the only sensible course). Since then, calls for revision have occurred with great regularity. See Christopher M. Fairman, Abdication to Academia: The Case of the Supplemental Jurisdiction Statute, 19 SETON HALL LEGIS. J. 157, 190 (1994) (calling for Congress to undertake “immediate minor revisions, major deliberative actions and a revised disclaimer”); infra note 8 (listing proposed revisions). For an account of the current reform efforts of the American Law Institute, see infra notes 69-74 and the accompanying text.
cent years as the federal courts have adopted a more text-centered approach to statutory interpretation. This rigorous textualism now threatens to reshape the rules of federal jurisdiction quite dramatically and to produce results that appear very much at odds with the relatively modest expectations of the enacting Congress.

As a consequence, a growing consensus of academic opinion now holds that Congress should revise the law of supplemental jurisdiction. One recent symposium featured articles from a variety of respected scholars, many of whom agree that section 1367 requires at least a tune-up if not a more substantial legislative overhaul. This

tered the rule of Zahn v. International Paper Co.; id, at 960 (suggesting that sympathetic interpretation of the statute might solve many of the problems identified by the critics); cf. Arthur & Freer, The Disaster, supra note 3, at 983 (suggesting that the call for "sympathetic" interpretation will not overcome statutory ambiguity). For more on the problem of Zahn, see infra notes 48-54.

Recent years have witnessed a resurgence of interest in the theory of statutory interpretation. A generation ago, courts made routine use of the legislative history of federal statutes, taking for granted the idea (now associated with the Legal Process school) that statutes respond purposively to some mischief that Congress has identified. Such purposive interpretation has faced two primary challenges: one from the public choice theorists and a second from those who emphasize the centrality of the text in the interpretive process. For an overview, see William N. Eskridge, Jr. & Philip P. Frickey, An Historical and Critical Introduction to The Legal Process, in HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCEss Ii (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994). On the rise of textualism generally, and its influence on both academics and federal judges, see William N. Eskridge, Jr., Textualism, The Unknown Ideal?, 96 MICH. L. REV. 1509, 1511-14 (1998) (book review) (identifying Justice Antonin Scalia as the Court's leading textualist in the course of reviewing his essay on interpretation).

Justice Scalia placed his interpretive stamp on the law of supplemental jurisdiction in his opinion for the Court in Finley v. United States, 490 U.S. 545, 552 (1989), emphasizing the absence of any textual basis for the assertion of pendent-party jurisdiction. Although Congress adopted section 1367 in response, the subsequent interpretation of the statute owes much to the literal textualism that Justice Scalia insisted upon in Finley. See infra text accompanying notes 36-42 (describing the Finley opinion authored by Justice Scalia).

Most of this disputation has centered on the application of the new statute to actions within the diversity jurisdiction of the federal district courts. See Stromberg Metal Works, Inc. v. Press Mech., Inc., 77 F.3d 928, 931 (7th Cir. 1996) (holding that section 1367 also changes amount-in-controversy rules in ordinary multi-party diversity litigation); In re Abbott Labs., Inc., 51 F.3d 524, 529 (5th Cir. 1995) (holding that section 1367 relaxes strict rules governing amount-in-controversy determinations in multiparty class action litigation). For a more detailed account of Abbott Laboratories and Stromberg, see infra text accompanying notes 50-59. See generally Freer, Life After Finley, supra note 3, at 475-76 (objecting to the statute's restrictive approach to diversity jurisdiction).

All four of the lead authors in the symposium support statutory change. See Richard D. Freer, Toward a Principled Statutory Approach to Supplemental Jurisdiction in Diversity of Citizenship Cases, 74 IND. L.J. 5, 17 (1998) (arguing for statutory revision to broaden supplemental jurisdiction in diversity proceedings); John B. Oakley, Integr-
movement for a statutory fix coincides with last year's action of the
American Law Institute ("ALI"). Acting at its meeting in May 1998,
the ALI approved Tentative Draft No. 2 of a fully revised version of
section 1367. The ALI Draft presents a new conceptual approach to
the issues of supplemental jurisdiction and traces the implications of
its new approach in illuminating and sometimes exhausting detail.
Approval of the draft lends the ALI's prestige as an agency of law re-
form to the movement for a statutory revision.

Although amendments to the statute may indeed prove necessary,
their adoption should await the judicial consideration of an alterna-
tive approach to the interpretation of section 1367. The alternative
presented in this Article, which I refer to as "sympathetic textualism,"
represents an attempt to fuse two competing approaches to the inter-
pretation of jurisdictional law, those of the legislative historians and of
the rigorous textualists. In the wake of section 1367's adoption, opin-
ion about the workability of the statute divided into two camps. On

The American Law Institute published Tentative Draft No. 1 of its Federal Judi-
cial Code Revision Project in April 1997. Although Draft No. 1 was returned for fur-
ther drafting, the Institute approved the supplemental jurisdiction proposals con-
tained in Tentative Draft No. 2 at its meeting in May 1998. See Civil Procedure—
Supplemental Jurisdiction: ALI Advocates Proposed Amendment to Supplemental Jurisdiction
For the details of the ALI proposal, see American Law Institute, Federal Judicial
one side were the drafters of the statute and its defenders, who argued that the federal courts could resolve textual problems with the statute through reliance upon legislative history. In contrast to those who invoked legislative history, the statute’s critics insisted upon the interpretive primacy of the text and argued that the statute might well unsettle jurisdictional law. The debate between the historians and the textualists over the meaning of the supplemental jurisdiction statute corresponded to a similar debate in legisprudential circles over the role of legislative history in statutory interpretation. Indeed, it was the Court’s leading textualist, Justice Antonin Scalia, who emphasized the centrality of the text in *Finley v. United States*, the jurisdictional decision that led to the adoption of section 1367.

In calling for a sympathetic textualism, I propose a reading of section 1367 that attempts to bridge the gap between these two schools of interpretive thought. My sympathetic approach takes the expressed purpose of Congress and the history of pendent and ancillary jurisdiction as valid (if not controlling) considerations and uncovers new interpretive possibilities in the language of the statute. I thus follow to some extent the lead of the more historically minded drafters of the statute, who first called for a “sympathetic” consideration of legislative purpose in the interpretation of the statute. The textualism I advance here may also appeal to the more rigorous textual critics of the statute. For even the most committed textualist will often invoke canons of statutory construction to aid the interpretive process, as Justice Scalia’s own textualist decision in *Finley* reveals. *Finley* invoked the canon that Congress, in revising and consolidating the laws, does not intend to change their effect unless such intention is “‘clearly expressed.’” Such a rule establishes a regime of continuity with the past, very much in keeping with the insights in Professor David Shapiro’s thoughtful defense of the use of the canons in the interpretive process. The canons can thus assist the textualist, as Professor Shapiro notes, in a “sincere and sympathetic effort” to uncover the meaning of a statute by reminding us all that statutes rarely produce unan-

---

11 See *supra* note 5 (noting the drafters’ preference for a sympathetic interpretation for the statute).
12 *Finley*, 490 U.S. at 554 (quoting Anderson v. Pacific Coast S.S. Co., 225 U.S. 187, 199 (1912)).
13 *See* David L. Shapiro, *Continuity and Change in Statutory Interpretation*, 67 N.Y.U. L. REV. 921, 925 (1992) (arguing that “the dominant theme running through most interpretive [canons] that actually influence outcomes is that close questions of construction should be resolved in favor of continuity and against change”).
nounced but revolutionary changes in the law.\footnote{Id. at 926.}

My sympathetic textualism produces a reading of section 1367 that fits well both with the legislative history's expressed desire to preserve the established rules of complete diversity and with the canonical emphasis on continuity with the past. In particular, my approach reads section 1367(a) as having incorporated the joinder and aggregation rules of complete diversity into its requirement that the district courts first obtain "original jurisdiction" of the claims in a civil action. On this account, the grant of supplemental jurisdiction in section 1367(a) does not supplant diversity's joinder and aggregation rules but comes into play in diversity proceedings only after those requirements have been satisfied. Similarly, the restrictions in section 1367(b) operate to prevent the erosion of the complete diversity requirement that might otherwise result from an expansive application of what was once termed the doctrine of ancillary jurisdiction. So read, the statute leaves in place differences that had marked the pre-codification operation of the doctrine of supplemental jurisdiction in federal-question and diversity matters, and occasions none of the unexpected changes in law that the current interpretive approach ascribes to the statute. In thus proposing a reading of section 1367 that corresponds to Congress's apparent design, this Article's "sympathetic textualism" may obviate the need for further legislative tinkering and restore the courts' role in the further elaboration of supplemental jurisdictional law.

The Article develops its case for a sympathetic interpretation of section 1367 in three parts. Part I reviews the origins of the supplemental jurisdiction statute and sketches its academic and judicial reception. I show that, beneath the surface of an ongoing debate over its proper interpretation, a broad consensus has developed concerning the meaning of the statutory text. Part II of the Article presents a more sympathetic alternative to the standard account of the text. Building on the important jurisdictional distinction between federal-question cases and diverse-party controversies, and the way that distinction informed the evolution of the judge-made doctrines of pendant and ancillary jurisdiction that became supplemental jurisdiction, the Article proposes and defends an interpretation of section 1367(a) that leaves the federal courts free to apply the tenets of diversity jurisdiction as they continued to evolve before and after the statute's adoption. Part III suggests that the Article's contrast between sympathetic
and unsympathetic textualism may shed some light on current debates over the proper role of the federal courts in the interpretation of jurisdictional statutes and on the need for further reform of the kind now contemplated in the work of the ALI.

I. THE STANDARD ACCOUNT OF SECTION 1367

Although fierce academic battles have marked much of the field of supplemental jurisdiction, some uncontested terrain remains. Perhaps most importantly, one finds in the literature a virtually universal and largely unspoken consensus about the best way to understand the interplay between the first two subsections of section 1367. The Article describes this reading as the standard account of section 1367(a).

15 See supra note 3 for its collection of citations giving an overview of the battleground. One author assessed the casualties by ascribing to his mother the comment that a "lot of fur" was flying. Rowe, supra note 8, at 53.

16 One finds the assumption that the broad grant of supplemental jurisdiction in section 1367(a) operates with the same impact in both federal-question and diversity matters in virtually every published writing on the subject. See ALI DRAFT, supra note 9, at 58, 76 (showing that the reference to Rule 20 occasions a decisive narrowing of an otherwise broad grant of supplemental jurisdiction by indicating that section 1367(a) would have had a "revolutionary" effect on diversity litigation but for the restrictions in subsection (b), and that the "plain meaning" of section 1367 trumps the aggregation rules in diversity only where the plaintiffs sue a single defendant); Arthur & Freer, The Disaster, supra note 3, at 982 (indicating that the grant of jurisdiction in subsection (a) "over all claims satisfying the constitutional test for supplemental jurisdiction" may overrule Strawbridge); Freer, Life After Finley, supra note 3, at 485 (noting that section 1367(a) extends supplemental jurisdiction to the constitutional limit while section 1367(b) creates exceptions; recognizing the omission of the exception for Rule 23; and so assuming full pendent-party jurisdiction was conferred in subsection (a)); Wendy Collins Purdue, The New Supplemental Jurisdiction Statute—Flawed But Fixable, 41 Emory L.J. 69, 77 (1992) (indicating that section 1367(b) fails to answer the question of jurisdiction presented by class action joinder under Rule 23 or by nondiverse plaintiff joinder under Rule 20 and thus revealing an assumption that subsection (a) authorizes supplemental jurisdiction of such matters in diversity); Rowe et al., A Reply, supra note 3, at 960 nn.90-91 (acknowledging that section 1367 might overrule Zahn and Strawbridge and thus interpreting section 1367(a) as a broad grant of pendent jurisdiction applicable to diversity matters); Joan Steinman, Section 1367—Another Party Heard From, 41 Emory L.J. 85, 95-96 (1992) (indicating that, in diversity matters and in the absence of an exception in subsection (b), the grant of supplemental jurisdiction in subsection (a) requires not an inquiry into consistency with section 1332 but an inquiry into transactional relationship under Article III); see also Lilly, supra note 8, at 184 (describing the broad grant of supplemental jurisdiction in section 1367(a) as operating without regard to the jurisdictional basis on which plaintiff grounds the action); Stephen C. Yeazell, Teaching Supplemental Jurisdiction, 74 Ind. L.J. 241, 246 (1998) (describing the broad grant of supplemental jurisdiction in section 1367(a) as partially retracted as to diversity matters in subsection (b)).
1367. To understand the standard account and to see how it influences modern interpretations of the statute requires some background on the nature of supplemental jurisdiction. After providing the necessary background, this Part sets out the standard account of section 1367, notes its influence with courts and commentators, and shows how it informs the revision project of the ALI.

A. The Origins of Supplemental Jurisdiction

Section 1367 represents an attempt by Congress to codify the doctrines of pendent and ancillary jurisdiction under the common rubric of supplemental jurisdiction.\(^7\) The Supreme Court had developed the two doctrines in a series of decisions running well back into the nineteenth century\(^8\) without much in the way of explicit guidance from Congress\(^9\) and without identifying an entirely satisfying conceptual or statutory basis for them.\(^10\) As a consequence, the judicial doc-

\(^7\) See McLaughlin, supra note 1, at 860 (expressing approval of Congress’s “beneficial” decision to abandon the old labels of pendent and ancillary jurisdiction in favor of the generic term “supplemental jurisdiction”).

\(^8\) Ancillary jurisdiction developed first, as the Supreme Court agreed to permit the assertion of jurisdiction over claims brought by intervenors who sought to perfect claims to property other litigants had previously brought within the custody of a federal court. See, e.g., Freeman v. Howe, 65 U.S. (24 How.) 450, 460 (1860) (upholding jurisdiction over “ancillary and dependent [claims] ... without reference to the citizenship or residence of the parties”). The Court extended ancillary jurisdiction in a series of subsequent cases. See Moore v. New York Cotton Exch., 270 U.S. 593 (1926) (upholding ancillary jurisdiction over defendant’s compulsory counterclaim under state law). See generally Freer, supra note 2, at 50-53 (arguing that the development of ancillary jurisdiction served the twin goals of allowing efficient packaging of cases and avoiding duplicative litigation).

Pendent jurisdiction developed along a separate track, as plaintiffs in federal-question cases came to join state-law claims as part of their “cause of action” against the defendant. See Hurn v. Oursler, 289 U.S. 238, 247 (1933) (describing federal copyright claim and state unfair competition claims as two grounds in support of the “same cause of action”); Siler v. Louisville & Nashville R.R., 213 U.S. 175, 192-93 (1909) (holding that jurisdiction over federal due process claim encompassed related state-law claim challenging rates as having exceeded state authority); McLaughlin, supra note 1, at 870-71 (describing how pendent-claim jurisdiction developed separately from ancillary jurisdiction).

\(^9\) See JACK H. FRIEDENTHAL ET AL., CIVIL PROCEDURE § 2.12-14 (1985) (describing pendent and ancillary jurisdiction as judicially created); CHARLES ALAN WRIGHT, THE LAW OF FEDERAL COURTS 103-06 (5th ed. 1994) (describing the judicial origin of pendent and ancillary jurisdiction doctrines); cf. Freer, supra note 2, at 55 (noting the “oxymoron[ic]” tradition of regarding supplemental jurisdiction as a “common law” doctrine of “subject matter jurisdiction” and arguing that such doctrines should be seen as interpretations of the statutory grant of jurisdiction over a civil action).

\(^10\) On the lack of an adequate statutory foundation, see Freer, supra note 2, at 55.
trine of supplemental jurisdiction showed some of the messy signs of case-by-case elaboration, with curious stops and starts along the way. Although the Court adopted a rather expansive approach to pendent jurisdiction in the federal-question context of United Mine Workers v. Gibbs, its more cautious approach in Owen Equipment & Erection Co. v. Kroger sought to prevent ancillary jurisdiction from eroding the rule of complete diversity.

The contrast between Gibbs and Kroger provides a useful introduction to the conceptual underpinnings of supplemental jurisdiction and to the debate that continues to swirl around the statute. In Gibbs, the plaintiff brought suit in federal court alleging both a secondary-boycott claim under federal labor law and a state-law claim for interference with advantageous business relations. In the current parlance of supplemental jurisdiction, we would refer to the federal-question claim as “jurisdictionally sufficient,” “jurisdiction conferring,” or, in the words of the ALI Draft, “freestanding,” to convey the notion that the claim supports an assertion of original jurisdiction on its own and without regard to any other claim in the action. By way of contrast, the state-law claim lacked this jurisdictional sufficiency or freestanding quality in light of the absence of complete diversity between the plaintiff and defendants. Nonetheless, the Gibbs Court agreed that the district courts may assert what was then known as pendent-claim jurisdiction over the state-law claim. The Court reasoned that the state-law claim arose from the “same common nucleus of operative fact” as the freestanding claim and the two claims thus formed

---

21 383 U.S. 715, 721-29 (1966) (holding that pendent jurisdiction exists over state-law claims which constitute part of the same constitutional “case” as one or more federal-question claims).

22 437 U.S. 365, 373-77 (1978) (citing the complete-diversity rule in refusing to permit plaintiff to join a claim against a nondiverse, impleaded third-party defendant).

23 See Gibbs, 383 U.S. at 718-20 (describing the plaintiff’s suit and the circumstances giving rise to it).

24 Courts and commentators have struggled somewhat in their search for the best shorthand expression to capture the concept of a claim that falls within the district court’s original jurisdiction and might provide a jurisdictional anchor for the assertion of supplemental jurisdiction. See Kroger, 437 U.S. at 372 n.11 (distinguishing between “federal” and “nonfederal” claims); Palmer v. Hospital Auth., 22 F.3d 1559, 1566 (11th Cir. 1994) (distinguishing between “anchor[]” claims and “supplemental” claims); cf. McLaughlin, supra note 1, at 869 (distinguishing the “jurisdiction . . . supporting” federal-question claim in Gibbs from the “jurisdictionally insufficient” state-law claim). I agree with the ALI drafters that the term “freestanding” captures the idea of jurisdictional sufficiency as well as any. See ALI DRAFT, supra note 9, at 35-43 (defining and illustrating the difference between “freestanding” and “supplemental” claims).
a single constitutional case for purposes of Article III. The Court also noted that a decision to permit the district courts to hear the claims would serve the interests of litigant convenience and judicial economy and help to secure the just and speedy resolution of disputes.

Although similar concerns with litigant fairness and judicial economy underlay the development of ancillary jurisdiction, the Court's decision in Kroger revealed a competing concern for the complete-diversity rule. In Kroger, the plaintiff brought suit against a single diverse defendant, asserting a state-law claim for personal injuries that satisfied the amount-in-controversy requirement. (The ALI Draft would treat such a claim as "freestanding." ) The defendant impleaded a third-party defendant in accordance with Rule 14 of the Federal Rules of Civil Procedure, arguing that the new defendant might bear responsibility for some portion of any award to the plaintiff. The plaintiff responded by asserting claims against this new defendant, as Rule 14 further contemplates. Both supplemental claims—that by the defendant and that by the plaintiff—satisfied the "common nucleus" test of Gibbs and a decision to permit jurisdiction

---

25 The Gibbs Court acknowledged that the two claims might not satisfy the test of Hurn v. Oursler, 270 U.S. 593 (1926), as two grounds in support of the same "cause of action" but dismissed the Hurn test as "unnecessarily grudging." Gibbs, 383 U.S. at 725. Instead, the Court held that the district courts may assert pendent-claim jurisdiction over a state-law claim that arises from the same "common nucleus of operative fact" as the federal-question claims. Id. This "common nucleus" test has been quite widely regarded as allowing the exercise of jurisdiction over claims that satisfy many of the transactional tests set forth in the Federal Rules of Civil Procedure and as extending the scope of jurisdiction to the boundaries of Article III of the Constitution. Cf. infra note 45 (describing the contention that Article III permits district courts to hear set-off claims that do not satisfy the common-nucleus test).

26 See id. at 726 (noting that the justification for pendent jurisdiction "lies in considerations of judicial economy, convenience and fairness to litigants").

27 See Kroger, 437 U.S. at 374 (noting that under a sufficiently expansive theory of supplemental jurisdiction, "a plaintiff could defeat the statutory requirement of complete diversity by the simple expedient of suing only those defendants who were of diverse citizenship and waiting for them to implead nondiverse defendants").

28 See id. at 367.

29 See ALI DRAFT, supra note 9, at 36-37 (noting that its definition of a "freestanding" claim encompasses claims within the original jurisdiction of the district courts on the basis of both the federal-question and diversity jurisdictional grants).

30 See FED. R. CIV. P. 14(a) (providing for the defendant to implead a third-party defendant on a claim that such defendant "is or may be liable" for some portion of the plaintiff's claim against the original defendant).

31 See id. (providing that the plaintiff may assert any claim against the third-party defendant arising out of the same transaction or occurrence that is the subject matter of the plaintiff's claim against the original defendant).
SUPPLEMENTAL JURISDICTION

over the entire “case” would have served the interests of convenience and economy. But the Kroger Court refused to go so far, emphasizing the availability of state court as a convenient, alternative forum. While the Court expressed a willingness to permit the assertion of jurisdiction over the defendant’s ancillary claims against a new party under Rule 14, it refused to extend such jurisdiction to the claims of the plaintiff against the new party. The Court worried that plaintiffs might omit nondiverse defendants from their initial complaint, await their predictable impleader under Rule 14, and then amend their complaint to assert the previously omitted claims, a strategy that might undermine the complete-diversity requirement.

Although the restrictive approach of Kroger attracted some negative reviews, it was the Court’s decision in Finley v. United States that gave rise to the enactment of section 1367. The Finley plaintiff brought suit against the United States for tort damages (a “freestanding” claim under federal law) and joined state-law claims against nondiverse defendants who allegedly bore responsibility for the airplane crash that led to the litigation. Despite the obvious transactional relationship among the claims, the Finley Court refused to permit the assertion of what was then termed pendent-party jurisdiction over the plaintiff’s state-law claims against the nondiverse defendants. The

32 See Kroger, 437 U.S. at 375 n.18 (citing with apparent approval decisions that authorize the assertion of ancillary jurisdiction over claims by defendants against impleaded third-party defendants).
33 See id. at 375-76 (emphasizing that the proposed assertion of ancillary jurisdiction over a claim by a plaintiff who had voluntarily chosen the federal diversity docket differed significantly from that over a claim by a defendant “haled into court against his will”).
34 See id. at 377 (“To allow the requirement of complete diversity to be circumvented as it was in this case would simply flout the congressional command.”).
35 See Freer, Life After Finley, supra note 3, at 459-61 (arguing that Kroger was wrongly decided); McLaughlin, supra note 1, at 880 (arguing that Kroger failed to define the “permissible limits of supplemental jurisdiction”).
36 490 U.S. 545 (1989). All of the scholarly and judicial commentary on the statute recognizes the decisive role that Finley played in leading Congress to adopt section 1367. See, e.g., CHARLES ALAN WRIGHT ET AL., 13 FEDERAL PRACTICE AND PROCEDURE: JURISDICTION AND RELATED MATTERS 2d, § 3523, at 65 (Supp. 1999) [hereinafter FEDERAL PRACTICE] (ascribing the origins of section 1367 to the decision in Finley); McLaughlin, supra note 1, at 889 (same).
37 See Finley, 490 U.S. at 546 (noting that plaintiff sued the United States under the Federal Tort Claims Act (“FTCA”) and then “moved to amend the federal complaint to include claims against the original state-court defendants, as to which no independent basis for federal jurisdiction existed”).
38 See id. at 553, 556 (stating that “[t]he statute here defines jurisdiction in a manner that does not reach defendants other than the United States” and declining to
Finley Court acknowledged the strength of the arguments of litigant convenience and judicial economy, arguments made stronger by the plaintiff's inability as in Kroger to join claims against all the defendants in a single state court proceeding. But the Court nonetheless refused to approve jurisdiction over the pendent parties, citing the absence of any written statutory authorization for such jurisdiction. Although the Court distinguished Gibbs and the situation of pendent-claim jurisdiction, its emphasis on the absence of a statute appeared to threaten many established forms of supplemental jurisdiction over additional parties.

permit the assertion of pendent-party jurisdiction).

See id. at 555 (acknowledging that its decision would sacrifice the efficiency and convenience of litigating multiple claims in a single action); cf. id. at 555-56 (refusing to adopt dicta in Aldinger v. Howard, 427 U.S. 1, 18 (1976), which had hinted that the exclusive jurisdiction conferred in the FTCA over claims against the United States provided a strong argument for the exercise of pendent-party jurisdiction).

See id. at 552 (emphasizing that the jurisdictional grant in question spoke of "civil actions on claims ... against the United States" and did not, by its terms, authorize claims against anyone else).

See id. at 549-51 (emphasizing the fundamental analytical difference between the pendent-claim jurisdiction in Gibbs and the joinder of parties not named in any claim that is independently cognizable); id. at 556 (describing Gibbs as "a departure from prior practice" that the Court had no intent to "limit or impair").

See Thomas M. Mengler, The Demise of Pendent and Ancillary Jurisdiction, 1990 BYU L. REV. 247, 258-60 (arguing that the rationale of Finley threatened ancillary jurisdiction as well as the pendent-claim jurisdiction recognized in Gibbs); cf. McLaughlin, supra note 1, at 887-89 (noting that Finley threatened Gibbs by characterizing the decision as a departure from prior practice and summarizing post-Finley judicial decisions).

Apart from the threat, Finley also issued an invitation to Congress. By emphasizing the need for jurisdiction conferred by "written law," Finley, 490 U.S. at 547 (quoting Ex parte Bollman, 8 U.S. (4 Cranch) 75, 93 (1807)), the Court obviously contemplated the possibility that Congress might supply the law in question. The Court also supplied a set of "interpretive rules" to enable Congress to "know the effect of the language it adopts." Finley, 490 U.S. at 556. The rules appeared earlier in the opinion, in the course of its rejection of the argument that existing statutory provisions conferred jurisdiction in terms flexible enough to support an exercise of pendent jurisdiction. This argument for flexible interpretation rested upon the recodification of the jurisdictional statutes in 1948 against the backdrop of decisions that had expanded the scope of jurisdiction over civil actions to take account of the growth in the litigation unit reflected in the Federal Rules. But the Finley Court rejected this flexible approach, noting that the real growth in pendent jurisdiction came in Gibbs—a decision that came down well after the 1948 recodification had occurred. See id. at 555 (noting that the liberalization of pendent-party jurisdiction effected by Gibbs occurred "nearly 20 years later" than the recodification the plaintiff relied on).
B. The Text and Meaning of Section 1367

Congress responded with section 1367. As commonly interpreted, the statute confers a broad grant of supplemental jurisdiction in subsection (a), allowing the district courts to hear nonfederal claims that bear an appropriately close relationship to the claims over which the court has original jurisdiction. The statute follows Gibbs in defining the scope of supplemental jurisdiction and expressly permits the assertion of jurisdiction over pendent parties, thus supplying the

---

Section 1367 reads as follows:

(a) Except as provided in subsections (b) and (c) or as expressly provided otherwise by Federal statute, in any civil action of which the district courts have original jurisdiction, the district courts shall have supplemental jurisdiction over all other claims that are so related to claims in the action within such original jurisdiction that they form part of the same case or controversy under Article III of the United States Constitution. Such supplemental jurisdiction shall include claims that involve the joinder or intervention of additional parties.

(b) In any civil action of which the district courts have original jurisdiction founded solely on section 1332 of this title, the district courts shall not have supplemental jurisdiction under subsection (a) over claims by plaintiffs against persons made parties under Rule 14, 19, 20, or 24 of the Federal Rules of Civil Procedure, or over claims by persons proposed to be joined as plaintiffs under Rule 19 of such rules, or seeking to intervene as plaintiffs under Rule 24 of such rules, when exercising supplemental jurisdiction over such claims would be inconsistent with the jurisdictional requirements of section 1332.

(c) The district courts may decline to exercise supplemental jurisdiction over a claim under subsection (a) if—(1) the claim raises a novel or complex issue of State law, (2) the claim substantially predominates over the claim or claims over which the district court has original jurisdiction, (3) the district court has dismissed all claims over which it has original jurisdiction, or (4) in exceptional circumstances, there are other compelling reasons for declining jurisdiction.

(d) The period of limitations for any claim asserted under subsection (a), and for any other claim in the same action that is voluntarily dismissed at the same time as or after the dismissal of the claim under subsection (a), shall be tolled while the claim is pending and for a period of 30 days after it is dismissed unless State law provides for a longer tolling period.

(e) As used in this section, the term "State" includes the District of Columbia, the Commonwealth of Puerto Rico, and any territory or possession of the United States.

---

4 See, e.g., Lilly, supra note 8, at 184 (describing the statute as opening in subsection (a) with a "broad conferral of supplemental jurisdiction" and following that grant with limits in subsection (b) on the scope of such jurisdiction in diversity matters); Yeazell, supra note 16, at 246 (explaining that the statute "establishes a broad jurisdictional grant in subsection (a) and then retracts much of that grant in subsection (b), whose restrictions apply to diversity-only cases").
statutory underpinning that the Finley Court identified as missing. Subsection (b) seeks to preserve certain of the limitations that had developed on the exercise of supplemental jurisdiction in the diversity cases. In particular, subsection (b) codifies the rule of Kroger by foreclosing the assertion of jurisdiction over claims by plaintiffs against persons made parties under Rule 14. Subsection (c) codifies the discretionary factors that the Gibbs Court had directed district courts to consider in determining whether to assert supplemental jurisdiction.

Although section 1367 has generated an enormous body of scholarly literature and a raft of conflicting lower court decisions, the proper reading of its text has become a matter of quite widespread consensus. Both the statute's detractors and defenders agree that subsection (a) provides a broad grant of supplemental jurisdiction that applies to both federal-question and diversity proceedings and extends to the limits of Article III. Both camps agree that the exceptions to this grant of supplemental jurisdiction appear in subsection (b) and operate primarily in diversity litigation. But despite widespread agreement about the interplay between subsections (a) and (b)—what I call the standard account of section 1367—scholars and courts disagree about the legal effect of the enacted words. Critics charge that the plain meaning of the statute makes or threatens a series of sweeping changes in the law of supplemental jurisdiction; defenders cite legislative history that clearly shows a desire on the part of Congress simply to overrule Finley and codify most of the pre-Finley status quo.

46 For useful discussions of the relationship between the constitutional limits of supplemental jurisdiction and the "common nucleus" test of Gibbs, see William A. Fletcher, "Common Nucleus of Operative Fact" and Defensive Set-Off: Beyond the Gibbs Test, 74 IND. L.J. 171 (1998), which argues in favor of the existence of constitutional power to adjudicate an unrelated defensive set-off, despite the absence of any common nucleus among the freestanding and supplemental claims. See also McLaughlin, supra note 1, at 890-95 (explaining the relationship between the constitutional case or controversy requirement and the supplemental jurisdiction statute).

47 On the statute's preservation of the rule of Kroger, see McLaughlin, supra note 1, at 936-40. Critics of the statute admit that it preserves the rule of Kroger, and focus their criticism on the decision to do so. See, e.g., Arthur & Freer, The Disaster, supra note 3, at 975-78 (arguing that Kroger was wrongly decided and that the statute not only preserved but extended its wrongheaded features); Freer, supra note 8, at 13-15 (same).

48 As the Supreme Court explained, subsection (c) "codifie[s] the principles" of economy, convenience, fairness, and comity that inform the discretionary regime of Gibbs. City of Chicago v. International College of Surgeons, 522 U.S. 156, 165 (1997). For doubts about the effectiveness of the codification, see ALI DRAFT, supra note 9, at 78-95 (arguing that the statute may have invited lower federal courts to refrain from hearing claims, such as "freestanding" and ancillary claims, to which the regime of discretion should not apply).
One can best understand this familiar debate over the proper roles of text and legislative history in statutory interpretation by considering the impact of section 1367 on the amount-in-controversy rule of *Zahn v. International Paper Co.* The *Zahn* Court held that, even where the claims of the named plaintiffs meet the threshold amount, diversity jurisdiction does not extend to those unnamed members of a plaintiff class whose individual claims fail to meet the amount-in-controversy requirement. The standard account of the statute holds that the plain meaning of section 1367 alters the outcome in *Zahn*. After jurisdiction attaches to the jurisdictionally sufficient claim of a single, named class representative (a "freestanding" claim), section 1367(a) provides supplemental jurisdiction over the jurisdictionally insufficient but related claims of the additional class members. Moreover, since 1367(b) does not specify an exception for claims joined under Rule 23 (class actions), the broad grant in subsection (a) controls the outcome.

Although the argument that the text of section 1367 overrules *Zahn* first appeared in Professor Freer’s critique of the statute, the Fifth Circuit has given it the force of law. Its decision in *In re Abbott Laboratories, Inc.* features the standard textual argument in the heart of its opinion: "Section 1367(a) grants district courts supplemental jurisdiction over related claims generally, and § 1367(b) carves exceptions. Significantly, class actions are not among the exceptions." Having made the textual argument, the *Abbott Laboratories* court acknowledged that the House had included a statement in the legislative history to the effect that the section was not intended to overrule *Zahn*. But the court refused to look behind the text, concluding that section 1367 authorized the assertion of jurisdiction rejected in *Zahn*.

---

49 See id. at 301 (holding that the district courts, sitting in diversity, may not assert jurisdiction over the claims of unnamed class members who fail to meet the amount-in-controversy requirement).
50 See Freer, *Life After Finley*, supra note 3, at 485-86 (criticizing the statute as a threat to *Zahn* and implying that the disclaimer of this interpretation in the legislative history is an insufficient precaution); cf. Freer, *supra* note 8, at 18 (opining that "[a]ll observers agree that the supplemental-jurisdiction statute, on its face, overrules *Zahn*.")
51 51 F.3d 524 (5th Cir. 1995).
52 Id. at 527.
53 See id. at 528 (citing H.R. REP. NO. 101-734, at 29 (1990), reprinted in 1990 U.S.C.C.A.N. 6860, 6875 n.17, which, the court states, "cites *Zahn* as a pre-*Finley* case untouched by the Act").
54 See id. at 528-29 (stating that "[w]e are persuaded that under § 1367 a district court can exercise supplemental jurisdiction over members of a class, although they
The Fifth Circuit expressly refused to enter into a discussion of the "wisdom of Zahn"; that was "not our affair."55

With the growing emphasis on textualism, and the accompanying distrust of legislative history, the widespread acceptance of the standard account of section 1367 appears likely to threaten a number of established jurisdictional rules. In Stromberg Metal Works, Inc. v. Press Mechanical, Inc.,56 the Seventh Circuit followed the lead of the Fifth in holding that section 1367 overrules the jurisdictional rule of Clark v. Paul Gray, Inc.57 The Clark court held that diversity jurisdiction attaches only to the claims of plaintiffs, joined under Rule 20, whose individual claims meet the amount-in-controversy threshold.58 The Stromberg court concluded that section 1367 altered that result by conferring a supplemental jurisdiction on the district courts broad enough to encompass the related, but jurisdictionally insufficient, claims of additional plaintiffs.59

Much the same textualism informs the Fourth Circuit's decision that section 1367 overrules the legal-certainty rule of St. Paul Mercury Indemnity Co. v. Red Cab Co.60 The St. Paul court held that diversity jurisdiction attaches at the threshold to claims that the plaintiff asserts in the complaint, unless it appears to a legal certainty that the claims will not satisfy the jurisdictional amount.61 In Shanaghan v. Cahill,62 the Fourth Circuit read section 1367 as supplanting the St. Paul rule. In particular, the Fourth Circuit concluded that the summary disposition of one of three aggregated claims—a disposition that brought the amount claimed below the jurisdictional amount—brought into play the discretionary power of the district court to decline jurisdiction over the remaining claims.63 The Fourth Circuit acknowledged that

---

55 Id. at 529.
56 77 F.3d 928 (7th Cir. 1996).
58 See Clark, 306 U.S. at 589 (holding that a claim asserted in a diversity case must separately meet the amount-in-controversy requirement).
59 See Stromberg Metal Works, Inc., 77 F.3d at 931.
60 303 U.S. 283 (1938).
61 See id. at 288.
62 58 F.3d 106 (4th Cir. 1995).
63 See id. at 112 (holding that "[i]f some event subsequent to the complaint reduces the amount in controversy, such as the dismissal of one count . . . , the court must then decide in its discretion whether to retain jurisdiction over the remainder of the case"). See also Stevenson v. Severs, 158 F.3d 1332, 1334 (D.C. Cir. 1998) (holding that when dismissal of claims reduces the amount in controversy below the statutory minimum,
this discretionary regime departed from the old rule, which had treated the amounts claimed in the complaint as controlling and ignored actual awards in determining whether the plaintiff satisfied the jurisdictional amount requirement. Nonetheless, the court concluded that section 1367 had effected a change in the law by making the discretionary regime of subsection (c) applicable to such aggregated claims.64

Scholars have identified a good many other settled jurisdictional rules that the statute, read in keeping with the standard account, might alter. Without elaborating all of the potentially affected areas, it seems plain that section 1367 might alter the complete-diversity rule of Strawbridge v. Curtis65 and could modify in important respects the manner in which supplemental jurisdiction operates in removed cases.66 The Court itself has already interpreted section 1367 as conferring jurisdiction on the district courts to hear pendent claims in the nature of cross-system petitions for appellate review of state administrative proceedings, a decision driven by the kind of textualism that informs the standard account.67 As a result of these and other possible changes, scholars who initially defended the statute against charges of ambiguity have now reluctantly concluded that some re-drafting may be necessary.68

"the district court has discretion to entertain the remaining claim if it so chooses," pursuant to section 1367(c)'s discretionary dismissal provision. But see Wolde-Meskel v. Vocational Instruction Project Community Serv., Inc., 166 F.3d 59, 65 (2d Cir. 1999) (holding that "[w]hen state law claims are aggregated, regardless of the amounts at issue, all of them together are 'original,' and none of the constituent claims are 'supplemental,'" making section 1367(c)'s discretionary-dismissal regime inapplicable).64

See Shanaghan, 58 F.3d at 111 (holding that "the strict St. Paul rule is inconsistent with the statutory framework of § 1367 and so must be modified to fit the contemporary congressional view of federal jurisdiction").

65 7 U.S. (3 Cranch) 267 (1806). The complete-diversity rule of Strawbridge v. Curtis holds that citizens of the same state may not appear on opposing sides of a diversity proceeding. See id. The failure of section 1367(b) to include the joiner of additional plaintiffs under Rule 20 could support a textual argument that the statute overrules Strawbridge. See, e.g., Rowe et al., A Reply, supra note 3, at 961 n.91 ("We can only hope that the federal courts will plug that potentially gaping hole in the complete diversity requirement . . .").

66 For a summary of the questions that scholars have raised about the application of section 1367 to actions removed to federal court, see infra notes 151-56 and accompanying text.

67 See City of Chicago v. International College of Surgeons, 522 U.S. 156, 168-69 (1997) (holding that a pendent claim seeking appellate review of a local administrative decision may fall within section 1367's grant of supplemental jurisdiction).

68 See, e.g., Rowe, supra note 8, at 53-54 (suggesting that the time has come to shift the focus of the debate regarding section 1367 from the problems with the current statute to the best way to redraft it). But cf. id. at 57-58 (worrying cogently that highly
C. The Proposal of the American Law Institute

The movement for reform received a boost from the decision of the American Law Institute to approve a new and more detailed draft of section 1367 (the "ALI Draft") for possible enactment into law by Congress. According to the project's Reporter, John B. Oakley, the ALI Draft of section 1367 proposes to reconceptualize the interplay between original and supplemental jurisdiction. Based upon the insight that the rules now governing the joinder of claims and parties in diversity jurisdiction are themselves rules of supplemental jurisdiction, the ALI Draft proposes a more general approach that distinguishes between "freestanding" and "supplemental" claims. Then, with admirable rigor and attention to detail, the ALI Draft specifies a set of rules to govern the exercise of supplemental jurisdiction. In the course of the work, the ALI Draft revises and extends the criticisms that others have made of the current version of section 1367.

Although they defy easy summary, the 165 pages that comprise the ALI Draft and its accompanying commentary, memoranda, and appendices reveal much the same commitment to a rigorous textualism that has characterized the Abbott Laboratories and Stromberg decisions.

literal textualism may undermine the creation of a sound, practical relationship between the federal courts and Congress.

As the ALI Draft acknowledges, reconceptualization comes at a price. In a candid and, to my mind, accurate description of the likely reaction of many judges and practicing lawyers, the draft admits that readers who try to take in the complexity of the statute may feel some "indigestion." ALI DRAFT, supra note 9, at xvii.

See id. (concluding that the rule of complete diversity is not in tension with the concept of supplemental jurisdiction but is a rule of supplemental jurisdiction that restricts its operation in diversity cases).

According to the Draft, "freestanding" claims are those that come "within the original jurisdiction of the district courts without reliance upon supplemental jurisdiction." Id. "Supplemental" claims are not freestanding but they bear a relationship to them such that, together, they form a single case or controversy within the meaning of Gibbs. See id. at 35-39 (setting forth the text of a proposed replacement for section 1367(a) which distinguishes in a definitional section between freestanding and supplemental claims). By breaking down the jurisdictional analysis into what it calls claim-specific terms and by treating both federal-question claims and diversity claims as potentially "freestanding," the ALI Draft makes clear that it intends to extend jurisdiction on a conceptually similar basis to supplemental claims in both federal-question and diverse-party proceedings. See id. at 31-37 (laying out the basis for a reconceptualization of supplemental jurisdiction).

To see that the ALI Draft proceeds upon the same interpretative assumption as the Abbott Laboratories and Stromberg courts, consider its discussion of proposed section 1367(c), which performs the same function as current section 1367(b) in reining in the grant of supplemental jurisdiction in an earlier section. See id. at 58-59. In the
The draft itself acknowledges as much, in the course of questioning the legitimacy of "the sort of pragmatic discretion rather than express legislative command that long colored the whole realm of diversity jurisdiction."73 Viewing the days of such pragmatic discretion as numbered, the ALI Draft notes that "[t]he legitimacy of this pragmatic discretion was questioned in [Finley], and the enactment the following year of present § 1367 substituted a new regime of close attention to the literal text and plain meaning of the statutory conferral of supplemental jurisdiction."74 The ALI Draft carries this new regime of close attention to literal text to its logical conclusion, specifying in some detail the way in which its rules of supplemental jurisdiction play out in a variety of different contexts. One can quibble with certain of the choices in the ALI Draft, and an enacting Congress might well tinker with its provisions. Ultimately, though, a decision to adopt something like the ALI Draft would move the law of federal jurisdiction decisively away from a reliance upon pragmatic discretion and decisively into the realm of textual literalism.

Perhaps the combination of Finley and section 1367 leaves us with no alternative to an increasingly detailed jurisdictional code. But before we endorse the textual literalism of the ALI Draft, this Article proposes that we explore a more sympathetic approach to the interpretation of current section 1367. Such an approach may make it possible to retain the statutory underpinning of supplemental jurisdiction and to preserve some portion of the pragmatic discretion that had informed the evolution of the rules of diversity before the codification. It certainly avoids the textualist overkill of such cases as Abbott Laboratories and Stromberg.

II. TOWARD A MORE SYMPATHETIC READING OF SECTION 1367

Carefully read, section 1367 reveals no clearly expressed intention to alter the laws of diversity jurisdiction. Indeed, quite the contrary. Section 1367(a) appears to assume that the existing rules of complete diversity will continue to apply and that the grant of supplemental ju-

73 ALI DRAFT, supra note 9, at 65.
74 Id.
risdiction will come into play only after the plaintiff has submitted claims that properly invoke such original jurisdiction. Consider again the language of the first sentence of subsection (a):

Except as provided in subsections (b) and (c) . . . in any civil action of which the district courts have original jurisdiction, the district courts shall have supplemental jurisdiction over all other claims that are so related to claims in the action within such original jurisdiction that they form part of the same case or controversy.\textsuperscript{75}

Literally read, this language provides for the assertion of supplemental jurisdiction only as to claims that bear a transactional relationship to the "claims" asserted in a "civil action" of which the district courts have "original jurisdiction." The statute thus appears to distinguish between joinder and aggregation issues that inform the existence of original jurisdiction and those that operate as part of the district courts' supplemental jurisdiction.

In this textual distinction lies the key to the sympathetic reading of section 1367. Before \textit{Finley}, the Court had developed doctrines of pendent and ancillary jurisdiction that operated quite differently in federal-question and diversity proceedings. In federal-question cases, original jurisdiction attached to any well-pleaded complaint that asserted a substantial federal claim; pendent jurisdiction came quickly into play to govern the plaintiff's initial joinder of additional claims along the lines the Court developed for pendent claims in \textit{Gibbs} (but rejected for pendent parties in \textit{Finley}). In the pre-\textit{Finley} world of diversity, by contrast, the rules of original jurisdiction (rather than of supplemental jurisdiction) governed a broad range of initial joinder and aggregation questions.\textsuperscript{76} Only after original jurisdiction attached in accordance with these fairly elaborate rules of complete diversity did the doctrine of ancillary jurisdiction come into play. One can thus read section 1367(a) as having incorporated the rules of complete diversity into the statute's requirement that the district courts first obtain original jurisdiction of the cause. Sympathetically read, the statute would overrule \textit{Finley} in federal-question cases but would still enable the federal courts to retain the pre-\textit{Finley} rules of diversity jurisdiction, in keeping with the views outlined in the House Report that accompanied the statute.\textsuperscript{77}

\textsuperscript{75} See supra note 43 (providing the text of section 1367).

\textsuperscript{76} For a summary of the rules that govern the determination of diversity of citizenship and amount in controversy, see infra notes 77-87 and accompanying text.

\textsuperscript{77} See H.R. REP. NO. 101-734, at 27-29 (1990) [hereinafter HOUSE REPORT] (explaining the need for and effects of the codification of supplemental jurisdiction). For
This Part of the Article presents the argument for a sympathetic reading of section 1367. It begins with a review of the rules of pendent and ancillary jurisdiction that developed before Finley, emphasizing the differing operation of those rules in federal-question and diversity proceedings. Next, this Part considers both textual and structural evidence that section 1367 preserves and incorporates these pre-Finley differences. Finally, this Part tests the sympathetic reading in light of predictable arguments against its adoption and concludes that the sympathetic reading outperforms the standard account of the statute.

A. Pre-Finley Distinctions in the Operation of Supplemental Jurisdiction

Despite conceptual similarities, the pre-Finley doctrines of pendent and ancillary jurisdiction remained quite distinct in the work of the Supreme Court. In addition to its preservation of the nominal distinction between pendent and ancillary jurisdiction, the Court flatly refused to apply pendent jurisdictional concepts to cases in diversity. Instead, the Court continued to apply rules of law that it had developed in the interpretation of 28 U.S.C. § 1332's (“section 1332”) provision for the exercise of diverse-party jurisdiction. As a practical

the HOUSE REPORT's discussion of the preservation of diversity, see infra notes 104, 112.

78 See supra text accompanying note 18 (discussing the Supreme Court's development of the doctrines of pendent and ancillary jurisdiction in a series of decisions running well back into the nineteenth century); see also PAUL M. BATOR ET AL., HART & WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 1685 n.1 (3d ed. 1988) (noting the distinction between pendent and ancillary jurisdiction: pendent jurisdiction applies to efforts by plaintiff "as in Gibbs to join with a federal claim a nonfederal claim over which the court has no independent basis of jurisdiction" while ancillary jurisdiction applies "with respect to claims asserted after the filing of the original complaint").

79 It was broadly recognized that the pre-Finley federal-question doctrine of pendent jurisdiction did not apply to cases based on diversity of citizenship. See WRIGHT, supra note 19, at 158 (noting that most pre-Finley courts "recognized that whatever the merits, or lack thereof, of the concept of pendent jurisdiction to bring in additional parties in cases in which there is a federal question or to overcome problems of amount in controversy, it was not properly used to avoid the longstanding requirement of complete diversity"); McLaughlin, supra note 1, at 869 (noting that in pre-Finley pendent-claim jurisdiction the freestanding claim was a "federal law claim" and distinguishing ancillary jurisdiction as applicable to both federal-question and diversity jurisdiction).

80 See, e.g., Zahn v. International Paper Co., 414 U.S. 291, 294-98 (1973). The Zahn Court concluded that its interpretation of the term "matter in controversy" in the diversity statute required each member of the class to set up a claim that met the amount-in-controversy requirement and so refused to apply the doctrine of supple-
matter, the refusal to import pendent jurisdiction concepts into diversity litigation meant that issues of transactional relationship and litigation convenience—the coin of the realm for supplemental jurisdiction—had far less to do with the scope of the claims a plaintiff might permissibly join in a diverse-party proceeding than the established rules of aggregation and complete diversity that the Court had worked out long before Gibbs came down in 1966.

Consider first the well-established complete-diversity requirement of Strawbridge v. Curtiss.\footnote{7 U.S. (3 Cranch) 267 (1806).} Chief Justice Marshall’s cryptic opinion in Strawbridge has come to stand for the proposition that citizens of the same state may not appear on opposing sides of an action within the diversity jurisdiction of the federal trial courts.\footnote{See id. (holding that jurisdiction based on diversity requires “complete diversity”); see also WRIGHT, supra note 19, at 156 (illustrating the complete-diversity rule).} Although it applies chiefly to section 1332 and does not control the scope of Congress’s power under the diversity grant in Article III,\footnote{See, e.g., State Farm Fire & Cas. Co. v. Tashire, 386 U.S. 523, 530-31 (1967) (holding that the complete-diversity rule does not apply to actions brought under the authority of the interpleader statute, 28 U.S.C. § 1335, and upholding the constitutionality of the minimal diversity approach of the interpleader statute).} the complete-diversity rule operated to preclude the assertion of many forms of pendent-party jurisdiction in the diversity context. Thus, a citizen of Illinois may sue a citizen of Texas and properly invoke diversity of citizenship as the basis of jurisdiction. But the Illinois citizen may not join an additional plaintiff from Texas, or an additional defendant from Illinois, even if the claims by and against these additional parties would satisfy the transactional tests of Rule 20 and Gibbs.\footnote{See WRIGHT, supra note 19, at 156 (illustrating the operation of the complete-diversity rule in similar terms).} In a complete-diversity inquiry, then, pendent-party jurisdiction would have no application.

The rules of aggregation similarly confirm the inapplicability of pendent jurisdiction to claims in diversity. Consider first the aggregation rule that permits a plaintiff to join a series of claims against a diverse defendant to satisfy the jurisdictional amount. Quite in contrast to the rules of pendent jurisdiction in Gibbs, the aggregation rules do not require any transactional relationship among the aggregated claims.\footnote{See id. at 210 (noting the absence of any requirement that claims aggregated to meet the amount-in-controversy threshold satisfy a test of transactional relationship).} Nor do they require that any one of the aggregated claims satisfy the jurisdictional threshold on its own; the law requires only
that the total value of all of the claims meets or exceeds the statutory minimum. Transactional relationship was equally irrelevant to the rules of aggregation that governed the joinder of additional parties. In Snyder v. Harris, the Court ruled that plaintiffs in a class action under Rule 23 may not aggregate the value of their several transactionally related claims to meet the amount-in-controversy requirement. The Court expressly based its decision on the statutory reference in section 1332 to the "matter in controversy" and held that each plaintiff's claim must meet the minimum amount. Similarly, as noted above, the decisions in Clark and Zahn preclude the assertion of jurisdiction over the jurisdictionally insufficient claims of co-plaintiffs, joined under Rules 20 and 23 respectively, even where one of the plaintiffs asserts a claim that meets the jurisdictional threshold. Like Snyder, Zahn refused to adopt a test of ancillarity, choosing instead to rely upon the established jurisdictional rules of diversity to determine the propriety of hearing the claims of the additional plaintiffs.

In Finley itself, the Court noted this essential distinction between the operation of supplemental jurisdiction in federal-question and diversity proceedings. Although the Finley Court was willing to assume that the plaintiff's claims satisfied the constitutional "common nucleus" test for the assertion of pendent jurisdiction identified in Gibbs, it emphasized that the Gibbs test did not always control supplemental jurisdictional issues in different jurisdictional contexts. Pointing to Zahn as an example, the Finley Court noted that a transactional rela-

86 See id. (explaining that unrelated claims, none of which meet the amount-in-controversy threshold, may be aggregated as long as the total value of the claims satisfies the threshold amount).
88 See id. at 336-39 (stating that "Congress has... consistently amended the amount-in-controversy section and re-enacted the 'matter in controversy' language without change of its jurisdictional effect against a background of judicial interpretation" which did not permit aggregation of "separate and distinct claims").
89 See Zahn v. International Paper Co., 414 U.S. 291, 294-98 (1973) (refusing to permit exercise of jurisdiction over the claims of unnamed class members, joined under Rule 23, that did not meet the jurisdictional threshold); Clark v. Paul Gray, Inc., 306 U.S. 583, 589 (1939) (refusing to permit exercise of jurisdiction over jurisdictionally insufficient claims of additional plaintiffs joined under Rule 20).
90 See Zahn, 414 U.S. at 299 (noting that the doctrine of aggregation rests not upon the transaction-based joinder rules of the Federal Rules of Civil Procedure but on "this Court's interpretation of the statutory phrase 'matter in controversy'"); Snyder, 394 U.S. at 336-37 (explaining that the doctrine of aggregation rests upon the Court's interpretation of the phrase "matter-in controversy" as precluding aggregation).
91 See WRIGHT, supra note 19, at 214 (describing Zahn as the "death blow" to the lower court trend toward the application of the ancillarity principle to the determination of aggregation questions).
tionship among claims did not alone suffice to establish the existence of jurisdiction over new parties. As the Finley Court noted in its account of Zahn, "we based this holding upon 'the statutes defining the jurisdiction of the District Court,' ... and did not so much as mention Gibbs."92

B. Rereading Section 1367 in Light of Pre-Finley Law

The Court's continued emphasis on the applicability of the rules of diversity in cases like Snyder, Clark, and Zahn may explain why section 1367(a) draws a sharp distinction between "original jurisdiction" and "supplemental jurisdiction." In federal-question cases such as Gibbs and its progeny, original jurisdiction attached to a federal-law claim in the plaintiff's well-pleaded complaint; joinder of additional nonfederal claims triggered the application of the rules of supplemental (pendent) jurisdiction. But in diverse-party litigation, section 1332 and its collection of complete diversity and aggregation rules controlled the plaintiff's ability to join additional claims and parties.93

---

93 Under the rules that govern the timing of jurisdictional determinations in diversity, the federal court bases its determination as to the citizenship of the parties on the facts in the original complaint and does so as of the date of the filing of the action. See Freeport-McMoRan, Inc. v. K.N. Energy, Inc., 498 U.S. 426, 428 (1991) (discussing "the well-established rule that diversity of citizenship is assessed at the time the action is filed"). But jurisdiction does not attach at the outset for all time. If the plaintiff subsequently proposes to amend the complaint to join a nondiverse party, the district court must either reject the amendment or dismiss (or remand) the action. See, e.g., Hensgens v. Deere & Co., 833 F.2d 1179, 1181-82 (5th Cir. 1987) (specifying test for scrutiny of post-removal motion by plaintiff to add nondiverse defendant). Similarly, the district court must dismiss the action if a nondiverse party proposed for joinder meets the test of indispensability. See Horn v. Lockhart, 84 U.S. 570, 579 (1873) (stating that where nondiverse parties are not indispensable, dismissing such parties is preferable to joining them and then dismissing the action for lack of subject matter jurisdiction). The district court, moreover, must align the parties in accordance with their real interest in the action and must dismiss the action if, as realigned, the parties do not satisfy the diversity requirement. See, e.g., Indianapolis v. Chase Nat'l Bank, 314 U.S. 63, 69, 74-75 (1941) (explaining that "[d]iversity jurisdiction cannot be conferred upon the federal courts by the parties' own determination of who are plaintiffs and who are defendants"). Finally, although the district and appellate courts can dismiss any non-indispensable parties whose presence would otherwise defeat diversity (jurisdictional "spoilers"), the remaining parties must satisfy the requirements of diversity. See Newman-Green, Inc. v. Alfonzo-Larrain, 490 U.S. 826, 837 (1989) (holding that "courts of appeal have the authority to dismiss a dispensable nondiverse party," but noting that, at the appellate level, such authority should be "exercised sparingly."). See generally WRIGHT, supra note 19, §§ 28-30, at 171-80 (explaining when parties' diversity is determined, which parties are considered in that determination, and a court's ability
District courts sitting in diversity in the pre-*Finley* era would have had no occasion to consider supplemental jurisdictional issues until after the plaintiff had filed a complaint with claims and parties aligned in ways that satisfied the settled rules of original jurisdiction in section 1332. The same statutory distinction between original and supplemental jurisdiction applies to federal-question litigation but carries less significance in that context. As we have seen, a plaintiff with a substantial federal-law claim under *Gibbs* was free to invoke the district court’s original jurisdiction over that claim and its supplemental jurisdiction over a related nonfederal claim in the same well-pleaded complaint. Original jurisdiction concepts still control and still operate distinctly from supplemental jurisdiction concepts, but the original jurisdiction inquiry remains quite discrete and focuses entirely upon the existence of a substantial federal-law claim. To be sure, the *Finley* Court refused to extend the concept of supplemental jurisdiction to encompass a pendent claim against a new party, finding a lack of statutory authority in the provision that conferred original jurisdiction on the district courts in claims against the United States. But Congress re-
sponded to that restrictive conception of supplemental jurisdiction by expressly declaring in section 1367(a) that "[s]uch supplemental jurisdiction shall include claims that involve the joinder or intervention of additional parties." Congress cured the Finley problem, in short, by redefining supplemental jurisdiction to include additional parties and left the rules of original jurisdiction alone.

This sympathetic reading of section 1367(a) in light of prior law has a series of important implications for the interpretation of supplemental jurisdiction under the new statute. Perhaps most significantly, the sympathetic reading preserves the rules of complete diversity and aggregation that the Court had developed in the course of construing section 1332. As a gloss on the provisions of section 1332, these rules were understood to govern the plaintiff’s joinder of claims and parties in diversity litigation brought within the original jurisdiction of the district courts. Section 1367(a) appears to incorporate all of these rules of joinder and aggregation by referring to civil actions "of which the district courts have original jurisdiction." The language of the statute suggests a neat, if not entirely logical or conceptually consistent, distinction between the rules of original jurisdiction in diversity that were to govern the plaintiff’s joinder and aggregation of claims and parties and the rules of ancillary jurisdiction that were to control the defendant’s joinder of claims and parties in subsequent pleadings.

In preserving a broader array of rules to govern original jurisdiction over additional claims and parties in diversity than in federal-question proceedings, section 1367(a) leaves in place a distinction between pendent and ancillary jurisdiction that had grown up in prior cases. As we have seen, the pre-Finley decisions steadfastly refused to apply pendent jurisdiction concepts to diversity matters. Rather, the established body of law governing complete-diversity and aggregation continued to govern the plaintiff’s initial assertion of claims. The ref-

---

97 28 U.S.C. § 1367(a) (1994). Commentators agree that the statute accomplished the goal of overturning Finley by including this explicit provision for the exercise of pendent-party jurisdiction. See, e.g., Freer, Life After Finley, supra note 3, at 473 (approving of the implicit overruling of Finley).

98 On this account, then, the statute directly overrules Finley by including additional parties within the ambit of supplemental jurisdiction that the plaintiff may invoke, as in Gibbs, at the initial pleading stage.

99 See supra note 43 (setting forth the text of section 1367).

100 For a criticism of this absence of conceptual consistency, see ALI DRAFT, supra note 9, at xvi, 58 (describing current section 1367 as “imperfectly articulated in claim-specific terms”).
ference in section 1367(a) to the necessity of first securing "original jurisdiction" makes it clear that these joinder and aggregation rules would continue to control in diversity litigation. The provision for the exercise of supplemental jurisdiction over claims related to those "within such original jurisdiction" can only sensibly refer, at least in the diversity context, to the established doctrine of ancillary jurisdiction, which traditionally applied to the claims and parties joined in subsequent pleadings filed by defendants and intervening parties.101

This sympathetic reading of section 1367(a) fits well with the language and structure of subsequent provisions of section 1367.102 Consider first the provisions of section 1367(b), which specify a variety of situations in which the district courts, sitting in diversity, may not exercise the supplemental jurisdiction that has been conferred upon them in subsection (a). The statute bars such jurisdiction "over claims by plaintiffs against persons made parties under Rule 14, 19, 20, or 24" of the Federal Rules and over "claims by persons proposed to be joined as plaintiffs under Rule 19" when exercising such jurisdiction would be inconsistent with the jurisdictional requirements of section 1332.103 The statute reflects a concern with the preservation of the rules of diversity and a desire to preclude the grants of supplemental (ancillary) jurisdiction from eroding those rules.

More than merely confirming a general spirit of cautious restatement, subsection (b) offers strong structural support for a sympathetic interpretation of the scope of subsection (a)'s grant of supplemental jurisdiction in diversity. The exceptions specified in subsection (b) operate as a bar to the assertion of jurisdiction over claims that plain-

---

101 See McLaughlin, supra note 1, at 874 (noting that ancillary jurisdiction applies to claims "asserted in an ongoing federal lawsuit after the filing of the original complaint"); Friedenthal et al., supra note 19, §§ 2.12, 2.14 at 65-67, 76-81 (same).

102 It bears noting that the distinctive operation of supplemental jurisdiction in federal-question and diversity matters parallels a well-known distinction in these two familiar sources of judicial power. As Chief Justice John Marshall noted, Article III of the Constitution distinguishes between federal-question "cases" (jurisdiction "depends on the character of the cause," and not on the identity of the parties) and party-based "controversies" ("jurisdiction depends entirely on the character of the parties"). Cohens v. Virginia, 19 U.S. (6 Wheat.) 264, 378 (1816). See generally James E. Pfander, Rethinking the Supreme Court's Original Jurisdiction in State-Party Cases, 82 CAL. L. REV. 555, 604-17 (1994) (discussing the distinction and its implications for the interpretation of Article III). Even though this constitutional distinction has no immediate relevance to the proper interpretation of the statute, it may help to explain why the doctrine of pendent parties met resistance in the diversity context. Cf. Freer, supra note 2, at 62-63 (noting the difference between the subject-matter focus of federal-question jurisdiction and the party-based focus of diversity jurisdiction).

103 See supra note 43 (setting forth section 1367).
tiffs would propose to bring, not in the first instance, but in response to other claims that have been inserted into the proceeding by other parties. Subsection (b)'s exception for claims by plaintiffs against persons made parties under Rule 14 offers a definitive illustration. All commentators agree that the reference to Rule 14 codifies the Supreme Court rule in *Kroger*, a case in which the Court refused to permit the district court to assert ancillary jurisdiction over the plaintiff's claim against a nondiverse party joined by the defendant under Rule 14. The *Kroger* Court laid special emphasis on the fact that the ordinary rules of diversity jurisdiction would bar the plaintiff from asserting claims against the nondiverse defendant in the first instance. Unwilling to permit an end-run around this rule of complete diversity, the *Kroger* Court adopted a narrow interpretation of ancillary jurisdiction. Although the Court seemingly agreed to permit the assertion of ancillary jurisdiction over the impleader claims of defendants under Rule 14, and in other settings, the Court refused to countenance the expansion of ancillary jurisdiction to encompass a Rule 14 claim by the plaintiff against the newly impleaded defendant.

Notice how closely the underlying structure of section 1367 follows the rationale of *Kroger*. Subsection (a), sympathetically read, preserves the rules of complete diversity and confers supplemental (ancillary) jurisdiction over new claims and parties joined in pleadings subsequent to the plaintiff's initial complaint. But subsection (b), like *Kroger*, creates a restriction on the scope of such ancillary jurisdiction to preserve the essential features of the complete diversity requirement. In other words, the ancillary focus of the exceptions in subsection (b) tends to confirm that the grant of supplemental jurisdiction in subsection (a) operates in effect as a grant of ancillary jurisdiction. Because Congress had preserved the rules of complete diver-

---

104 See, e.g., McLaughlin, *supra* note 1, at 935-38 (describing the manner in which section 1367(b) codifies and extends the rule of *Kroger*).
105 See Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 377 (1978) (holding that the "policy of [section 1332] calls for its strict construction" and that "[t]o allow the requirement of complete diversity to be circumvented as it was in this case would simply flout the congressional command" (citations omitted)).
106 Id. at 373-74 (citing *Strawbridge* and *Zahn*).
107 Id. at 375 n.18 (citing with apparent approval lower court decisions that had asserted ancillary jurisdiction over compulsory counterclaims, cross-claims, impleader claims, and claims by intervenors as of right).
108 The remaining exceptions in section 1367(b) also focus on the containment of ancillary jurisdiction. In foreclosing jurisdiction over claims by plaintiffs against persons joined under Rules 19 and 24, subsection (b) seeks to preserve the rule that ancillary jurisdiction does not extend to necessary parties (Rule 19) and to produce the
sity and aggregation in subsection (a), it was simply unnecessary for Congress to establish exceptions in subsection (b) other than those necessary to rein in the scope of ancillary jurisdiction and to preserve the result in *Kroger*.

C. Testing the Sympathetic Reading

1. The Grant of Supplemental Jurisdiction

Notwithstanding the tight fit between the sympathetic reading and the language, structure, and legislative history of the statute, one can imagine plausible, if ultimately unconvincing, arguments against the interpretation. Consider first the argument that, by using the term “supplemental jurisdiction” in section 1367(a), Congress must have intended to bring the doctrines of pendent and ancillary jurisdiction into congruence rather than to preserve their differences under a new label. On this account, the plenary grant of supplemental jurisdiction in section 1367(a) operates identically in both federal-question and diversity cases and extends to the boundaries of a const-

same result with respect to those who intervene as defendants (Rule 24). *See generally* Rowe et al., *A Reply, supra* note 3, at 955-59 (discussing the statute’s resolution of the necessary-party/intervention anomaly). Original diversity jurisdiction did not attach to such claims under pre-*Finley* decisional law; rather, such claims were analyzed under the rules of ancillarity. *See, e.g.*, 7 Charles Alan Wright et al., *FEDERAL PRACTICE AND PROCEDURE* § 1610, at 150-54 (2d ed. 1986) (treating, in pre-*Finley* discussion, the Rule 19/24 anomaly as a question of ancillary jurisdiction). For an ancillary interpretation of the reference to Rule 20, *see infra* notes 113-19.

If the language and structure of the statute reveal a desire on the part of Congress to maintain the rules of diversity and to authorize ancillary jurisdiction only so far as that jurisdiction poses no threat to those rules, then the legislative history provides additional confirmation. The House Judiciary Committee Report explains that the statute responds to *Finley* by providing statutory authority to hear supplemental claims, including claims involving the joinder of additional parties. But the Report also suggests quite clearly that its codification of the rules of supplemental jurisdiction will preserve an existing distinction between federal-question and diversity litigation:

In federal-question cases, it broadly authorizes the district courts to exercise supplemental jurisdiction over additional claims, including claims involving the joinder of additional parties. In diversity cases, the district courts may exercise supplemental jurisdiction, except when doing so would be inconsistent with the jurisdictional requirements of the diversity statute.

*HOUSE REPORT, supra* note 77, at 28. The Report thus confirms that the statute leaves in place the rules governing the assertion of original jurisdiction in diversity proceedings, including the rules of joinder and aggregation that govern original jurisdiction of the plaintiff’s initial complaint. The same conclusion emerges from the Report’s further statement that the “section is not intended to affect the jurisdictional requirements of 28 U.S.C. § 1332 in diversity-only class actions, as those requirements were interpreted prior to *Finley*.” *Id.* at 29.
tutional "case" under Article III.\textsuperscript{110} Something like this argument, or this unspoken assumption, informs virtually every published account of section 1367 and deserves to be taken seriously.\textsuperscript{111} Indeed, many observers who have otherwise found much to criticize in the statute have applauded Congress's decision to end the pendent/ancillary distinction through adoption of the supplemental label.\textsuperscript{112}

I share the view that we should strive for doctrinal coherence in the application of supplemental jurisdiction concepts, but I do not believe that such a goal requires that we view the use of the term supplemental jurisdiction as reflecting a congressional decision to import pendent jurisdictional concepts into diversity litigation. On my account of the statute, supplemental jurisdiction can operate just as it did in the pre-Finley era. In federal-question litigation, the rules of original jurisdiction require only a substantial federal-law claim for supplemental (pendent) jurisdiction to support the adjudication of the plaintiff's additional, nonfederal claims. In diversity litigation, by contrast, the rules of original jurisdiction govern joinder and aggregation issues in the plaintiff's complaint and supplemental (ancillary) jurisdiction applies to the subsequent joinder of claims and parties. In conceptual terms, supplemental jurisdiction operates in the same way in both settings, coming into play only after the demands of original jurisdiction have been satisfied and applying to claims that satisfy the transactional relationship test of Article III. On this account, the same test of transactional relationship might well apply in both the federal-question and diversity contexts, thus achieving a measure of doctrinal coherence. At the same time, section 1367(a)'s preservation of the rules of original jurisdiction would, at least in diversity proceedings, defer the application of this supplemental jurisdictional analysis until after the complaint passed muster. It may well prove useful to continue to talk of pendent and ancillary jurisdiction to keep these distinctions straight, as many scholars have done in writing about the operation of the statute.\textsuperscript{113} Still, the crucial statutory dis-

\textsuperscript{110} See supra note 43 (setting forth the text of section 1367).
\textsuperscript{111} See supra note 16 (citing articles which view subsection (a) as a broad grant of supplemental jurisdiction).
\textsuperscript{112} See, e.g., Freer, Life After Finley, supra note 3, at 473 (noting that the use of the term "supplemental jurisdiction" reflects a "clear trend" in case law and academic literature); McLaughlin, supra note 1, at 860 (referring to the abolition of the pendent and ancillary labels as "beneficial").
\textsuperscript{113} See, e.g., McLaughlin, supra note 1, at 925 (noting that the statute can be understood by referring to the former doctrines of pendent-claim, pendent-party and ancillary jurisdiction).
tinction lies in its incorporation of the rules of original jurisdiction, which govern a much broader array of joinder and aggregation issues in diversity than in federal-question litigation. Once jurisdiction attaches, the statutory grant of supplemental jurisdiction may operate in much the same way in a variety of settings. What we used to think of as ancillary jurisdiction, for example, would continue to be available in both federal-question and diversity litigation.¹¹⁴

One can, in short, read Congress's decision to provide for the exercise of "supplemental jurisdiction" as something other than a directive to achieve precisely the same outcomes in both federal-question and diversity litigation. Congress did not express any desire to change the manner in which the doctrine operates in discrete cases (aside from its provision for the exercise of supplemental jurisdiction over additional parties). After all, the statute did not coin the term "supplemental jurisdiction"; scholars had previously used the term to refer to both pendent and ancillary jurisdiction.¹¹⁵ The statute's use of the phrase, therefore, conveys no desire to change the established operation of the underlying doctrine. Indeed, the Finley decision itself invoked a canon of construction under which a codification of existing jurisdictional practices would presumptively carry forward past interpretations unless that statute contains some clear expression of congressional intent to depart from the settled rule.¹¹⁶ One can scarcely discover an intent to change the law in Congress's decision to use the supplemental label standing alone, particularly in view of its drafters' avowed desire to avoid controversy. If one were inclined to consider it, moreover, the legislative history expressly disclaims any intention to make broad changes in the law.¹¹⁷

¹¹⁴ See id. at 874 (noting that, unlike pendent jurisdiction, ancillary jurisdiction applied whether the original jurisdiction claim was founded on a federal question, diversity of citizenship, or some other basis).

¹¹⁵ See supra note 2 (citing earlier use of the term "supplemental jurisdiction").

¹¹⁶ See supra note 42 (discussing Finley).

¹¹⁷ The clearest evidence in the legislative history appears in the House Judiciary Report, which includes the following statement:

The doctrines of pendent and ancillary jurisdiction, in this section jointly labeled supplemental jurisdiction, refer to the authority of the federal courts to adjudicate . . . [nonfederal] claims [that meet the test of Gibbs]. . . .

. . . . This section would authorize jurisdiction in a case like Finley, as well as essentially restore the pre-Finley understandings of the authorization for and limits on other forms of supplemental jurisdiction.

HOUSE REPORT, supra note 77, at 27-28. These statements, in a Report that emphasizes a desire on Congress's part to restore understandings of the "limits" on other, pre-Finley forms of supplemental jurisdiction, provide the backdrop for a sympathetic read-
The congressional disclaimer corresponds to the conclusion that emerges from a consideration of the correspondence among the drafters of the supplemental jurisdiction statute. In the Explanation that accompanied a discarded version of the statute, the authors expressly stated that the "purpose of the proposal is to codify the judicially created doctrines of pendent and ancillary jurisdiction as the Federal Courts Study Committee recommended." In reviewing that draft, one of the three principal drafters of the final version of section 1367 expressed "complete support" for the proposal to "codify the doctrines of pendent and ancillary jurisdiction." Although the terms of the codification changed over the course of the next few weeks, the correspondence of the drafters reveals no departure from this fundamental desire to codify pendent and ancillary jurisdiction. Indeed, the apparent focus of the drafters was to rewrite the draft statute to preserve the rule in *Kroger* and to bring the rules governing the assertion of diversity jurisdiction over parties joined under Rules 19 and 24 into congruence with *Kroger's* spirit.

118 Hearing on H.R. 5381 Before the Subcomm. on Courts, Intellectual Property, and the Administration of the Justice of the House Comm. on the Judiciary, 101st Cong. 689 (1990) [hereinafter Hearing] (Explanation of the Proposal to Codify Supplemental Jurisdiction accompanying Letter from Arthur D. Wolf, Professor, Western New England College of Law, to Robert W. Kastenmeier, Chairman, Subcommittee on Courts, Intellectual Property, and the Administration of the Justice (June 8, 1990)). Although the Wolf draft appeared in H.R. 5381 as late as September 6, 1990, Congress later chose to adopt a substitute measure that was largely the work of Professors Mengler, Rowe, and Burbank. See Fairman, supra note 4, at 160-70 (recounting the progression of drafts from that of Wolf, to the substitute of Judge Weis, and finally to the September 11, 1990 draft that became law).

119 Letter from Thomas M. Mengler, Professor, University of Illinois at Urbana-Champaign College of Law, to Robert W. Kastenmeier, Chairman, Subcommittee on Courts, Intellectual Property, and the Administration of the Justice (June 13, 1990), in Hearing, supra note 118, at 701.

120 One can see this focus on the problems associated with diversity jurisdiction and the preservation of *Kroger* in the correspondence of Professor Burbank, Professor Mengler, and Professor Kramer, a consultant to the Federal Courts Study Committee. See Letter from Stephen B. Burbank, Professor, University of Pennsylvania Law School, to Thomas M. Mengler, Professor, University of Illinois at Urbana-Champaign College of Law (Aug. 14, 1990), in Hearing, supra note 118, at 706-07 (raising the question of whether *Kroger* would be overruled by the proposed statute); Letter from Thomas M. Mengler to Stephen B. Burbank (August 24, 1990), in Hearing, supra note 118, at 708-09 (affirming that the proposed language would overrule *Kroger*); Letter from Larry Kramer, Professor, University of Michigan Law School, to Joseph F. Weis Jr., Judge, United States Court of Appeals for the Third Circuit (Aug. 51, 1990), in Hearing, supra note 118, at 713, 714-15 (arguing that an early draft of section 1367(b) that had appeared in the working papers of the Federal Courts Study Committee and that ultimately became the vehicle the drafters relied upon in crafting the statute avoided the problem of overruling *Kroger* by "preserving pre-Finley limitations on pendent jurisdiction."
One finds an echo, perhaps unconscious, of the drafters' desire to preserve the distinction between pendent and ancillary jurisdiction in the terms of the statute itself. The statute refers to the existence of original jurisdiction over the "claims" in a "civil action" and thus appears to reject the notion that a single, jurisdictionally sufficient claim will support the exercise of plenary pendent jurisdiction in diversity matters. Previous decisions as to the scope of diversity jurisdiction scrutinized the claims and parties in the action as a whole for compliance with the requirements of section 1332; as noted above, a single, jurisdictionally sufficient claim would not have supported the joinder of nondiverse parties or jurisdictionally insufficient claims. In its focus on the need for jurisdiction over the claims in a civil action, section 1367 decisively differs from the conceptual apparatus of the ALI Draft. The ALI Draft aims to transform the operation of supplemental jurisdiction by treating any jurisdictionally sufficient claim (including one in diversity) as a "freestanding" claim that will, by definition, support the assertion of jurisdiction over supplemental claims that satisfy the case-or-controversy test of relatedness. There may be good reasons to shift from an action-specific focus to a claim-specific focus, as the ALI Draft suggests, but there is no reason to believe that Congress had anticipated that change in the law when it enacted section 1367. Rather, the statute appears to preserve the action-specific focus of the rules of complete diversity and aggregation that had evolved up to that point.

2. Jurisdiction over Civil Actions and the Overruling of Finley

To the extent that section 1367(a) operates in diversity only after the district court first secures original jurisdiction over the civil action, one might doubt its effectiveness in overruling Finley in federal-question cases. Finley held that the jurisdiction of the district courts under the FTCA extends to claims against the United States but does not reach related state-law claims against nondiverse pendent parties. Section 1367(a) set out to overrule Finley by conferring supplemental jurisdiction on the district courts and by making it clear that such jurisdiction includes claims against new parties. But as we have seen, the statute provides for the assertion of supplemental jurisdiction only after the district court has acquired original jurisdiction over the civil action. So while a district court was free (before section

\[^{121}\] I am indebted to David Shapiro for drawing this possibility to my attention.

\[^{122}\] See supra notes 36-42 and accompanying text (describing Finley's holding).
1367's enactment) to exercise original jurisdiction over a federalquestion claim against a single defendant, Finley itself concluded that such original jurisdiction did not extend to related state-law claims against additional, nondiverse parties. Lacking original jurisdiction over all of the plaintiff's claims in the civil action, the district court in a case like Finley may appear to lack the statutory predicate for bringing section 1367(a)'s grant of supplemental jurisdiction into play. In other words, as one treatise noted in analyzing the problem addressed here, "if the courts applied the language of § 1367(a) literally, it would defeat the main purpose of the statute—overruling Finley."

In contrast, I believe that the literal terms of section 1367(a) can both preserve the rules of original jurisdiction in diversity and secure the overruling of Finley. Although Finley ruled out pendent-party jurisdiction in federal-question cases, it did leave pendent-claim jurisdiction and the rule of Gibbs intact. The preservation of Gibbs, however, coexists uneasily with the Finley Court's emphasis on the need for written statutory authority; the absence of written authority that Finley read to foreclose jurisdiction over claims against pendent parties appeared to many to plague the assertion of pendent-claim jurisdiction as well. The drafters of section 1367(a) thus set out to provide statutory authority for both Gibbs and Finley and did so in that portion of the statute that confers supplemental jurisdiction over "all other claims that are so related to claims in the action within such original jurisdiction that they form part of the same case or controversy under Article III." This language, clearly meant to codify Gibbs, assumes that the district court will hear pendent claims in a federal-question proceeding not as part of the court's original jurisdiction over a civil action but as part of the statutorily conferred grant of supplemental jurisdiction. The statute thus resolves the anomalous status of Gibbs.

---

125 See Finley, 490 U.S. at 549-51, 556 (distinguishing pendent-claim from pendent-party jurisdiction and expressing no desire to impair the continuing vitality of Gibbs, despite the fact that it departed from past practice in asserting jurisdiction without statutory authority).
126 See supra note 42 and accompanying text (describing concerns over the breadth of Finley's holding).
127 Most observers agree that the grant of supplemental jurisdiction over related claims in section 1367(a) codifies the pendent-claim jurisdiction rule in Gibbs. See, e.g., McLaughlin, supra note 1, at 925. Yet the logic of the action-specific focus that Professors Teply and Whitten rely upon in doubting the statute's effectiveness in overruling Finley would also render the statute essentially irrelevant to pendent-claim jurisdiction. See TEPLY & WHITTEN, supra note 124, at 123-24. The Teply-Whitten approach reads
after *Finley* by defining pendent-claim jurisdiction as part of the district court's supplemental, not original, jurisdiction.\(^2\)

The statute's treatment of pendent claims as lying within the district court's supplemental jurisdiction helps to make clear that the statute authorizes supplemental jurisdiction over pendent parties as well. Immediately following the grant of supplemental jurisdiction, the statute provides that "[s]uch supplemental jurisdiction shall include claims that involve the joinder or intervention of additional parties."\(^2\) In effect, then, the statute provides for pendent-party jurisdiction by providing a statutory foundation for *Gibbs*, and by extending such jurisdiction to include the addition of new, pendent parties. It thus assumes that a single "freestanding" or jurisdictionally sufficient federal-question claim will bring into play the district court's supplemental jurisdiction over related claims; in other words, the statute continues and codifies the claim-specific approach to pendent jurisdiction in federal-question cases that the *Gibbs* Court had developed. Such an approach belies the argument that district courts must defer their inquiry into the existence of supplemental jurisdiction until after the rigorous original jurisdiction demands of *Finley* have first been met.

3. The Rule 20 Wrinkle

One might also argue that the sympathetic reading of the statute cannot well account for the appearance in subsection (b) of a provision that restricts supplemental jurisdiction over claims by plaintiffs

---

\(^{123}\) In describing *Gibbs* as anomalous after *Finley*, I simply mean to note that the rationale of *Finley*'s insistence upon written statutory authorization raised doubts as to whether existing jurisdictional statutes conferred pendent-claim jurisdiction on the district courts. On the one hand, *Finley* expressly refused to treat the statutory grant of jurisdiction over a "civil action" as having acquired a judge-made gloss that included supplemental jurisdiction. See supra note 42. On the other hand, the *Finley* Court expressed no desire to impair *Gibbs*. See supra note 125. *Finley* thus left it uncertain as to whether a district court's pendent-claim jurisdiction under *Gibbs* would come into play as part of its original jurisdiction or as part of supplemental jurisdiction.

against persons made parties under Rule 20.\textsuperscript{130} The Rule 20 exception plays a major role in the standard account of section 1367, operating as an important restriction on the otherwise broad grant of supplemental jurisdiction in diversity matters conferred in subsection (a).\textsuperscript{131} In contrast to the standard account, the sympathetic reading posits that section 1367(a) incorporates the complete-diversity rules that preclude plaintiffs from joining additional, nondiverse defendants under Rule 20. If subsection (a) already precludes the plaintiffs from joining nondiverse defendants, one might plausibly ask why subsection (b) also includes language to foreclose the assertion of claims by plaintiffs against persons made parties under Rule 20. The standard account of the reference to Rule 20 in section 1367(b), in short, appears to undermine the sympathetic reading of section 1367(a).

But one can develop a sympathetic alternative to the standard account of the Rule 20 reference that fits well with the interpretation of section 1367(a) proposed in this Article. Recall that section 1367(b) seeks to protect complete diversity from situations, like that in \textit{Kroger}, in which plaintiffs seek to rely upon the district court’s ancillary jurisdiction over claims they assert against nondiverse defendants that others have joined to the litigation. In its effort to prevent an erosion of complete diversity through such ancillary jurisdiction, section 1367(b) deliberately uses the passive voice. Subsection (b) does not directly bar plaintiffs from joining nondiverse defendants under Rule 20; rather, it declares that plaintiffs may not assert claims against “persons made parties” under Rule 20. The statute thus can be read to contemplate that the persons in question will have been made parties by someone other than the plaintiff; the language of section 1367(b) literally addresses itself to the plaintiff’s subsequent assertion of claims.

\textsuperscript{130} Rule 20 provides in relevant part that all parties may join as plaintiffs in an action “if they assert any right to relief . . . arising out of the same transaction, occurrence, or series of transactions or occurrences and if any question of law or fact common to all these persons will arise in the action”; the same transactional test governs the joinder of multiple defendants. \textit{Fed. R. Civ. P. 20(a).}

\textsuperscript{131} See, e.g., Stromberg Metal Works, Inc. v. Press Mech., Inc., 77 F.3d 928, 931-32 (7th Cir. 1996) (noting the Rule 20 limits on the otherwise broad scope of section 1367(a)); ALI DRAFT, supra note 9, at 68 (describing the effect of the reference to Rule 20 as withdrawing supplemental jurisdiction over claims against additional defendants joined under that rule, but as failing to withdraw such jurisdiction over claims by additional plaintiffs joined under the same rule and ascribing the difference in treatment to an error in drafting); Rowe et al., \textit{A Reply}, supra note 3, at 961 n.91 (relying on the reference to Rule 20 in arguing that the statute preserved so much of the complete-diversity requirement as relates to the joinder of nondiverse defendants).
against these new parties.\footnote{One can probably best explain the reference to Rule 20 as a drafting error; indeed, most accounts of the statute adopt that view. See, e.g., Rowe et al., \textit{A Reply}, supra note 3, at 961 n.91 (describing the omission of claims by plaintiffs, joined under Rule 20, as a "far more serious" problem than others the critics had identified); Steinman, \textit{supra} note 16, at 100-01 (raising doubts about the nature of the error and maintaining that Rules 15 and 21, rather than Rule 20, are the proper rules for adding parties post-filing). It appears that the error resulted from an excess of caution on the part of the drafters. Although they had framed a statute that sought to preserve the complete-diversity requirement, last-second concerns led them to add a reference to Rule 20 in the limiting provisions in section 1367(b). \textit{Compare} Letter from Thomas M. Mengler, Professor, University of Illinois at Urbana-Champaign College of Law, to Thomas D. Rowe, Jr., Professor, Duke University School of Law (Aug. 28 1990), \textit{in Hearing, supra} note 118, at 716-17 (describing his changes to a draft of 1367(b) and listing exceptions for claims by plaintiffs against parties joined under Rules 14, 19, and 24, but omitting any reference to Rule 20), \textit{with} Draft, prepared by Thomas D. Rowe, Jr., Stephen B. Burbank & Thomas M. Mengler, Sept. 11, 1990, \textit{in Hearing, supra} note 118, at 722 (including in section 1367(b) an exception for claims against persons made parties under Rule 20 and suggesting that the addition of Rule 20 occurred in early September). \textit{See also} Fairman, \textit{supra} note 4, at 166-69 (tracing the progression of drafts and noting in particular that the reference to Rule 20 first appeared in the September 11 draft of Rowe, Burbank, and Mengler). Similar last-second concerns prompted the drafters to worry about their failure to include a restriction for claims joined under Rule 23. They caught the Rule 23 implications too late, however, to address with a change to the statutory language and so relied upon a curative reference in the legislative history instead. \textit{See} Rowe et al., \textit{A Reply}, supra note 3, at 960 n.90 (describing the attempt to address the Rule 23 problem through a curative reference in the legislative history). These concerns reflect an acceptance of what I have called the standard account, resting as they do on the view that section 1367(a) conferred a broad grant of supplemental (pendent-party) jurisdiction in diversity matters. The curative efforts reveal that the drafters did not read section 1367(a) itself as an adequate defense of the complete-diversity requirement.}

The Federal Rules furnish examples of situations in which defendants may join additional parties under the transactional test of Rule 20: Rule 13(a) specifies that a defendant must assert an available, transactionally related "compulsory" counterclaim against an opposing party and Rule 13(g) authorizes a cross-claim against a co-party.\footnote{\textit{FED. R. CIV. P.} 13(a), 13(g).} (Rule 14 incorporates similar joinder rules in permitting impleaded third-party defendants to set up counterclaims and cross-claims under Rule 13.)\footnote{\textit{FED. R. CIV. P.} 14(a).} Once a new claim has been asserted under these Rules, Rule 13(h) expressly permits a defendant to join additional defendants as parties to these counterclaims and cross-claims,\footnote{\textit{FED. R. CIV. P.} 13(h) (specifying that persons "other than those made parties to the original action may be made parties to a counterclaim or cross-claim in accordance with the provisions of Rules 19 and 20"). \textit{Cf. TEPLY \\& WHITTEN, supra} note 124, at 132-33 n.295 (suggesting a similar interpretation of the Rule 20 reference as applicable to ancillary jurisdiction over parties joined under Rule 13(h)).} so long as
the claims against such new parties satisfy the transactional test of Rule 20.136 Section 1367(a) extends the district court’s ancillary jurisdiction to the claims that defendants assert against such newly joined parties. But the Rule 20 exception in section 1367(b) can be read to follow Kroger in refraining from permitting the plaintiffs to assert claims against them in circumstances that might erode the complete diversity requirement.137 So read, the Rule 20 exception conforms to the general thrust of section 1367(b), which, sympathetically read, operates not as a constraint on what the plaintiff does in the initial complaint but on what the plaintiff does later with respect to subsequently joined parties.

Although a sympathetic reading of the Rule 20 exception does not conform to the drafters’ understanding of section 1367(b)’s operation, it may make as much sense as the standard account. For one thing, the standard account of the Rule 20 exception contains its own shortcomings. Most importantly, the exception fails to preclude the plaintiff from filing a diversity action and later moving to join additional, nondiverse plaintiffs under Rule 20 as parties within the grant of supplemental jurisdiction in section 1367(a). The sympathetic account of the statute avoids what the drafters rightly termed this “gapping hole” in the complete-diversity rule.138 Moreover, the sympathetic account of the Rule 20 exception makes better sense of the subsequent language in section 1367(b). Recall that section 1367(b) does not operate as a flat bar to the assertion of supplemental jurisdiction.

136 FED. R. CIV. P. 13(h). To be sure, some observers have taken issue with functional interpretation of the reference in section 1367 to Rule 20, arguing that its rule of transactional relationship for the joinder of parties does not actually provide a measure of the propriety of such joinder in the context of pending litigation. See Steinman, supra note 16, at 100-01 (arguing that Rule 20 joinder does not apply to the addition of parties post-filing). These observers argue that the statute should have referred instead to Rule 15 as the rule that actually provides the standard for leave to amend the pleadings to add additional parties. Id. I do not share this criticism of the statute’s reference to Rule 20. All of the rules referred to in the statute set forth a test of transactional relationship. It is these tests of transactional relationship, and not general standards for amendment of the pleadings, that inform the questions of joinder and supplemental jurisdiction in the course of motions for leave to amend under Rule 15.

137 Of course, one might argue that the proposed interpretation of the Rule 20 reference suffers from redundancy insofar as it incorporates the joinder rules in Rule 14, which appears as a separate exception in section 1367(b). Yet one can see the rules as performing discrete functions: the reference to Rule 14 incorporates the test for the impleader of third-party defendants (as in Kroger) and Rule 20 governs the assertion of claims against parties added to the litigation in connection with counterclaims and cross-claims allowed under Rule 13.

138 See supra note 65 (quoting the statute’s drafters).
SUPPLEMENTAL JURISDICTION

One has difficulty seeing how a district court could possibly find that an action in which plaintiffs have invoked Rule 20 to join nondiverse defendants would be consistent with the complete diversity rules of Strawbridge v. Curtiss. But the assertion of claims by plaintiffs against persons made parties under Rule 20 by someone other than the plaintiff might have a less dramatic tendency to erode the complete diversity rule, at least so long as the proposed joinder did not present the threat of strategic manipulation that persuaded the Court to enforce the complete diversity rule in Kroger.

D. Applying the Sympathetic Reading

Sympathetically read, section 1367 produces results that closely conform to the House Report's assertion that the statute was meant to preserve the jurisdictional requirements of the diversity statute. Recall that section 1367(a) provides for the assertion of supplemental jurisdiction only after "original jurisdiction" has attached to the claims in the plaintiff's complaint. This means that section 1367(a) does much of the work of preserving diversity by incorporating the rules of complete diversity into that term. While section 1367(a) makes ancillary jurisdiction available after the rules of complete diversity have been satisfied and jurisdiction attaches, section 1367(b) creates exceptions to the scope of such ancillary jurisdiction to provide further protection for the rule of complete diversity.\footnote{28 U.S.C. § 1367(b) (1994).}

\footnote{Professors Tidmarsh and Transgrud have suggested an "alternate" reading of section 1367(b) that would make its restrictive provisions inapplicable to cases such as Abbott Laboratories and Stromberg. See JAY TIDMARSH & ROGER H. TRANSGRUD, COMPLEX LITIGATION AND THE ADVERSARY SYSTEM 355-56 (1998). In constructing their alternative, Professors Tidmarsh and Transgrud emphasize the language that limits the application of section 1367(b) to matters founded "solely on section 1332." Id. at 355. They note that Abbott Laboratories and Stromberg fail to satisfy the demands of complete diversity (as understood in cases decided before the statute became law) and could on this account fall outside section 1367(b) power such that the provisions in section 1367(b) do not come into play. Id. They also note that Snyder may present a different question from that in Zahn, inasmuch as none of the claims in Snyder came within the district court's original diversity jurisdiction. Id. at 356.}

While this alternate reading focuses on the interplay between sections 1367(a) and (b) and offers an interesting textual alternative to Abbott Laboratories and Stromberg, it differs in important respects from (and thus presents problems that do not plague) the sympathetic account. For example, the authors of the alternate reading struggle with what they see as a problem of tautology; if section 1367(a) permits less than complete
1. Preserving Zahn and Clark

The sympathetic reading of section 1367 produces results wholly consistent with the legislative history and remains true to the language of the statute that Congress adopted. It thus dissolves the apparent tension, reported in cases such as Abbott Laboratories and Stromberg, between the supposedly clear results demanded by the literal text and the more modest claims in the legislative history. The Abbott Laboratories court mistakenly assumed that section 1367(a) conferred a plenary grant of pendent jurisdiction on the federal courts, sitting in diversity. With such a grant in place, the court looked for exceptions in section 1367(b). Finding no exception there for claims brought by additional parties joined as plaintiffs under Rule 23, the Abbott Laboratories court concluded that Congress had unthinkingly overruled Zahn. The Stromberg court ascribed the same interpretive significance to the omission of the joinder of plaintiffs under Rule 20 in the course of its decision to regard section 1367 as a legislative overruling of the Clark rule.

Both courts erred in looking for an explicit exception on the face of section 1367(b). The exceptions appear instead in the requirement that "original jurisdiction" attach to the complaint of the plaintiffs in section 1367(a). The established rules of Zahn and Clark form a part of the jurisdictional requirements in diversity proceedings, and they were both incorporated into section 1367(a) by reference. Congress thus had no reason to create any explicit exception in 1367(b) for claims by plaintiffs joined under either Rule 23 or Rule 20. Jurisdiction over claims entailed in the joinder of additional plaintiffs had always required that those plaintiffs meet the amount-in-controversy requirement; the Zahn Court had explicitly refused to substitute the
doctrine of pendent jurisdiction as the measure of such joinder and section 1367 seemingly preserves that choice in language preserving the rules of original jurisdiction.

Recent decisions, which explicitly reject the approach of Abbott Laboratories and Stromberg, provide some basis for optimism. In Leonardt v. Western Sugar Co., the Tenth Circuit rejected the textualism of the two prior circuit decisions and concluded that Zahn remains good law. Citing an early draft of this Article, the court found that the text of section 1367 can be read "literally, and unambiguously" to incorporate the established rules governing the amount in controversy requirement for diversity jurisdiction. Some months later, in Meritcare Inc. v. St. Paul Mercury Insurance Co., the Third Circuit followed the Tenth Circuit's lead in rejecting the argument that section 1367 overrules Zahn. Although the Third Circuit found the Tenth's account of the text attractive, it ultimately based its decision on the legislative history (as had the Tenth) following the established rule that courts may defer to the legislative history in the face of ambiguities in the relevant text. Writing for the court, Judge Joseph Weis—the judge who had chaired the Federal Courts Study Committee and had represented the Judicial Conference in testimony about the statute in 1990—made a compelling case that the history pointed to the preservation of the complete diversity rules.

2. Preserving St. Paul

The sympathetic reading also helps to expose the flaws in the Fourth Circuit's decision in Shanaghan v. Cahill, which read section 1367 as replacing the bright-line rule of St. Paul with the discretionary regime of section 1367(c). Shanaghan proceeded by assuming that

141 160 F.3d 631, 640 (10th Cir. 1998) (concluding that "a literally and textually faithful reading of section 1367(a) leads to the opposite conclusion from that of [Abbott Laboratories and Stromberg]").

142 Id. at 639 n.6, 640.

143 166 F.3d 214, 222 (3d Cir. 1999) (holding that section 1367 "preserves the prohibition against aggregation . . . and thus maintains the traditional rules governing diversity").

144 See id. at 222 (finding "sufficient ambiguity in [section 1367] to make resort to the legislative history appropriate").

145 58 F.3d 106, 112 (4th Cir. 1995) (holding that under section 1367 courts should "weigh convenience and fairness to both parties, as well as the interests of judicial economy" in deciding whether to retain jurisdiction over claims that fall below the amount-in-controversy requirements post-filing).
the assertion of jurisdiction over claims aggregated to meet the jurisdictional threshold represented an assertion of supplemental jurisdiction under section 1367(a). In truth, however, the federal courts have long applied their aggregation rules, not as a matter of supplemental jurisdiction, but as a matter of original jurisdictional analysis. Since the district court's power to hear the additional claims in Shanaghan did not, strictly speaking, derive from the grant of supplemental jurisdiction in section 1367(a), the Fourth Circuit had no basis for bringing the discretionary regime of section 1367(c) into play. Subsection (c), after all, permits the district court to refrain from exercising "supplemental," not original, jurisdiction. One might defend the outcome in Shanaghan as part of a considered review of the messy aggregation rules that govern original jurisdiction, but such a review would have to confront the controlling language of St. Paul more directly.

3. Joinder in Alienage Cases

The sympathetic reading also lays to rest the concern that section 1367 might have unwittingly foreclosed the assertion of pendent-party jurisdiction in alienage cases. The sympathetic reading teaches that

---

146 See supra text accompanying notes 80-92 (discussing the distinction between aggregation rules and supplemental jurisdiction).

147 For a well-reasoned rejection of Shanaghan that emphasizes the distinction between rules of original and supplemental jurisdiction in diversity matters, see Wolde-Meskel v. Vocational Instruction Project Community Services, Inc., 166 F.3d 59, 64-65 (2d Cir. 1999). But see Stevenson v. Severs, 158 F.3d 1332, 1334 (D.C. Cir. 1998) (following Shanaghan).

148 It was Professor Freer who first argued that section 1367(b) flatly bars the use of pendent-party jurisdiction in alienage matters. See Freer, Life After Finley, supra note 3, at 474-75 ("Although no one objects to pendent parties jurisdiction in [alienage cases], the statute outlaws its use, probably through inadvertence."). Professor Freer based his argument on the standard account of the Rule 20-reference in section 1367(b), which he took to apply to any attempt by plaintiffs to join additional defendants in matters within the scope of the jurisdiction conferred by section 1332. Seeing the reference to section 1332 as requiring complete diversity in alienage jurisdiction, as well as citizen-citizen diversity, Professor Freer argued that the drafters had inadvertently eviscerated pendent-party jurisdiction. The drafters responded by arguing that section 1367(b) did not require such an outcome, pointing to the language at the end of the provision that forecloses supplemental jurisdiction only where inconsistent with the jurisdictional requirements of section 1332. See Rowe et al., A Reply, supra note 3, at 954-55 (noting that after the passage of section 1367, federal courts remain free to "abolish the complete diversity rule for alienage jurisdiction"). According to the drafters, the statute simply refused to address the question of pendent parties in alienage, leaving the matter for resolution ultimately by the Supreme Court. See id. Compare Arthur & Freer, The Disaster, supra note 3, at 978-81 (expressing doubt that the statute's text can be construed to leave the matter to judicial discretion), and Arthur & Freer, Close
the rules governing original jurisdiction in section 1332 matters remain intact, having been incorporated by reference into section 1367(a). Just as it preserves the rules of complete diversity in suits brought under section 1332(a)(1), section 1367(a) can be read sympathetically to preserve the role of the federal courts in giving shape to the rules of diversity in alienage cases under section 1332(a)(2). The rules of diversity in alienage matters have arisen by virtue of the federal courts' interpretation of the demands of original jurisdiction; they do not depend on the grant of supplemental jurisdiction in section 1367(a) and do not come within the ambit of the restrictive language in section 1367(b). One can thus conclude that the statute preserves the status quo, just as its defenders have argued, without having to rely upon the language in section 1367(b) that relaxes its otherwise flat prohibition of the exercise of supplemental jurisdiction in specified situations under section 1332.

4. Supplemental Jurisdiction After Removal

Early comments on the supplemental jurisdiction statute expressed a variety of concerns about its application to removed actions. Some observers, for example, questioned whether the statute would apply to removed cases at all, noting that the final version of

Enough, supra note 3, at 1012 (suggesting that courts interpreting the statute will not reach uniform conclusions regarding its application), with Rowe et al., A Coda, supra note 3, at 998-99 (restating the claim of statutory agnosticism on the issue of alienage jurisdiction).

140 The question of jurisdiction in alienage cases continues to evolve. See Rowe, supra note 8, at 59-61 (summarizing the rules of complete diversity that have emerged in section 1332(a)(2) cases but arguing that the rules make no sense and inappropriately restrict access to federal courts for aliens who may suffer bias in state courts). Professor Rowe acknowledges that the Supreme Court might yet rewrite these rules but considers the prospect unlikely and on that basis urges statutory reform. See id. at 61. I have no quibble with the proposed reform and simply note that the current statute does not codify a complete-diversity rule for alienage matters but leaves the matter for resolution by the federal courts, and ultimately the Supreme Court, in the interpretation of the grant of original jurisdiction.

150 For an account of the drafters' argument for the preservation of the status quo, relying upon the language in section 1367(b), see supra note 148.

151 See Freer, Life After Finley, supra note 3, at 485 (questioning whether the statute applies to removed cases); McLaughlin, supra note 1, at 949-52 (pointing out inequities to plaintiffs in removed actions if the statute is applied); Karen Nelson Moore, The Supplemental Jurisdiction Statute: An Important but Controversial Supplement to Federal Jurisdiction, 41 EMORY L.J. 31, 58-60 (1992) (discussing the statute's possible effects on removal cases). But see Joan Steinman, Supplemental Jurisdiction in § 1441 Removed Cases: An Unsurveyed Frontier of Congress' Handiwork, 35 ARIZ. L. REV. 305, 308-10 (1993) (supporting the applicability of the statute).
the statute dropped the specific reference to removed actions that had appeared in an early draft. Others expressed doubts that section 1367(b) applied to actions removed on the basis of diversity; these doubts stemmed from the perception that removal jurisdiction depends upon the interaction of sections 1332 and 1441 and not "solely" on section 1332 as section 1367(b) specifies. Still others worried about the unfairness that might result from the application of Kroger (and other restrictive rules of complete diversity) to plaintiffs who had not chosen the federal forum. Finally, some observers argued that the restrictive references to specific federal rules in section 1367(b) might have no application to litigation governed by state rules of procedure. Coupled with the standard account of section 1367(a), the inapplicability of section 1367(b) limits to removed cases appeared to have threatened some broadening of federal removal jurisdiction.

Sympathetic interpretation can overcome these interpretive problems. The removal statute provides for removal by the defendants of any "civil action" of which the district courts have "original jurisdiction." This test of original jurisdiction focuses on the plaintiff's state court complaint and considers whether its allegations meet the standards for the assertion of federal-question or diversity jurisdiction. It thus appears plausible to conclude that many of the same rules that govern the interplay between original and supplemental jurisdiction in actions instituted in federal court will also apply to actions removed.

---

152 See, e.g., Moore, supra note 151, at 59 (noting that the specific reference to removal cases dropped out during the drafting process).
153 See Freer, Life After Finley, supra note 3, at 485 (arguing that removal jurisdiction is not founded solely on section 1332 and hence is not within the provisions of section 1367(b)); Moore, supra note 151, at 58 (same). But see Steinman, supra note 151, at 328 (considering and rejecting the argument for the inapplicability of 1367(b) to removed cases).
154 See Moore, supra note 151, at 58-59 (arguing that because the plaintiff in a removed case had initially chosen state court, the reasons for limiting supplemental jurisdiction emphasized in Kroger do not apply).
155 See Steinman, supra note 151, at 330 (arguing that section 1367 does not apply to claims against persons made parties under state law prior to removal).
156 See Steinman, supra note 8, at 103 (noting the possibility that section 1367 could be read to give federal courts jurisdiction over state-law class actions regardless of the citizenship and amounts in controversy of class members); Steinman, supra note 151, at 331 (concluding that section 1367(b)'s inapplicability to removed diversity cases could expand federal jurisdiction over class actions).
158 See generally WRIGHT, supra note 19, at 224 (describing the rules that govern the determination of original and removal jurisdiction as "equated" and "linked" though not entirely coincident).
to federal court. Whether the action begins in federal court or comes there on removal, the plaintiff's complaint must allege a civil action over which the district courts have original jurisdiction. By its terms, section 1367(a) applies to "any" such civil action and confers jurisdiction over related claims.

The parallel structure of the removal and supplemental jurisdiction statutes dissolves many of the uncertainties that others have identified. First, it appears quite clear, as the Supreme Court recently held, that the supplemental jurisdiction statute applies to removed actions. Second, the sympathetic reading of sections 1367(a) and (b) essentially eliminates any concern that supplemental jurisdiction will unduly expand the scope of diversity jurisdiction in removed cases. If, as sympathetically construed, section 1367(a) incorporates the demands of complete diversity into its provision for the assertion of original jurisdiction, then the same rules that govern plaintiff's joinder and aggregation of parties and claims will apply to diversity matters initiated in and removed to district court. Only after the complaint satisfies such rules can the defendants remove and bring the doctrine of federal supplemental (ancillary) jurisdiction into play. As a consequence, section 1367(b)'s arguable inapplicability to actions in state court does not present an inordinate threat of broadened supplemental jurisdiction in removed diversity proceedings. Section 1367(a), as we have seen, preserves the complete diversity rules and section 1367(b) attempts to prevent ancillary jurisdiction from eroding them. Facing time constraints, most defendants will have removed the action before moving to implead or otherwise add parties whose joinder would present post-removal questions of ancillary jurisdiction.
III. SYMPATHETIC TEXTUALISM AND JUDGE-MADE JURISDICTIONAL LAW

So far, this Article has suggested that a sympathetic textualism can produce answers to discrete interpretive issues that fit well with the text of the statute and with the apparent understanding of Congress. This Part of the Article steps back from the particular issues to argue that the sympathetic reading of section 1367 also provides a stronger foundation on which to build a workable body of jurisdictional law. By reading section 1367 as a general grant of pendent and ancillary jurisdiction and as a decision otherwise to leave many jurisdictional issues untouched, the sympathetic account narrows the range of issues that the federal courts must regard as ones that Congress has definitively resolved. So read, the statute leaves a much broader range of issues for resolution by the federal courts, inviting them, in effect, to exercise the kind of pragmatic discretion that had long characterized judge-made jurisdictional law in the years preceding Finley. By preserving a role for the federal courts in the development of jurisdictional law, the sympathetic reading may avoid the stubborn textualism of Abbott Laboratories and the elaborate detail of the ALI Draft.

Of course, one can argue that Congress should, as a matter of first principles, play the lead role in the development of jurisdictional law. Congress, after all, bears responsibility for the decision to create inferior federal courts and has broad, if not unfettered, control over the scope of their jurisdiction.162 Many thoughtful observers have emphasized the importance of legislative control of jurisdiction and have worried that the courts will take too many liberties with their jurisdictional grants. Professor Martin Redish, for example, speaks for many when he argues that the federal courts have no principled basis on which to abstain from the exercise of jurisdiction that Congress has conferred upon them.163 Similar arguments for legislative primacy underlie criticisms of other doctrines in which the courts exercise

162 For an account of the Madisonian Compromise and its embodiment in provisions of Article III that authorize but do not require Congress to create lower federal courts, see BATOR ET AL., supra note 78, at 11. See also Glidden Co. v. Zdanok, 370 U.S. 530, 551 (1962) (describing the "great constitutional compromise" that authorized but did not obligate Congress to create inferior courts and noting that, once created, such courts remained "subject to jurisdictional curtailment").

some discretion in defining what matters to hear.  

Among the most persuasive responses to this emphatic argument for legislative primacy, Professor David Shapiro has advanced a subtle and, to me, quite persuasive argument that the federal courts properly exercise a principled discretion in giving more particular content to general jurisdictional statutes. As Professor Shapiro notes, experience and tradition teach that the question whether to exercise jurisdiction and decide the merits of a particular dispute may defy general legislative definition; courts may enjoy functional advantages over the legislature in the necessary fine tuning. In suggesting the need for a dialogue between the courts and the legislature, Professor Shapiro argues for the widely held view that the business of defining the contours of judicial power represents a shared responsibility of the First and Third Departments.

Certainly as a matter of history, much jurisdictional law has grown out of an unspoken partnership between the legislative and judicial branches of government. Congress has tended to provide relatively general jurisdictional grants and the Supreme Court has often played a fairly active role in shaping what we now think of as jurisdictional law. In a range of familiar cases, the Court has adopted interpretations of jurisdictional statutes that redefine the scope of federal power along lines scarcely visible in the legislative text. Many observers de-

---

164 Compare ALEXANDER M. BICKEL, THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS 111-98 (1962) (defending the Court's use of justiciability doctrines to exercise discretion to avoid the merits of some controversial issues) with Gerald Gunther, The Subtle Vices of the "Passive Virtues"—A Comment on Principle and Expediency in Judicial Review, 64 COLUM. L. REV. 1, 25 (1964) (criticizing such discretionary avoidance as an unprincipled refusal to exercise jurisdiction that Congress has conferred on the Court).

165 See David L. Shapiro, Jurisdiction and Discretion, 60 N.Y.U. L. REV. 543, 574 (1985) (arguing that courts are uniquely qualified on jurisdictional matters and should continue to have "measured authority" to decline jurisdiction).

166 See id. at 577 (advocating a "productive dialogue").

167 Professor Friedman has seconded Shapiro's call for dialogue between the courts and the legislature. See Barry Friedman, Dialogue and Judicial Review, 91 Mich. L. Rev. 577, 580-81, 668-69 (1993) (arguing against rigid separation of powers thinking and in favor of a dialogic approach to defining the judicial role). The Supreme Court's decision in Quackenbush v. Allstate Insurance Co., 517 U.S. 706, 731 (1996) (holding that Burford abstention can only be applied when "equitable or otherwise discretionary" relief is sought), reaffirmed the existence of the district court's equitable discretion to refrain from adjudicating particular disputes.

168 See, e.g., Louisville & Nashville R.R. Co. v. Mottley, 211 U.S. 149, 152-54 (1908) (holding that only those federal questions that appear on the face of the well-pleaded complaint can support the district court's assertion of arising-under jurisdiction); Osborn v. Bank of the United States, 22 U.S. (9 Wheat.) 738, 823 (1824) (holding that
scribed the evolution of pendent and ancillary jurisdiction in the pre-
Finley years as simply another illustration of the way in which the Court
adapted jurisdictional law to take account of evolution in our under-
standing of the proper scope of a civil action.\(^{169}\)

Practical considerations help to explain the role of the federal
courts in making jurisdictional law. For a variety of reasons having to
do with the nature of the political process, Congress simply has not
done an effective job of keeping jurisdictional rules in good repair.\(^{170}\)
Partly this congressional neglect reflects the inability of federal judges
to play an institutionally effective role in securing jurisdictional legis-
lation; partly it reflects the absence of well-organized interest group
support; partly it reflects the relatively specialized nature of the sub-
ject matter and its inaccessibility to those without special competence
in the subjects of civil procedure and federal courts.\(^{171}\) For all these
reasons, Congress has been content, absent a crisis, to leave the elabo-
ration of jurisdictional rules to the courts and similar specialists.

Even those who agree that the courts play an appropriate, and not
just an inevitable, role in the development of jurisdictional rules may

---

Article III empowers Congress to assign federal trial jurisdiction over any claim in
which a federal ingredient appears); Strawbridge v. Curtiss, 7 U.S. (3 Cranch) 267, 267
(1806) (holding, despite the absence of any clear constitutional or statutory require-
ment to that effect, that minimal diversity would not support federal trial jurisdiction).

\(^{169}\) See Freer, supra note 2, at 55-60 (arguing that one can best rationalize the judge-
made doctrines of pendent and ancillary jurisdiction as an interpretation of the statu-
tory grant of original jurisdiction over "civil actions").

\(^{170}\) For a summary of Congress's work on the judicial code of the United States, see
Wright, supra note 19, at 21-25 (describing the codifications of 1911 and 1948; noting
the absence of any systematic updating of jurisdictional provisions; questioning the
skill of the initial drafting and the quality of subsequent amendments; and opining
that the time has come for the preparation of a new judicial code). In keeping with
these concerns, Professor Wright supports, in his capacity as president of the ALI, the
ALI's current judicial code revision project.

Although it has left jurisdictional rules in disrepair, Congress has taken an in-
creased interest in the rule-making process in recent years. See Charles Gardner Geyh,
Paradise Lost, Paradigm Found: Redefining the Judiciary's Imperiled Role in Congress, 71
N.Y.U. L. Rev. 1165, 1169, 1187-91 (1996) (describing the period from 1973 to the
present as one of heightened interaction between the judiciary and a Congress that is
more willing to suspend and modify proposed procedural rules and expressing some
optimism about the prospects for more effective interactions in the future).

\(^{171}\) For a general account of the difficulties that arise from interactions between
Congress and the judiciary in the course of the extrajudicial making of jurisdictional
and procedural law, see Geyh, supra note 170. Professor Geyh persuasively argues that,
despite some questions of self-interest, federal judges participate effectively in the
lawmaking process because they are "extraordinary lawyers" who may understand the
law and the implications of proposed reforms better than their legislative counterparts.
Id. at 1219 (noting judges' unique expertise in matters such as procedure and judicial
administration).
view Finley as signaling the beginning of a new era. With its emphasis on the necessity for written authority, the Finley Court made what some have seen as a decisive break with the past. On this account, Finley brought to a close the free-wheeling jurisdictional days of Gibbs and inaugurated an era of close attention to statutory text. If the Finley Court foreswore the exercise of what the ALI Draft perceptively describes as pragmatic discretion in fashioning jurisdictional rules, then decisions such as that in Abbott Laboratories arguably proceed with appropriate deference to the command of their judicial superior in hewing closely to the textualist line.

The argument makes sense as far as it goes but the Finley decision proceeds upon the assumption that the ultimate responsibility for the content of jurisdictional law rests with Congress, not the Supreme Court. In a world of avowed legislative supremacy, a decision by Congress to reestablish the partnership with the federal courts by delegating some responsibility for jurisdictional law to the federal courts would seemingly answer any doubts about the legitimacy of the judicial role. No constitutional principle forbids Congress from making such a delegation and it thus seems apparent that a post-Finley delegation of law-making authority from Congress to the federal courts would trump the new emphasis on literal textualism in Finley itself.

---

172 For the argument that Finley and section 1367 signal a new era of literal textualism in the law of supplemental jurisdiction, see Arthur & Freer, The Disaster, supra note 3, at 979 (arguing that the courts may not ignore the literal commands of a statute clear on its face); Freer, supra note 8, at 7 (arguing that the statutory model adopted in section 1367 "locks" future efforts into the same statutory model and precludes reliance on a regime of judicial discretion); and Purdue, supra note 16, at 74-75 (opposing development of the law through lower court decisions and arguing instead that the statute should provide clear and concrete answers, as in the model of Treasury Regulations). Cf. Rowe, supra note 8, at 54-58 (expressing general support for an interpretive approach that preserves a measure of judicial discretion and maintains a practical, working relationship between the federal courts and Congress but nonetheless concluding that certain features of the existing statute require a fix).

173 For the ALI Draft's view of the demise of pragmatic discretion in judicial interpretation of jurisdictional grants, see supra notes 69-74 and accompanying text (describing the ALI Draft's approach to the statute).

174 See Finley v. United States, 490 U.S. 545, 547-48 (1989) (arguing from implicit separation of powers grounds that the federal courts may not exercise jurisdiction without an appropriately clear grant of legislative authority).

175 Consider, for example, Justice Frankfurter's famed dissent in Textile Workers Union v. Lincoln Mills, 353 U.S. 448, 460 (1957) (Frankfurter, J., dissenting). Although Frankfurter argued against the constitutionality of a provision of the Taft-Hartley Act that enabled federal courts to adjudicate suits for violation of collective bargaining agreements, he did so based upon his contention that the statute merely conferred
Precisely such a delegation can be said to flow from the relatively open-ended grant of supplemental jurisdiction in section 1367. Read sympathetically, the statute supplies the grant of statutory authority that the *Finley* Court had identified as missing and otherwise attempts to reestablish the role of the federal courts in working out the details of supplemental jurisdiction. Three provisions in particular appear to restore a measure of judicial discretion. First, section 1367(a) incorporates the rules of original jurisdiction in diversity, rules that had previously developed through the exercise of some judicial discretion and remained in flux to some degree. Second, section 1367(b) establishes that the district courts may exercise ancillary jurisdiction in diversity matters only where they find it to be consistent with the jurisdictional principles of section 1332. This test of consistency incorporates the pragmatic analysis of the *Kroger* decision, where the Supreme Court balanced the threat to the complete diversity requirement against the arguments for litigant fairness in determining whether to hear the claim. Third, section 1367(c) incorporates the rules governing the discretionary decline of jurisdiction that the *Gibbs* Court had earlier specified. The statute thus provides a framework within which the federal courts may continue to work out sensible jurisdictional rules.

The question of alienage jurisdiction provides an excellent illustration of the statute’s operation as a jurisdictional framework. Ac-
According to most observers, the federal courts had yet to decide whether to allow pendent-party jurisdiction in alienage cases when section 1367 became law. It takes a great leap of interpretive faith to conclude that Congress addressed that question one way or another in section 1367. It makes far more sense to conclude that Congress simply left the matter in the hands of the courts through its provision in section 1367(a) for the assertion of ancillary jurisdiction only after the demands of original jurisdiction had been satisfied. The statute does not specify the contours of that original jurisdiction and thus leaves the process of determining the proper boundaries between original and ancillary jurisdiction to the federal courts.

Whatever one's view of the merits of that question, I believe that the preservation of a judicial role provides a better prescription for the long-term health of jurisdictional law than the rigorous textualism of Abbott Laboratories. Recall that in Abbott Laboratories, the Fifth Circuit not only refused to give effect to the legislative history but also quite deliberately refused to consider the "wisdom" of its conclusion that section 1367 had effectively, if unwittingly, overruled Zahn. In the end, the Abbott Laboratories court refused to take any responsibility for the jurisdictional rule that it applied; it simply declared the legislative text clear on its face and walked away. Perhaps the court meant to teach Congress a lesson, as Judge Pollak suggested in characterizing the opinion as an example of interpretive "gotcha." In any case, the

---

179 See supra note 149 (noting the continued evolution of alienage jurisdiction).

180 Some may wonder if the district courts would enjoy the same authority to re- think the rules of complete diversity as part of the process of elaborating the meaning of original jurisdiction in section 1367(a). In my view, the question turns not on the meaning of the statutory language but rather on the nature of the district court's obligation to respect the unamended decisional law of the Supreme Court. The Court itself has created the rules that govern diversity of citizenship and amount-in-controversy and those decisions bind lower courts. In the alienage context, by contrast, the Court has yet to speak and the process of lower court development may continue. Of course, the lower courts may have reached a consensus that will reduce the likelihood of any further review by the Court. See Rowe, supra note 8, at 59, 61 (noting consensus among lower courts on the alienage jurisdiction question and viewing Supreme Court involvement as unlikely).

181 See supra notes 50-55 and accompanying text (discussing the decision in Abbott Laboratories).

182 According to Judge Pollak, writing in Russ v. State Farm Mutual Automobile Insurance Co., 961 F. Supp. 808 (E.D. Pa. 1997), the issue presented by the supposedly clear text of section 1367(b) was "almost impenetrable to any audience other than specialists in civil procedure in the federal courts." Id. at 819. In such a context, Judge Pollak suspects that most members of Congress would have relied upon the legislative history as the definitive extrinsic aid to statutory meaning. See id. (noting that the House Judiciary Committee issued a detailed report discussing section 1367). For the federal
decision certainly proceeds upon the assumption that Congress has asserted control over the rules of supplemental jurisdiction and must deal with the consequences of its own mistakes. A similar assumption underlies the ALI Draft.

I suspect the law of federal jurisdiction will lose much if this rigorous textualism takes hold. Consider that the doctrine of supplemental jurisdiction—so essential that Congress acted quickly in the wake of Finley to provide the necessary statutory authority—evolved through the exercise of judicial discretion in the interpretation of jurisdictional law. One need not applaud each instance of this law to recognize that the rigorous textualism of Finley and Abbott Laboratories will ultimately displace the judicial role. The loss from such a pruning of judicial discretion could be considerable, leaving Congress with little guidance on jurisdictional issues and few incentives to approach them. One certainly has difficulty in seeing what would have moved Congress to work out the details of pendent-claim jurisdiction had the Court been unwilling to show the way in Gibbs.

CONCLUSION

I am unwilling to present the story of section 1367 as a morality play, in which the forces of good interpretation went to war with the forces of bad interpretation and lost. Reality rarely conforms to such simplistic assessments, and we have already witnessed some of the unfortunate consequences of rhetorical overkill. Yet it remains true that a statute Congress designed to achieve rather modest goals may well result in unsettling and confusing the law of supplemental jurisdiction to such an extent that it will require further congressional repairs. I

courts, to assert jurisdiction against the weight of such clear history was to say to Congress: "We know what you meant to say but you didn't quite say it... [B]etter luck next time." Id. at 820 (internal quotations omitted). Judge Pollak notes that under such a "gotcha" approach, Congress and the courts play antagonistic rather than coordinate roles in the government of the United States. See id. (drawing support for his decision from the need for Congress and the courts to work together as "parts of a single government").

A good many observers share this concern. See Cooper, supra note 8, at 153 (supporting a regime of broadened judicial discretion and arguing that the detailed codification of the ALI cannot answer every question); Lilly, supra note 8, at 189 (suggestions that because "issues of supplemental jurisdiction arise in such varied contexts... their resolution is ill-suited to statutory treatment"); Shapiro, supra note 8, at 218 (questioning the need for detailed codification and continuing to support development through case law).
have presented an alternative, sympathetic interpretation in the hope that these results can still be avoided.
ELECTION 2000: POINT / COUNTERPOINT SERIES

As the presidential election of 2000 is upon us, we present to the readers of Volume 148 a series of essays addressing law and the American political system. Each issue of the volume will include two essays—in the form of a point/counterpoint debate—on a given subject. We have developed this series in an effort to raise the level of civic discourse during an election cycle critical to our nation's future.

Future issues of Volume 148 of the University of Pennsylvania Law Review will include discussions of the following topics:

Issue 2: Campaign Finance Reform
Issue 3: Urban Sprawl
Issue 4: Regulation of Guns as Consumer Products
Issue 5: War Powers Act

The following is the first exchange in our series. These essays address the proper role of legal academics in the political process. Here, Professors Neal Devins and Cass Sunstein present a provocative debate framed by the letter-writing campaign by law professors to influence the 1998 impeachment of President Bill Clinton. Please note that Professor Sunstein responds directly to Professor Devins's critique. Professor Devins's rebuttal is contained in the footnotes of his piece.
BEARING FALSE WITNESS: THE CLINTON IMPEACHMENT AND THE FUTURE OF ACADEMIC FREEDOM

NEAL DEVIŃS

Academic freedom may prove to be one of the casualties of the Clinton impeachment. By signing letters about the constitutional standards governing impeachment, an issue most of them know very little about, many academics placed partisanship and self-interest above all else. The logic of academic freedom, however, cannot be squared with academics who see celebrity and power as more important than the pursuit of truth. Grounded in the belief that academics searching for knowledge in free universities will strengthen a free society, academic freedom insulates the academy from political attack. It also gives credibility to the writings, testimony, court filings, talk show appearances, and other activities of academics who seek to influence public policy. At the least, academic freedom conveys the message that scholars who speak out on public issues know something about those issues. When academics join forces to send a purely political message, their reputation as truth-seekers will diminish and, with it, their credibility. While that day has not yet arrived, it is rapidly approaching. Accusations of political correctness run amok and goofiness are becoming increasingly mainstream. Unless academics

† Goodrich Professor of Law and Lecturer in Government, College of William and Mary. Special thanks to Mary Sue Backus for her research assistance, counsel, and good cheer. Thanks also to John Duffy, Lou Fisher, Mike Fitts, Mike Gerhardt, Mike Klarman, John McGinnis, Alan Meese, Tom Merrill, Bob Nagel, David Rabban, Jeremy Rabkin, Suzanna Sherry, and participants at a University of Georgia workshop for their help.

1 Another casualty, of course, is the independent counsel statute, the Ethics in Government Act, which expired on June 30, 1999. See David Johnston, Attorney General Taking Control as Independent Counsel Law Dies, N.Y. TIMES, June 30, 1999, at A1 (noting that support for the statute eroded because of accusations by both parties that it was being used as a “political weapon against incumbents”).

2 See infra notes 15-19, 83-85 and accompanying text (arguing that society accords academics, and perhaps judges, privileges it accords no one else).

3 See Ray Suarez, Too Many in Academe Stayed Grandly Above the Fray, CHRON. OF HIGHER EDUC., Mar. 6, 1998, at B8 (saying that in the academy, “[e]very example of
can answer these charges, they risk becoming irrelevant. Consequently, when a significant number of law professors and historians hold themselves out as experts when they are not, they mislead, and all academics pay a price. For this very reason, academics can ill afford another nail to be placed in the coffin of the dispassionate academic expert. Rather, they must hold politically motivated professors accountable for abusing academic freedom.

* * *

The practice of bolstering a political argument with the help of an academic expert was on full display during the Clinton impeachment. In explaining why he could not have committed a "high crime and misdemeanor," the President asked rhetorically: "[W]hy did nearly 900 constitutional experts say that they strongly felt that this matter was not the subject of impeachment?" The President, of course, was referring to two open letters sent on his behalf—one from 400 historians; the other from more than 430 law professors.

The President was hardly alone in referring to these letters. His lawyers entitled a section of their House Judiciary Committee submission "Recent Statements by Historians and Constitutional Scholars Confirm that No Impeachable Offense is Present Here." Along the same lines, Democrats contended that the "scholarly support for the [President] . . . is overwhelming, and it cannot be ignored" while Republicans complained that the letters "substitute[d] political opinion for scholarly analysis."

A decade earlier, 2,000 legal academics banded together in an...
other letter-writing campaign. The subject: Robert Bork. The objective: to communicate that Bork’s thinking was outside the constitutional mainstream. This campaign also hit pay dirt. Finding Judge Bork “[o]utside the [t]radition of Supreme Court [j]urisprudence,” the Senate Judiciary Committee Report highlighted this “unprecedented” opposition to the nomination.10

Beyond Bork and impeachment, academics have written joint letters on abortion, affirmative action, bankruptcy reform, campaign finance, copyright reform, gun control, international human rights, supermajority requirements, the nomination of federal court judges, and much more.11 Portraying their signatories as “concerned legal scholars,” “constitutional scholars,” “professors of law,” and “professors of bankruptcy and commercial law,”12 these letters are intended to communicate the consensus opinion of academic experts.13 While these letters have not overtaken expert testimony and individual letters to Congress, letter-writing campaigns have become an increas-

---

10 Senate Comm. on the Judiciary, Nomination of Robert H. Bork to Be an Associate Justice of the United States Supreme Court, Exec. Rep. No. 100-7, at 6-7, 13 (1987). By highlighting the attention these letters received, I am not suggesting that academic letter writing changes votes in Congress. Rather, academic letter writing provides rhetorical cover to members of Congress who have already made up their minds. See infra notes 63-68 and accompanying text (discussing the prominent role played by congressional staffers in organizing academic letter-writing campaigns).


12 See Hearings, supra note 6, at 374 (referring to “professors of law”); Scholars’ Statement, supra note 11, at 1711 (referring to “constitutional scholars”); infra note 63 (noting that the signatories were “concerned legal scholars”); see also Draft Letter to Senators Orrin Hatch and Patrick Leahy (referring to “professors of bankruptcy and commercial law”) (on file with the University of Pennsylvania Law Review).

13 The fact that only academics can sign these letters also signals that letter signers are speaking as experts, not concerned citizens. For further discussion, see infra notes 88-91 and accompanying text.
ingly important mechanism for academics to send a message to Congress. Why, though, do people pay attention to these letters? Why treat these letters with more deference than, say, a petition from the ACLU or the NRA? The answer, of course, is that academics have a reputation for placing the search for truth ahead of partisanship. Unlike movie stars, interest groups, or the person on the street, the credibility of academics is tied to their purported willingness to speak "[t]ruth to [p]ower."

Society, acting on this vision, accords academics certain privileges that it accords no one else (except perhaps judges). Academic freedom, tenure, sabbaticals, and the like encourage academics to think independently and to challenge prevailing norms through their scholarship. At the same time, the trust that society has placed in academics, as well as the resources it has provided them, are grounded in certain assumptions about academic conduct. Academics, for example, have an obligation "to speak truthfully about the issue at hand, because they have a detached cast of mind as well as a large stock of relevant and reliable knowledge on the subject at issue." Correspondingly, before speaking as experts, academics have an obligation to read and to think about arguments on both sides of an issue. The ways of the scholar, as Alexander Bickel put it, "appeal to men's better natures" because they are about the leisure of thinking, training, and insulation, not "the moment's hue and cry."

Whether or not academics live up to this obligation, Bickel's vision still resonates with much of the public. For this very reason, policy

---


15 Like academics, judges have a reputation for dispassionate expertise and can ill afford for that reputation to give way to the view that judicial decision-making is simply another form of politics. See THURMAN W. ARNOLD, *THE SYMBOLS OF GOVERNMENT* 199-206 (Yale Univ. Press 1935) (expressing the view that society accepts the judgment of judges so long as they think that judges are acting like lawyers—reading law and applying it—instead of politicians).

16 See David M. Rabban, *Academic Freedom, in ENCYCLOPEDIA OF THE AMERICAN CONSTITUTION* 12-13 (Leonard W. Levy et al. eds., 1986) (exploring the theoretical underpinnings of academic freedom). For further discussion, see infra notes 70-72 and accompanying text (arguing that an assumption of dispassionate truth-seeking is the "first obligation" of academics).


19 In a 1997 Gallup Poll, college teachers were ranked fourth highest for honesty
makers and media outlets seek out academics on many of the issues that divide the nation. Academic letter-writing campaigns likewise capitalize on the academic's reputation for dispassionate expertise. Consider, for example, the anti-impeachment letters. Writing "neither as Democrats nor as Republicans" (but as "professors of law"), these citizen scholars saw the drive to impeach the President as a threat to "our constitutional order." Signed by many of the nation's most prominent law professors and historians, it is no wonder that these letters were taken seriously by the President's supporters as well as his foes. Upon closer inspection, however, these letter-writing campaigns are little more than a testament to the willingness of many academics to pawn off fake knowledge.

Of the 900 signers of the anti-impeachment letters, for example, it is doubtful that many had thought seriously about the constitutional standards governing impeachment. Impeachment, at least until this past year, is a subject that is rarely written about and rarely taught. Indeed, nearly all of the legal academics who testified before the House Judiciary Committee were better known for their allegiance to either liberal or conservative causes than for their scholarship about impeachment. For this very reason, these academics have been
roundly criticized both for the quality of their constitutional analysis and for "conducting a transparently political debate in constitutional terms." 24 Far more significant, most of the historians who signed the letter were not constitutional specialists. 25 Among the law professors, only one-third of the signatories teach constitutional law. 26

Even among professors of constitutional law, moreover, there is no reason to think that these individuals have "some expertise on the topic of impeachment." 27 Consider, for example, professors (such as myself) who have used Cass Sunstein's constitutional law casebook. Just over one page of this 1800-page tome considers the constitutional standards governing impeachment. 28 And that one page provides ab-

---

24 Michael J. Klarman, Constitutional Fetishism and the Clinton Impeachment Debate, 85 VA. L. REV. 631, 631 (1999); see also John O. McGinnis, Impeachable Defenses, POL'Y REV., June & July 1999, at 27, 27-29 (criticizing the reasoning and methodology of law professors' positions on Clinton's impeachment). It is also noteworthy that scholars who argued against the appropriateness of originalism at the Bork hearings made use of originalism (and little else) in arguing against the Clinton impeachment. See Philip Elman, Shame on the Partisan Professors, LEGAL TIMES, Nov. 16, 1998, at 21 (noting this reversal of position).

25 See Jesse Lemisch, Anti-Impeachment Historians and the Politics of History, CHRON. OF HIGHER EDUC., Dec. 4, 1998, at B6 (discussing the advertisement taken out by scholars opposing the impeachment of President Clinton).

26 According to the biographical data contained in the 1998-99 AMERICAN ASSOCIATION OF LAW SCHOOLS' DIRECTORY OF LAW TEACHERS, 130 of the 452 signers of the law professor letter list constitutional law as among the subjects that they teach. For this very reason, the anti-impeachment letter is cut from a different cloth than, say, letters on bankruptcy or copyright reform. In those cases, only individuals familiar with the letter's subject matter were allowed to sign and, consequently, some mechanism was in place to screen the bona fides of letter signers.

27 Cass R. Sunstein, Professors and Politics, 148 U. PA. L. REV. 191 n.6 (1999). Sunstein "assume[s]" that anyone who teaches "some aspect of constitutional law as part of their curricular responsibilities" has impeachment-related "expertise." Id. Whether or not corporations professors who teach about corporate free speech would qualify, Sunstein's list of assumed experts certainly includes faculty who teach courses on church and state, freedom of speech, or state constitutional law. Sunstein also assumes that, in the wake of Watergate, many law professors "developed genuine, if fairly general, views on the appropriate meaning of 'high crimes and misdemeanors.'" Id. at text accompanying notes 12-13. But this assumption is implausible. For some law professors (myself included), Watergate is a distant prepubescent memory. And for those who do remember, there is little reason to think that their undoubtedly "genuine" views measure up to the standard of expert academic opinion (especially since perjury on a private matter was not a critical part of the Watergate drama). See also infra notes 95-96 and accompanying text (detailing the paucity of impeachment-related expertise).

28 See GEOFFREY R. STONE ET AL., CONSTITUTIONAL LAW 423-24 (3d ed. 1996) (discussing constitutional standards governing impeachment). As to why so little space is given to impeachment: Casebooks focus on case law. The substantive standards governing impeachment are a nonjusticiable political question and, consequently, there is no case law on the subject.
olutely no guidance in assessing the appropriateness of the Clinton impeachment. Consequently, whether or not the question raised in the Clinton impeachment was "close," it is doubtful that professors of constitutional law—let alone all law professors—were well positioned to render an expert opinion on the subject.

At one level, the lesson here is simple. Many of the law professor and historian signatories were animated by partisanship and self-interest, not scholarship. Needless to say, there is a real temptation for academics who want to be part of the fray, who want to see their names in print, who want to tell their families that they did something that mattered, to sign a mass letter. Other academic letter signers may not care at all about celebrity. They may, however, care a great deal about the President's ability to pursue his agenda. In particular, partisan Democrats who voted for the President and support his policies may sign the letter for political reasons. As it turns out (surprise),

---

29 Instead, the focus of this entry is that "there is no clear answer to the central question: What is the meaning of the phrase 'high Crimes and Misdemeanors'?" Id. at 423. More striking, the Teacher's Manual accompanying this casebook asks: "Why has the House used the impeachment route so rarely? One might explore the possibility that impeachment for quasi-political reasons might be a good idea . . . ." GEOFFREY R. STONE ET AL., TEACHER'S MANUAL TO CONSTITUTIONAL LAW 71 (3d ed. 1996). The Sunstein casebook is hardly unique. Some casebooks do not mention impeachment at all. See, e.g., DANIEL A. FARBER ET AL., CONSTITUTIONAL LAW: THEMES FOR THE CONSTITUTION'S THIRD CENTURY (2d ed. 1998). Other casebooks dedicate only a page or two to the topic. See, e.g., GERALD GUNTER & KATHLEEN M. SULLIVAN, CONSTITUTIONAL LAW 411-13 (13th ed. 1997).

30 Sunstein, supra note 27, at text accompanying note 4. For arguments that the constitutional question was anything but obvious, see Klarman, supra note 24, at 657-58, and RICHARD A. POSNER, AN AFFAIR OF STATE (1999). See also Symposium, Background and History of Impeachment, 67 GEO. WASH. L. REV. 601, 601 (1999) (Foreword by Representative Henry Hyde, Chairman, House Committee on the Judiciary) (highlighting divergent thinking among "academic experts" on this question). If something is truly debatable, moreover, it is especially important that letter signers are truly experts. Otherwise, there is great risk that their conclusions simply will be a reflection of their personal preferences. More significant, the fact that some experts do not think the constitutional question is close does not mean that those who are not experts can sign on as experts. For further discussion, see infra notes 95-96 and accompanying text (arguing that law professors are not experts on all legal issues merely by virtue of their status as law professors).

31 Sunstein recognizes that "[f]or some of the signatories, perhaps this is true." Sunstein, supra note 27, at text accompanying notes 12-13. Sunstein does not consider the possibility that mass letters may register bias and, consequently, are a poor device for communicating scholarly expertise.

32 The fact that the academy's "A" team signed onto these letters made it easier to join the chorus. After all, if Arthur Schlesinger, Laurence Tribe, and others are willing to put their stamps of approval on these letters, there is reason to think that the letter's argument is at least plausible.
the academy is overwhelmingly left-liberal, overwhelmingly Democratic.\textsuperscript{33} Take the case of the law professors.\textsuperscript{34} "Only [ten] percent of [them] characterize themselves as conservative to some degree,"\textsuperscript{35} while more than eighty percent of them are registered Democrats.\textsuperscript{36} Therefore, many legal academics see Kenneth Starr—who argued against abortion rights and affirmative action as the Bush administration’s Solicitor General—as their nemesis. At a panel on impeachment at the 1999 American Association of Law Schools convention, for example, law professors loudly booted when it was revealed that one of the panelists, John McGinnis, clerked for then-D.C. Circuit Judge Starr. For these professors, signing a letter that would place Starr’s views on impeachment “outside the legal mainstream” would be manna from heaven.

This is not to deny that some letter signers signed on for reasons other than partisanship.\textsuperscript{37} But it simply begs the question to “assume” (as do defenders of the letter-writing campaign) that dispas-

\textsuperscript{33} A 1989 survey by the Carnegie Foundation for the Advancement of Teaching found that, at research universities, 67% of faculty identify themselves as liberal, and 17% identify themselves as conservative. See Seymour Martin Lipset, \textit{The Sources of Political Correctness on American Campuses}, in \textit{THE IMPERILED ACADEMY} 71, 79 (Howard Dickman ed., 1993) (arguing that political opinions on American campuses are shifting to the left). In the humanities and social sciences, less than five percent of faculty identify themselves as conservative. See \textit{RICHARD PIPES, PROPERTY AND FREEDOM} 272 (1999) (discussing political diversity in university faculty).

\textsuperscript{34} Historians too have a reputation for left-leaning partisanship. See Thomas Haskell & Sanford Levinson, \textit{Academic Freedom and Expert Witnessing: Historians and the Sears Case}, 66 TEX. L. REV. 1629, 1630 (1988) (describing and analyzing “vilification” of award-winning historian Rosalind Rosenberg for providing expert testimony for Sears, Roebuck in a sex discrimination lawsuit).


\textsuperscript{36} James Lindgren, Measuring Diversity, Speech to the National Association of Scholars (Jan. 5, 1997). More striking, although 14.9% of full-time working women are Republicans, 0.5% of women law faculty are Republicans. Id. For Pierre Schlag, it is “[o]nly because legal scholars speak within a stratified and homogeneous community can they present their views as the ‘work of reason itself.’” Book Note, \textit{Let Us Reason Together}, 112 HARV. L. REV. 958, 959 (1999) (quoting PIERRE SCHLAG, \textit{THE ENCHANTMENT OF REASON} 38 (1998)).

\textsuperscript{37} Of the 452 letter signers, Cass Sunstein identifies three who are “not normally characterized as ‘left-liberal . . .’” Sunstein, \textit{supra} note 27, at note 21 and accompanying text. No doubt, there are others (non-left liberals) in addition to these three. I am equally confident, however, that the overwhelming majority of signatories are—like the academy itself—left-liberal. Furthermore, for some letter signers, the anti-impeachment effort served their self-interest. Specifically, since (mostly liberal) academics embrace both “judicial activism” and “big government,” the scholarship and advocacy of academics is taken more seriously by Clinton appointees than by Republicans in Congress.
sionate expertise animated law professor signatories. With no evidence of preexisting expertise on impeachment and with ample evidence that most law professors are left-liberal Democrats, the possibility of partisanship seeping into the anti-impeachment campaign is anything but remote. For this reason, letter organizers cannot rely on assumptions; instead, they must explain why it is that letter signers were qualified experts. Otherwise, accusations of partisan bias (like the one levied in this article) may well stick.

For their part, organizers of the letter-writing campaign paid far more attention to increasing their ranks than to screening the bona fides of letter signers. At my law school, for example, a professor of civil procedure (and signer of the letter) sent a faculty-wide e-mail distributing the letter and explaining how to sign on to it. Through an e-mail sent to me (and most other professors of constitutional law) by Cass Sunstein, I was also invited to sign on to a companion letter. Although telling me that "[e]very signature really counts," this invita-

---

53 This, of course, is what Cass Sunstein does in his response. For Sunstein, we should "assume that those with whom we disagree are acting in good faith on the basis of evidence that they honestly believe to be sufficient." Sunstein, supra note 27, at text accompanying note 24.

54 See supra notes 33-37 and accompanying text (discussing political views of legal academics, including evidence that partisanship may have figured into law professor opposition to the Starr Report); infra notes 47, 50 and accompanying text (discussing the role played by People for the American Way in both the historians' anti-impeachment campaign and the law professors' campaign against Bork); infra notes 63-66 (noting the critical role played by members of Congress in spurring on academic letter-writing campaigns).

40 And, the easier it is to typecast the academy this way, the more vulnerable the academy is to attack. See infra notes 97-106 and accompanying text (arguing that "[o]ver time, the academic ethic may give way to the view that self-interest and partisanship are the coins of the academic realm").

41 The historian letter-writing campaign used e-mail. The letter's author, Sean Wilentz of Princeton, sent the letter to 30 or 40 like-minded historians who then distributed it to their colleagues. See Mike Feinsilber, 400 Historians, Rodino Resist Impeachment, BUFFALO NEWS, Oct. 29, 1998, at A8 (noting that Wilentz e-mailed 30 or 40 historians he thought might be dismayed at the implications of an attempt to impeach Clinton, and within three days, 300 historians agreed to sign a statement against such an attempt).

42 A not-so-random survey of constitutional law professors suggests that individuals disinclined to sign the e-mail were kept off the distribution list. For example, only one of the four conservative law professors whom I contacted (Steve Calabresi, John McGinnis, Mike Paulsen, and Eugene Volokh) had received the e-mail. Also, Mike Gerhardt, an impeachment scholar who refused to take sides, did not receive the e-mail.

43 E-mail from Cass Sunstein to Neal Devins (Oct. 29, 1998) (on file with the University of Pennsylvania Law Review). The text of the letter reads as follows:

The undersigned professors of law come from different political parties and
tion did not contain any analysis to support its one-sentence recommendation. Finally, through a posting on the JURIST web page, the organizers of the law professor letter circulated a follow-up letter (that any law professor or historian could sign) which called upon the President to resist calls for his resignation rather than give in and allow his resignation to "fundamentally transform the impeachment device." 

None of these open letters made expertise a prerequisite for signing. The reason, of course, was that impeachment was too politically charged for a letter signed by, say, twelve leading academics to make a difference. And making a difference is what letter organizers cared most about. Along the same lines, recognizing that their individual views on impeachment were of little consequence, letter organizers decided that they had to act like an interest group. For this very reason, letter organizers did more than gather names. They worked hard at publicizing their efforts. Anti-impeachment historians, for example, made effective use of a press conference to release their letter. Also, with the assistance of People for the American Way, the histori-
ans took out a full page advertisement in *The New York Times.*47 Mincing no words, the advertisement argued that, if the President were convicted, the presidency would be "permanently disfigured" and the Constitution "undermine[d]."48

While organizers of the anti-impeachment campaign made greater use of hyperbole and technology than previous lobbying efforts had, they sought inspiration and guidance from the highly successful academic campaign against Robert Bork. The anti-Bork campaign revealed what grass roots lobbyists have long known, namely, that there is strength in numbers. How better to communicate that Robert Bork was outside the constitutional mainstream than for 2,000 law professors openly to oppose the nomination? At first, organizers approached law school deans and professors who had taught constitutional law for five years or more ("except those who were known to be supporting Bork")49 to join the anti-Bork campaign. This effort proved so successful that it was expanded to all law professors. A contact person at most law schools was identified and that contact person solicited signatures from her colleagues. I was contacted this way, as were most of my colleagues.

Partisanship, of course, figured prominently in the campaign against Bork. To begin with, the academic campaign was spearheaded by Ricki Seidman of People for the American Way and William Taylor of the Leadership Conference on Civil Rights.50 The key to the anti-Bork letter-writing campaign, however, was the disdain in which the legal academy held Bork. Unlike the standards governing impeach-

47 See Nat Hentoff, *An Entirely New Impeachment Case*, WASH. POST, Mar. 6, 1999, at A21 (noting that People for American Way assisted Historians in Defense of the Constitution by enabling the list price of the New York Times advertisement to be reduced from $75,948 to $56,000).


49 NORMAN VIEIRA & LEONARD GROSS, *SUPREME COURT APPOINTMENTS: JUDGE BORK AND THE POLITICALIZATION OF SENATE CONFIRMATIONS* 143-44 (1998). For further discussion, see *infra* note 109 (discussing Sunstein's view that people should only sign petitions if they can defend the relevant position publicly but that they need not necessarily defend their positions as academics).

50 See MARK GITENSTEIN, *MATTERS OF PRINCIPLE: AN INSIDER'S ACCOUNT OF AMERICA'S REJECTION OF ROBERT BORK'S NOMINATION TO THE SUPREME COURT* 160 (1992) (noting that Seidman and Taylor led the "effort among the anti-Bork forces to recruit academics"). These two had previously coordinated a similar (but unsuccessful) campaign against Daniel Manion, a conservative nominated to the Seventh Circuit Court of Appeals. *See id.* (indicating that, while Seidman and Taylor had "little trouble drumming up opposition" to Manion's nomination in academic circles, Manion's nomination was ultimately approved with a one-vote margin).
ment, Bork's theories were well known to many academics. Nevertheless, it is doubtful that experts in commercial law, evidence, tax, securities, and the like were well versed in Bork's theories. Rather, many academics unfamiliar with Bork's writings opposed him because of where he would take the Court and because they feared that Bork's confirmation would strengthen the then burgeoning Federalist Society and, with it, the power of conservatives in the legal academy. In other words, left-leaning academics saw Bork as a threat to their status and influence. In particular, his confirmation would make their scholarship and advocacy less relevant because his views did not mesh with their own. Bork, moreover, antagonized many legal academics during his tour of duty in the Nixon Justice Department, which included the firing of Watergate special prosecutor Archibald Cox and the pursuit of Nixon's vendetta against Warren Court liberalism.

Partisan letter-writing campaigns are likely to continue, especially among legal academics. Not only are these letters highly visible and somewhat influential, but they also allow the rank and file of the academy to join forces with the academy's glitterati in a common cause. What better way to make oneself part of the "A" team than to sign off on the constitutional analysis of Ackerman, Sunstein, Tribe,

51 For this reason, many of the people who signed letters opposing the nomination did so, in part, because they thought Bork's writings were too rigid and too self-righteous for him to succeed on the Court. An excellent treatment of this subject (which does not pass judgment on whether or not Bork should have been confirmed) can be found in ROBERT F. NAGEL, JUDICIAL POWER AND AMERICAN CHARACTER 27-43 (1994), which discusses Bork's relationship with what his critics defined as the mainstream, and more specifically, how Bork's "critics define the mainstream in terms of principle and accuse Bork of standing outside it as a covert practitioner of conservative politics through judicial power." Id. at 30.

52 See GITENSTEIN, supra note 50, at 161 (noting that professors were concerned about the possibility of Bork's appointment because they saw him as a symbol of the unraveling of the civil rights and civil liberties the Supreme Court had expanded over the previous 30 years). For Bork, that battle over his confirmation pitted "left-liberal" "intellectual class values" against populists (like him?) who believe in the primacy of elections. See ROBERT H. BORK, THE TEMPTING OF AMERICA 337 (1990) ("The battle was ultimately about whether intellectual class values, which are far more egalitarian and socially permissive, which is to say left-liberal, than those of the public at large and so cannot carry elections, were to continue to be enacted into law by the Supreme Court.") (footnote omitted).

53 For a provocative argument that anti-Bork academics had a moral duty to oppose the nomination of Anthony Kennedy (but did not do so because they held a grudge against Bork and only Bork), see George Kannar, Citizenship and Scholarship, 90 COLUM. L. REV. 2017, 2060-61 (1990) (book review) (arguing that the American public's interest in making fully informed decisions—even wrong ones—required legal scholars who had opposed Bork's nomination to take what was a nonconformist stand and publicly oppose Kennedy's nomination).
and the like? Along these lines, Richard Posner described the law professor anti-impeachment letter as "a form of herd behavior (the 'herd of independent minds') by the animal that likes to see its name in print." The appeal of letter writing, moreover, is fueled by the proliferation of media outlets and, with it, the opportunity for many academics to achieve their fifteen minutes of fame. Today, academics seek fame through talk show appearances, op-ed pieces, and trade press books. In this era of sound bite scholarship, it is little wonder that being part of the story is far more appealing than writing about it some years later.

In contrast, there is a growing perception among academics that court-ordered social reform is a hollow hope. Relatedly, the continuing conservatism of the Supreme Court and of many federal courts of appeal suggests that law review scholarship calling for novel judicial solutions to social problems will fall on deaf ears. Perhaps for this

54 POSNER, supra note 30, at 242. Some of the letter's signatories, of course, signed on because their analysis of both the constitutional standards governing impeachment and the Starr Report convinced them that the President's conduct did not warrant removal from office. Many academics did not sign on to the letter because they thought it unduly partisan. Nevertheless, Judge Posner is correct in referring to the herd mentality of the legal academy. A significant portion of the legal academic community signed on to these letters without independent knowledge of the constitutional standards governing impeachment. See supra notes 20-38 and accompanying text (discussing the possible political motivations and constitutional expertise of academics who signed letters about impeaching Clinton). Rather than stand as a roadblock to such partisanship, cultural norms within the academy encouraged these letter signers to see themselves as an interest group, not as free thinkers committed to the pursuit of truth. See infra notes 41-45 and accompanying text (discussing how letter organizers cared only about increasing the number of signatories); infra notes 87-94 and accompanying text (discussing the strong incentives of hegemony within the academy).

55 I agree with Cass Sunstein that none of these things is inappropriate. See Sunstein, supra note 27, at text accompanying notes 4-11. Like Sunstein, I also think it is perfectly fine for academics (through testimony, letter writing, whatever) to speak as experts before Congress. But academics should only speak as experts about matters on which they have invested the time and energy necessary to hold themselves out as experts. See infra notes 95-104 and accompanying text (contrasting Sunstein's definition of academic experience from my own).

56 See GERALD N. ROSENBERG, THE HOLLOW HOPE: CAN COURTS BRING ABOUT SOCIAL CHANGE? 338 (1991) ("U.S. courts can almost never be effective producers of significant social reform. At best, they can second the social reform acts of the other branches of government.").

57 For left-leaning academics, populist constitutional discourse is now preferred to Court-centered social policymaking. See CASS R. SUNSTEIN, ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT 3-6 (1999) (identifying and advocating the current Court's "judicial minimalism"); MARK TUSHNET, TAKING THE CONSTITUTION AWAY FROM THE COURTS 181-82 (1999) (explaining constitutional law as a populist narrative and advocating a populist constitutional legal system based on
reason, a 1998 report of the Twentieth Century Fund deemed letter-writing an essential ingredient of its social reform agenda.58 For the most part, these campaigns, like the academy itself, advance liberal causes.59 Indeed, conservatives are best off not going head-to-head against liberal letter writers. Why gather 100 signatures in support of Bork, as did the right of center Ad Hoc Committee for Principled Discussion of Constitutional Issues?60 A twenty-to-one disparity, rather than serve as effective counter-speech, will simply prove an embarrassment.61 With that said, right-of-center interests are still likely to

the Constitution and Declaration of Independence); Robin West, The Aspirational Constitution, 88 Nw. U. L. Rev. 241, 245-46 (1993) (acknowledging that while such a shift is unlikely to occur, it would constitute a shift from a judicially-enforced Constitution of limits to a congressionally-enforced Constitution of aspirations and would help modern progressive causes). For conservatives the legal academy has only itself to blame: “it has become a heavily normative body of advocacy scholarship targeted at the federal courts with the goal of influencing them to do things that they are extremely unlikely to do in the current political and social climate.” Steven G. Calabresi, The Crisis in Constitutional Theory, 83 Va. L. Rev. 247, 266 (1997) (reviewing LOUIS MICHAEL SEIDMAN & MARK V. TUSHNET, REMNANTS OF BELIEF (1996)).

58 TWENTIETH CENTURY FUND WORKING GROUP ON CAMPAIGN FINANCE LITIGATION, BUCKLEY STOPS HERE: LOOSENING THE JUDICIAL STRANGLEHOLD ON CAMPAIGN FINANCE REFORM, 100-01 (1998) (referring to a letter sent by more than 200 legal academics on campaign finance reform and contending that legal scholars could play a critical role in campaign finance reform by signing on to such statements).

59 Up until now, the most visible letters have all advanced left of center causes: affirmative action, abortion rights, international human rights, the defeat of conservative judicial nominees, the defeat of supermajority rules, and the defeat of the Clinton impeachment. No doubt, as Sunstein argues in his response, the principles enunciated in the anti-impeachment letter would apply to a future Republican president. See generally Sunstein, supra note 27, at note 5 (contending that the academics involved in congressional discussions of the possibility of impeaching Clinton were not working “for” or “with” the White House). But would Cass Sunstein organize a letter-writing campaign to save, say, Ronald Reagan? And if he did, would the same 452 law professors sign on? See Sunstein, supra note 27, at note 19 (recognizing that it is possible but speculative that personal opinion would affect letter-writing campaigns).

60 This letter, sent to Senators Robert C. Byrd and Robert Dole, is reprinted in 133 Cong. Rec. S28853 (daily ed. Oct. 22, 1987). With respect to impeachment, a coalition of 96 academics and former government officials, including Bork, William Bennett, and Edwin Meese, wrote to Congress that the Starr Report did, in fact, support the President’s impeachment. See Don’t Let the President Lie With Impunity, WALL ST. J., Dec. 10, 1998, at A22 (reprinting a letter signed by academics, lawyers, and former government officials which was distributed to the House Judiciary Committee and urged the House of Representatives to impeach Clinton). Needless to say, my criticism of the anti-Bork and anti-impeachment letter-writing campaigns might well apply to the pro-Bork and pro-impeachment campaigns. In particular, if signatories signed on as non-expert partisans, those signatories (and quite possibly the organizers of these campaigns) would have violated the academic ethic.

61 Indeed, the 100 or so pro-Bork, pro-impeachment letter signers were so outnumbered that they appeared well outside of the mainstream and, consequently,
launch some letter-writing campaigns. Like anti-Bork and anti-impeachment academics, conservative activists may well seek strength in numbers, not academic expertise. For example, in June 1999, a letter opposing proposed gun control legislation was distributed to law professors and other academics over e-mail. Recipients of this e-mail were told: "If everyone who we are sending this to can get even a couple other people in your department to sign this, we will end up with well over a few hundred signatures." It did not matter whether possible signatories were well versed in the particulars of proposed legislation, in the "real costs" of waiting periods, or anything else. Like the anti-Bork and anti-impeachment letters, expertise played second fiddle to the bottom line, that is, to sending an effective political message.

* *

The question remains: Why are these letters taken seriously? While it never referred to the lobbying efforts of People for the American Way, the NAACP, or pro-choice activists, the Senate Judiciary Committee's report on Robert Bork trumpeted the views of legal academics. Likewise, the President and his supporters singled out the historian and law professor letters. Is the myth of academic expertise a drug that somehow blinds policymakers to the partisanship of (at least some) academic letter signers?

Not at all. Politicians and their staffers know full well that the citizen scholars who send these letters often have a political ax to grind. In fact, rather than being duped by make-believe academic experts, elected officials sometimes sponsor these missives! Two years ago, Congressman John Conyers's office asked me to join a "community of concerned legal scholars" in writing to Congress about intelligence spending. The draft letter was addressed to none other than Congressman Conyers. Equally striking (and far more significant), Senate

looked especially politically motivated.

62 E-mail from John R. Lott, Jr. to James Lindgren (June 3, 1999) (on file with the University of Pennsylvania Law Review).

63 Draft Letter to Congressman Conyers, United States House of Representatives (June 1997) (on file with the University of Pennsylvania Law Review). I was asked to sign on because a friend suggested my name to Conyers's office. Although Conyers's office was coordinating this letter-signing campaign, potential signatories were directed to contact the Center for International Policy. I do not know whether the Center for International Policy's involvement was a subterfuge, intended to conceal Conyers's involvement.
Judiciary Committee chair and Bork foe Joseph Biden asked Chris Schroeder, a law professor working for the committee, to help drum up law professor opposition to Bork.64 Three other law professors working with Biden—Walter Dellinger, Philip Kurland, and Laurence Tribe—negotiated with People for the American Way over the text of the law professor letter.65 Along the same lines, members of Congress were actively involved in at least one of the law professors’ anti-impeachment letters. The invitation to join the ranks of letter signers made clear that the letter was consequential because “[s]ome members of the House of Representatives have suggested that it would be very valuable for the House to hear from a large group of teachers of constitutional law on the impeachment issue.”66

For anti-impeachment House members and anti-Bork Senators, academic letters were a kind of salve. Rather than appear overly ideological and overly partisan, lawmakers can take cover in a letter signed by a thousand or more academics.67 In contrast, a letter submitted by the AFL-CIO, the National Abortion Rights Action League, or (for that matter) the Family Research Council would call attention to, not cloak, possible biases.68 In other words, rather than affect the thinking

64 See Gitenstein, supra note 50, at 161 (“Chris Schroeder, at Biden’s direction, spent much of the late summer and early fall of 1987 on the phone with scores of law professors.”).  
65 See id. at 161 (“By late August, Seidman had negotiated the text of a letter acceptable to Kurland, Tribe, and Dellinger, and began to circulate it.”).  
66 Sunstein, supra note 43 (quoting one of the letters signed by anti-impeachment law professors and sent to Congress). Sunstein never says who these members of Congress are, if he responded to the requests of both Democratic and Republican leadership, or, alternatively, if the letter-writing campaign was the brainchild of one or the other side.  
67 Lawmakers likewise can take cover behind academics who testify at hearings. These academics often are contacted because they can be counted on to state a position that supports the person who invited them to testify, usually the ranking majority or minority member of a committee or subcommittee. For example, before I was asked to testify about line-item veto legislation, a Senate Judiciary staff member called me to confirm that I still subscribed to a position articulated in an article of mine. Along the same lines, it was no accident that Republican witnesses at the House impeachment hearings testified that the President’s conduct was impeachable whereas Democratic witnesses testified that it was not. See Hearings, supra note 6 (suggesting that the split on whether the President’s conduct was impeachable was primarily along party lines).  
68 For this very reason, interest group representatives were not called on to testify at the Bork confirmation hearings. Rather, “numerous witnesses from the legal academy presented the Senators with the same critique of Bork that the interest groups would have offered, but from a more ‘disinterested’ perspective . . . .” Kannar, supra note 53, at 2041. See also Mary McGrory, The Supreme Sacrifice, WASH. POST, Oct. 6, 1987, at A2 (recounting the Democrats’ strategy of not calling upon members of inter-
of politicians, academic letter-writing enables politicians to offer high-minded reasons for saying and doing what they otherwise would have said and done. Consider, for example, the anti-impeachment letters. The votes in the House and the Senate were on largely partisan lines. It simply strains credibility to believe that the law professors’ and historians’ letters moved fence sitters one or the other way. What these letters did do, however, was to enable the President’s defenders to tell the public that their votes were cast for nonpartisan reasons (the Starr Report did not state sufficient grounds for impeachment), not for partisan reasons (Democrats need to stick together).

Here, I think, is where the true significance of these letters lies. Politicians feed off of these letters because of the so-called academic ethic, that is, the notion that “the first obligation of the university teacher is to the truth.” Academics, likewise, feed off of this reputation in justifying these letter-writing campaigns. “We law professors are free from a client’s interest, free from a place in a hierarchy, free to say exactly what we think,” explained Barbara Babcock, a former Carter Administration Justice Department official and signer of both the Bork and impeachment letters. For Sean Wilentz, principal drafter of the historians’ letter, “[t]his is not a political effort at all. . . . [It is] historians speaking as historians.” Likewise, Susan Bloch and Jed Rubenfeld, two of the organizers of the law professors’ letter, spoke of their efforts as “nonpartisan” and of the need for legal academics “to come out of our comfortable law review role to make a point.”

Letter organizers, whether or not they believe their own rhetoric, have no choice but to perpetuate this myth. Otherwise, their missives

---

69 See Peter Baker & Helen Dewar, Clinton Acquitted, WASH. POST, Feb. 13, 1999, at A1 (noting the partisanship present in the vote on the perjury and obstruction of justice charges); Peter Baker & Juliet Eilperin, Clinton Impeached, WASH. POST, Dec. 20, 1998, at A1 (noting that vote on the first article of impeachment was “largely along party lines”).

70 SHILS, supra note 17, at 49.


would be no different than those of interest groups who wear their partisanship on their sleeves. But there is good reason to doubt whether academics still think of themselves as truth seekers. In particular, the traditional image of the academic has given way to "postmodernism, multiculturalism, and political correctness." Consequently, rather than see these letter-writing campaigns as a departure from their scholarly endeavors, many academics increasingly see scholarship and partisanship as inextricably linked. In this way, the willingness of academics (who know next to nothing about impeachment) to sign on to an anti-impeachment letter is understandable. For similar reasons, tax and commercial law experts did not blink when signing a letter condemning Judge Bork's interpretive theories. Specifically, if non-self-interested knowledge does not exist, it is unavoidable that academics will embrace one partisan position or an-

---

75 David M. Rabban, Can Academic Freedom Survive Postmodernism?, 86 CAL. L. REV. 1377, 1378 (1998) (reviewing THE FUTURE OF ACADEMIC FREEDOM (Louis Menand ed., 1996)). For this very reason, defenders of academic freedom no longer speak of impartial scholars discovering objective truths; instead, as Ronald Dworkin puts it, academic freedom concerns the fundamental ethical "responsibility of [academics] to find and tell and teach the truth as they see it." Ronald Dworkin, We Need a New Interpretation of Academic Freedom, in THE FUTURE OF ACADEMIC FREEDOM 181, 190 (Louis Menand ed., 1996); see also Rebecca S. Eisenberg, Academic Freedom and Academic Values in Sponsored Research, 66 TEX. L. REV. 1363, 1363 (1988) (contending that academic truth seeking is compromised by corporate and government funding of research).

76 Before postmodernism, of course, legal realists argued that constitutional interpretation is inescapably value-laden and, as such, that constitutional analysis would always be driven by "a particular set of policy preferences that cannot be distinguished from the preferences expressed in other political forums." LOUIS MICHAEL SEIDMAN & MARK V. TUSHNET, REMNANTS OF BELIEF 42 (1996); see also James O. Freedman, The Bully Lectern, HARV. MAG., Jan.-Feb. 1999, at 36, 36-37 (discussing some of the ways that early twentieth-century college presidents involved themselves in the world of politics). Perhaps more significant, many academics who sign letters do so for reasons that have nothing to do with post-modernism. See infra notes 86-91 and accompanying text (detailing some of the reasons, other than agreement with the letters' underlying reasoning, why academics signed onto the anti-impeachment letter).

77 It is true, however, that tax and commercial law scholars (if they inquired) could spot substantial deviations between Bork's constitutional views and those of their liberal constitutional law colleagues. For this reason, tax and commercial law scholars could place Bork outside of the constitutional mainstream within the academy. But the claim of the anti-Bork campaign was that Bork was outside of the constitutional mainstream as defined not just by the academy, but as defined by the Supreme Court. Here, some expertise about differences between Bork's writings (including his decisions as a D.C. Circuit judge) and Supreme Court decision making would be required. More to the point, anti-Bork letter signers would need to be able to explain why it is that White House claims that Bork's writings were within the mainstream were incorrect. See The White House Report: Information on Judge Bork's Qualifications, Judicial Record & Related Subjects, reprinted in 9 CARDOZO L. REV. 187, 188 (1987) (describing why Bork should be considered in the "mainstream of American jurisprudence").
other.

This postmodernist dilemma is especially acute in the legal academy. Compare, for example, the academy’s reaction to conflicting arguments over the attainability of truth through legal scholarship. When Michael Seidman and Mark Tushnet wrote that it is “apparent to everyone [that all constitutional] arguments can [and will] be manipulated to advance the particular policy goal of the advocate who makes them,” no one rose up to complain. In contrast, Paul Carrington prompted a near crisis in the academy by arguing that law professors should believe that law and legal texts matter.

This discomfort with truth-seeking is easily explained. Unlike chemistry or psychology, law is not a science. Instead, lawyers translate the knowledge, experience, and expertise of other professionals.

---

78 Seidman & Tushnet, supra note 76, at 90. In assessing this claim, Steve Calabresi suggested that the cause of the problem was not the unattainability of truth but the desire of legal academics to use their scholarship to advance normative objectives. See Calabresi, supra note 57, at 266 ("[T]he real explanation for the loss in prestige of constitutional scholarship is that it has become a heavily normative body of advocacy scholarship targeted at the federal courts . . . .").


60 During the late nineteenth century, Christopher Langdell and others sought to make law a science through the categorization of decisional law. See C.C. Langdell, A Selection of Cases on the Law of Contracts vi (1871) (“Law, considered as a science, consists of certain principles or doctrines.”). This effort, of course, proved unsuccessful. See Thomas C. Grey, Langdell’s Orthodoxy, 45 U. PITT. L. REV. 1, 45-46 (1983) (describing the practical problems judges and lawyers encountered with Langdell’s method).

Thanks to the adversary system, moreover, partisanship figures prominently in that translation. Legal academics then have good reason to see themselves as being in the business of making arguments. That is what they were trained to do, and that is what they teach their students to do. That their scholarship also would be argumentative comes as no surprise. The surprise, instead, is that the academic ethic is sufficiently strong to provide them with cover for such endeavors.

Whether or not traditional notions of truth exist, academics should still understand that pretending to have an expert opinion on something they know next to nothing about is a deception. Postmodernism helps explain but does not excuse this deceit. Another source of this deceit, ironically, is the special place of academics in our constitutional order. Linked to the academics' reputation as truth seekers, academic freedom empowers academics to speak out on public issues without sanction. For this reason, academics sometimes see themselves as supercitizens, entitled to speak out on issues by virtue of their status.

Membership in the academy, however, has its responsibilities as well as its rewards. Advocacy for advocacy's sake, while certainly enti-

---

82 Anthony Kronman, in an effort to combat the confluence of scholarship and advocacy, argued that "law teachers have a moral responsibility" to pursue truth in their scholarship. Anthony T. Kronman, Foreword: Legal Scholarship and Moral Education, 90 YALE L.J. 955, 967 (1981). Otherwise, their students will see all law as advocacy "which in turn encourages a cynical carelessness about the truth, thus undermining the important good of community." Id. For this very reason, I disagree with Cass Sunstein's attempt to portray all legal academics as potential experts on any (law-related) subject. Instead, speaking (as a "scholar" or "professor") on a politically hot issue requires particular expertise beyond being a lawyer or law professor. See infra notes 95-96 and accompanying text (questioning Sunstein's definition of academic expertise).

83 One of the responsibilities of the academic, according to Edward Shils, is "the obligation not to betray the trust which is given to him when laymen look to him for objective knowledge and responsible judgment." Shils, supra note 17, at 106. This traditional view is echoed by the American Association of University Professors (AAUP), which argues that academics "should remember that the public may judge his profession and his institution by his utterances. Hence he should at all times be accurate, should exercise appropriate restraint, [and] should show respect for the opinion of others." Academic Freedom and Tenure: A Handbook of the American Association of University Professors 132 (Louis Joughin ed., 1969) (quoting American Ass'n of Univ. Professors, 1940 Statement of Principles on Academic Freedom and Tenure, available in American Ass'n of Univ. Professors, 1940 Statement of Principles on Academic Freedom and Tenure with 1970 Interpretive Comments (visited Oct. 26, 1999) <http://www.aaup.org/1940stat.htm>). Along the same lines, the AAUP Statement on Professors and Political Activity specifies that faculty members "should be free to engage in political activities so far as they are able to do so consistently with their obligations as teachers and scholars." American Ass'n of Univ. Professors, Policy
tled to constitutional protection, is not entitled to the special protections of academic freedom. More precisely, the Supreme Court’s willingness to treat academic free speech as more important than other free speech claims is linked to an academic’s fiduciary duty to maintain "standard[s] of professional integrity." Put another way, academic freedom is a quid pro quo. On the one hand, it protects academics from outside political pressures. On the other hand, it is a “contingent privilege” justified by an academic’s willingness to be held accountable at a “professional level for the ethical integrity of his work.”

But do the responsibilities of academic freedom attach to joint letters? After all, no one expects that each and every signatory has played a hand in the letter’s drafting. For similar reasons, it is to be expected that many signatories agree with the conclusions but not the reasoning of the letters they sign. Moreover, with the academy’s glitterati spearheading these letter-writing campaigns, it is to be expected that some signatories (who care that the letter’s reasoning be well thought out but know nothing about impeachment, gun control, or whatever) sign on because they assume that these leading lights would not lead them astray. Finally, some signatories consider the letter’s reasoning beside the point. Their signature, instead, is about partisanship and nothing else. Being able to explain why academics (who cannot defend the reasoning of these letters) sign these missives does not justify this practice. Rather, these letters go out of their way to


Van Alstyne, supra note 84, at 76. In his response, Sunstein claims that I do not discuss what academic freedom is for. I respectfully disagree. See supra notes 15-18, 82-83 and accompanying text (arguing that academic freedom encourages academics to think independently and to seek truth).

85 See supra note 45 and accompanying text (illustrating that signatories do not necessarily agree with the texts of the letters they sign).

86 See supra note 92 (noting that it is easier to sign a letter written by someone you admire).
make clear that they are sending a professional, not a political, message. Writing as "scholars,"88 "historians,"89 "law professors,"90 and "teachers of constitutional law,"91 these letters tout the self-described academic expertise of their signatories. While it is to be expected that the academics signing these letters support the outcomes they advocate, it is not to be expected that many of them cannot defend (and may well not support) the letters' reasoning. Indeed, it is the reasoning of academics—not the conclusions they reach—which justifies academic freedom. It is therefore a perversion of academic freedom to treat professional expressions of expert opinion as nothing more than a plebiscite of personal preferences.

Widespread abuses of academic freedom, unfortunately, now seem to be a fixture of the modern academy. Most tellingly, academics are likely to do a poor job of checking their own excesses. Peer review—the mechanism by which the academy polices itself—requires an openness to different ways of thinking. But with more than three-fourths of the legal academy "characteriz[ing] themselves as 'moderately' or 'strongly' liberal or left,"92 there are strong incentives to agree with prevailing norms. In part, the hegemony within the academy ensures that like-thinking individuals will validate the arguments of other like-thinking individuals (no matter how sound or silly they may be).93 This is especially true among academics who think that power, not truth, holds the key to governmental reform. For these individuals, what matters most is that the right result is reached. In other words, rather than encourage counterspeech, the academy and, with it, peer

88 Scholars' Statement, supra note 11, at 1712.
89 Hearings, supra note 6, at 334-39.
90 Id. at 374-83.
91 Sunstein, supra note 48.
92 Merritt, supra note 35, at 780 n.54.
93 For Pierre Schlag, the sameness of viewpoints and methodologies among legal academics explains why law professors all agree that flag burning laws are unconstitutional. But outside the legal academy, say, before an American Legion in Des Moines, the "deployment of the scam will probably not work very well." PIERRE SCHLAG, THE ENCHANTMENT OF REASON 35 (1998). Within the academy, moreover, there is a real risk of opprobrium for those who do not toe the company line. For example, after publishing articles that questioned the efficacy of critical race scholarship and the purported arrival of the "Asian American Moment" in legal scholarship, the Harvard and Iowa law reviews published symposia filled with condemnatory essays. See generally Colloquy, Responses to Randall Kennedy's Racial Critiques of Legal Academia, 103 HARV. L. REV. 1844 (1990); Colloquy, 81 IOWA L. REV. 1467 (1996). I must confess that, as I write these words, I feel the pressure of nonconformity bearing down on me. By taking to task a significant chunk of the legal academy for their behavior in advancing (within the academy) politically popular causes, I too may find myself in a hornets' nest.
review may well impose "sharp limits on the range of respectable opinion within its ranks."94

* *

What does it mean to speak as an academic expert? Is it enough, as defenders of the anti-impeachment letter argue, that an academic "believed that they knew enough" by speaking with other academics who "probably believed that they knew enough . . . ?"95 If this is true, every law professor can speak as an expert on any issue. Take the recent Microsoft antitrust case. Based on my watching of television news, my conversations with colleagues, and my fuzzy (almost two decades old) recollection of an antitrust class, I might believe that I know enough to develop "genuine, if fairly general, views"96 on the legality of Microsoft's practices. Ditto the decisions of the Securities and Exchange Commission, the Internal Revenue Service, the Environmental Protection Agency, and the National Labor Relations Board. According to this view, by reading the newspaper and hanging out in the faculty lounge, I can hold myself out as an expert on all these things.

This recalibration of what it means to be an academic expert comes at a price. Over time, the academic ethic may give way to the view that self-interest and partisanship are the coins of the academic realm. Conservative critics of the academy have worked hard at portraying it "as a taxpayer-financed bunker inhabited by an army of Birkenstock-shod Marxists."97 And these critics have achieved more than a modicum of success. A recent crop of books (including Illiberal Education,98 The Closing of the American Mind,99 Impostors in the Temple,100

94 GEORGE J. STIGLER, THE INTELLECTUAL AND THE MARKETPLACE 68 (enlarged ed. 1984). For this reason, Stigler argues that academic freedom must look "inward," so that the academy is protected from its own corrupting influences as well as those of outsiders. See also Arthur O. Lovejoy, Academic Freedom, in 1 ENCYCLOPEDIA OF THE SOCIAL SCIENCES 384 (Edwin R. A. Seligman ed., 1930) (stating that academic freedom is "rendered impossible if the work of the investigator is shackled by the requirement that his conclusions" conform to prevailing norms); David M. Rabban, Does Academic Freedom Limit Faculty Autonomy?, 66 TEX. L. REV. 1405, 1407-08 (1988) (arguing that peer review should limit faculty autonomy).

95 Sunstein, supra note 27, at text accompanying note 13.

96 Id. at text accompanying notes 12-13.

97 Suarez, supra note 3, at B8.


99 ALAN BLOOM, THE CLOSING OF THE AMERICAN MIND: HOW HIGHER EDUCATION
Telling the Truth,\textsuperscript{101} and ProfScam\textsuperscript{102}) have argued that political correctness, like-minded thinking, and intolerance have corrupted the academic ethic. Columnists David Broder and Nat Hentoff, as well as Judge Richard Posner, condemned the impeachment letter-writing campaign for this very reason.\textsuperscript{103} Whether this criticism will undercut the saliency of future letter-writing campaigns remains to be seen. Nevertheless, the academy has good reason to fear the perception that it is filled with citizen partisans, not citizen scholars. Not only will academic freedom suffer a body blow, professors will be handicapped in their efforts to affect public discourse through their scholarship.\textsuperscript{104}

Aside from becoming irrelevant, the academy runs another risk. Over the past few years, social conservatives have urged donors and university trustees to play more active roles in the life of the academy. Among other things, trustees and donors have been urged to combat political correctness and moral relativism.\textsuperscript{105} And some trustees and donors are listening to this message. At Virginia’s George Mason University, for example, faculty have castigated the school’s board of trustees for trying to impose a conservative political agenda. Specifically, by shifting funds to programs with a conservative reputation and by dissolving a nontraditional educational program, these trustees (all

\textsuperscript{100} MARTIN ANDERSON, IMPOVERISHED THE SOULS OF TODAY’S STUDENTS (1987).

\textsuperscript{101} LYNN V. CHENEY, TELLING THE TRUTH: WHY OUR CULTURE AND OUR COUNTRY HAVE STOPPED MAKING SENSE—AND WHAT WE CAN DO ABOUT IT (1995).


\textsuperscript{103} See Posner, supra note 30, at 240-42 (likening the law professors’ campaign to “a form of herd behavior”); David S. Broder, The Historians’ Complain, WASH. POST, Nov. 1, 1998, at C7 (arguing that some activist academics, including organizers of the historians’ letter, risk looking “ridiculous” by “heedlessly” plunging into political debate); Nat Hentoff, Breeding Contempt for the Law, WASH. POST, Nov. 21, 1998, at A21 (depicting the signers of the historians’ letter as a “herd” who employed “embarrassingly contorted reasoning”).

\textsuperscript{104} For those who see elitist discourse as harmful to the commonwealth, of course, this changing image of the academy will be salutory. See BORK, supra note 52, at 337 (arguing that the public explosion at Bork’s nomination was driven by liberals at issue with a more conservative general public); NAGEL, supra note 51, at 27-43 (arguing that the legal culture is properly concerned with ideas, but the political culture is concerned with the consequences of ideas, as seen within the Bork Supreme Court nomination hearings); SCHLAG, supra note 93, at 35-38 (arguing that legal thinkers tend to be almost entirely center-left democrats and it is not surprising when they all agree).

\textsuperscript{105} See ANDERSON, supra note 100, at 194-206 (arguing that university trustees seldom use their potential power and leadership effectively); NATIONAL ENDOWMENT FOR THE HUMANITIES, TELLING THE TRUTH 49-51 (1992) (arguing that although trustees of universities seldom use their authority to exercise leadership, they should exert more influence on colleges and universities).
but one of whom were appointed by conservative Republican governors) have broken ranks with university administrators. Will other trustees follow George Mason's lead? Perhaps not. But the more political the academy is perceived, the more likely it is that governors will appoint political trustees.

It may be that this is the fate the academy deserves. After all, the partisanship and misdirection of some academic letter-writing campaigns contradicts some of the most basic tenets of academic freedom. More fundamentally, the willingness of so many academics to pawn off fake knowledge suggests that the conditions supporting academic freedom have dissipated. Among other things, it is difficult to square academic freedom with ideological conformity, the advent of postmodernism, the rise of sound bite scholarship, and, especially at law schools, the nexus between celebrity status and partisanship. What is truly amazing here is that the academy is risking so much to accomplish so little. In the cases of Bork and impeachment, for example, the battle lines were drawn before the academics entered the fray. More than anything, the academic participants in these wars were stage props—brought into the drama to demonstrate that politicians take the Constitution seriously.

What then can the academy do to rescue itself? To start, academics ought not to remove themselves from the world of politics. The very reason that academics possess tenure, academic freedom, and the like is so they may speak "truth to power." Academics, however, must be cautious in their utterances. It is not enough, for example, that some of their colleagues might "indulge a principle of charity" that would allow them to express an expert opinion without toiling with research, writing, and the like. The price of academic freedom is that scholars must use reason, thought, and care in defending their positions, whether political or not. Devices that allow academics to register positions without doing the necessary work undermine academic freedom for all. When it comes to letter writing, for example, academics should only sign letters that they could (if asked to) defend in public. Beyond letter writing, academics should embrace both

---

106 See Victoria Benning, Faculty, Board Clash at George Mason, WASH. POST, May 21, 1999, at B1 (reporting that the George Mason University faculty is battling with the university trustees over the trustees' conservative political agenda).
107 Schlesinger, supra note 14, at 28. Moreover, with the Supreme Court signaling its disinclination to reshape society through sweeping judicial edicts, it is sensible that legal academics would pay more attention to elected government reforms.
108 Sunstein, supra note 27, at text accompanying note 24.
109 In his response, Sunstein claims that he "emphatically" agrees with me that
ideological diversity and dialectic reasoning (where each thesis is challenged by a counter-thesis). That way the academy can better live up to its marketplace-of-ideas reputation.

Politically motivated academics must come to grips with a grim reality, that "[i]n a world where there are no rules of scholarship or journalism or evidence, where everything is opinion and all opinions are alike, the market wins." That market, of course, is far more conservative than the academy. For this very reason, politically motivated academics should see academic freedom as a bunker from which to fight battles, not as a relic of times past. But to preserve academic freedom, politically motivated academics must honor it, not abuse it. For their part, academics who steer clear of partisan causes—that is, most academics—must hold their politically active colleagues accountable for abusing academic freedom. Otherwise, they too will pay the price of membership in a once-revered profession increasingly held in disrepute.

"academics should not sign letters that they could not defend publicly." Sunstein, supra note 27, at text accompanying note 32. For Sunstein, however, it is not necessary that they could defend these positions as academics—i.e., defend the substance of the letter with a commanding knowledge of the relevant sources. See supra notes 94-96, 108 and accompanying text (discussing what it means to be an academic expert).

One way of encouraging such diversity, of course, is to hire professors who—because of academic training or ideology—see the world differently from one another. At the least, academics should share their work (in draft) with individuals who may well disagree with them. For this reason, the screening out of likely naysayers from letter writers' distribution lists is inappropriate. See supra notes 42, 49 and accompanying text (suggesting that organizers of the anti-Bork and anti-impeachment letters did not circulate those letters to likely naysayers).


For this very reason, the academics who suffer the greatest harm from purely political letter-writing campaigns are those who only sign letters on subjects on which they are experts (and organizers of mass letters who limit signatories to individuals who have subject matter expertise).
PROFESSORS AND POLITICS

CASS R. SUNSTEIN

In the last few years, a number of law professors have been involved in some highly public issues, not least through the circulation of letters with multiple signatories, expressing a view on some issue of national importance. It is not clear that there is anything like a trend in this direction; certainly teachers of law have participated in what some would consider "partisan politics" for many decades. But Neal Devins is right to raise questions about the legitimacy and consequences of political involvement by academics. It is also interesting to consider the relationship between such involvement and academic freedom; if law professors are not concerned with the pursuit of truth, the case for academic freedom is certainly weakened. The Clinton impeachment provides the immediate motivation for Devins's discussion. I offer a few remarks here on academic involvement in the impeachment debate, with a few references to my own experience, and

---

† Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, Law School and Department of Political Science, University of Chicago. Professor Sunstein testified on the constitutional issues relating to the impeachment of President Clinton, see Testimony Before the United States House of Representatives, Comm. on the Judiciary, Subcomm. on the Constitution (Nov. 9, 1998) (statement of Cass R. Sunstein, Professor of Law), available in LEXIS, News Library, Transcripts File. I am grateful to Martha Nussbaum and Richard Posner for valuable comments on an earlier draft, and to Brooke May for research assistance.

1 See, e.g., Constitutional Scholars' Conference, Joint Statement, Constitutional Scholars' Statement on Affirmative Action After City of Richmond v. J.A. Croson Co., 98 YALE L.J. 1711, 1712 (1989) (arguing that municipalities should not scrap affirmative action programs in the wake of City of Richmond v. J.A. Croson); An Open Letter to Congressman Gingrich, 104 YALE L.J. 1539, 1539 (1995) (urging that Congressman Gingrich rescind his proposal to amend the House Rules to require three-fifths vote for enacting laws that increase income taxes); infra notes 6-7 (providing text of a one-sentence letter opposing impeachment and a letter opposing Clinton's possible resignation).

2 See Neal Devins, Bearing False Witness: The Clinton Impeachment and the Future of Academic Freedom, 148 U. PA. L. REV. 165, 166 (1999) (arguing that "when a significant number of law professors and historians hold themselves out as experts when they are not... all academics pay a price").

3 The editors have asked me to provide a few details for background; readers who dislike this sort of thing are encouraged to skip the footnotes.
with an eye toward drawing some general lessons about professors and politics, including presidential politics.

I. PARTICULARS

With respect to the President, the United States has now had three serious impeachment inquiries in its entire history. As the nation began discussion of the impeachment of President Clinton in 1998, many constitutional law professors concluded that no legitimate grounds for impeachment had been identified. On their view, the House of Representatives had started to embark on a constitutionally impermissible path, one with potentially significant consequences. This seemed unfortunate not merely because of the potential removal of the President, and hardly because of the nature of President Clinton’s particular policies, but because of the constitutional illegitimacy and potentially destabilizing effect of resort to the impeachment mechanism.

Some law professors believed that, as a technical matter, the constitutional question was not even close. They believed that this interpretation was not a “partisan” view, that it had nothing to do with approval of President Clinton in general, and that it was the best reading of the Constitution—a reading that would attract support from a variety of possible approaches to interpretation and that would apply regardless of the political affiliation of the President. Before the hearings, members of the House of Representatives—themselves quite uncertain about the issue—asked, in private meetings with law professors: What do law professors generally think? Where are they? Why

---

4 For a discussion of the Impeachment Clause, see Cass R. Sunstein, *Impeaching the President*, 147 U. PA. L. REV. 279, 280-92 (1998). Devins is wrong to say that “scholars who argued against the appropriateness of originalism at the Bork hearings made use of originalism (and little else) in arguing against the Clinton impeachment.” Devins, *supra* note 2, at note 24. Professor Tribe’s argument, for example, was largely textual, see *Defining “High Crimes and Misdemeanors”: Basic Principles* (visited Oct. 24, 1999) <http://www.house.gov/judiciary/22398.htm> (advocating that defining impeachment properly “means taking seriously exactly what the Constitution says”), and while I did not argue against the appropriateness of originalism at the Bork hearings, my argument was typical in stressing a range of interpretive theories. SeeSunstein, *supra*.

5 I draw on personal experience here. There was a series of private, informal meetings with several law professors who met, in small groups, with a large number of representatives, many of whom were unsure about how to think about, or vote on, the impeachment question. A vivid recollection: The representatives listened politely to the constitutional arguments, and were clearly interested in them, but some of their eyes really lit up only during discussion, in which law professors were not involved, of the underlying political dynamics.
aren't they saying anything? An obvious question was whether it might conceivably be useful for members of Congress, and the public, to know that there was widespread professional opinion to the effect that impeachment would be unconstitutional.

These were the circumstances in which a number of professors (including me) were willing to circulate and to sign two letters opposing impeachment. One of the letters, not limited to teachers of constitutional law but circulated to law professors generally, provided an extended analysis of the legal issues; the other, limited to teachers of constitutional law, offered a one-sentence conclusion to the effect that impeachment would not be "appropriate" in light of the charges made by Judge Starr.

Many of those who signed or circulated these letters accepted (as I do) all or most of Devins's general views about the need to separate professional opinion from partisan politics. To say the least, most of those of us who were involved in all this really do not enjoy participat-

It is a considerable overstatement for Devins to say that "members of Congress were actively involved" in any academic anti-impeachment letter. Devins, supra note 2, at text accompanying note 66. Certainly the remarks did not lead to the letters. But I agree with Devins about the general importance of a degree of independence on the part of academics involved in these matters. Those of us who agreed with the White House on the inappropriateness of impeachment made very clear—especially in several informal discussions with lawyers there involving the constitutional standard for impeachment—that we were not working "with" or "for" the White House in any way.

Because this letter was limited to teachers of constitutional law, I think Devins is wrong to say that none of the relevant letters "made expertise a prerequisite"—at least if we assume that teachers of constitutional law generally have some expertise on the topic of impeachment. Devins, supra note 2, at text accompanying note 45.

I wrote and circulated the one-sentence letter after many discussions with other law professors. Some people did not like this approach, on the ground that it was without reasoning; others, including me, thought that it would be worthwhile and relevant, if only because it would show a widespread judgment on the general constitutional question without committing people to a particular argument about which they might have reservations. Here is its content:

The undersigned professors and teachers of constitutional law* have diverse political convictions and disagree on many political and legal issues; but we agree that the possible grounds for impeachment recently identified by Kenneth Starr and David Schippers are not an appropriate basis for impeaching a President under Article II, Section 4 of the Constitution.

*All of the signatories to this letter have taught the general constitutional law course or some aspect of constitutional law as part of their curricular responsibilities.

There was also a subsequent letter, circulated after impeachment, opposing resignation on the ground that it would do damage to the constitutional structure. See Bernard J. Hibbitts, Law Professors Solicit Signatures for Anti-Resignation Letter (visited Oct. 27, 1999) <http://jurist.law.pitt.edu/resig.htm> (containing the full text of the letter, dated Dec. 19, 1998).
ing in petition drives. Most of us thought that the process of sending and responding to emails and collecting signatures was a boring, unpleasant, and tedious business—part of a most unwelcome (even if occasionally hilarious\(^9\)) diversion from our academic jobs. Nothing here seemed glamorous, a source of "fame,"\(^10\) or in anyone's self-interest. But the relevant law professors believed that this was an exceedingly unusual event, that the House of Representatives was on the verge of acting unconstitutionally, and that it was appropriate, and maybe not inconceivably useful, for the public and elected representatives to be aware of a widespread (though not universal) professional opinion.

I am not entirely sure what, in particular, Devins thinks was wrong with all this, or what lesson he wants to draw from it.\(^11\) He does not argue that the impeachment of President Clinton was constitutionally acceptable (though he may believe this). He does not demonstrate nor even claim that the law professors who signed one or another letter were wrong on the merits. Nor does he argue for the (implausible) proposition that on principle, law professors should never speak publicly about important public issues.

Devins suggests that developments of this kind may endanger academic freedom, an empirical claim that seems to me quite doubtful, even absurd. Is it imaginable that the tenure system for law professors would be eliminated if law professors frequently signed petitions on public issues? In any case, Devins does not discuss what academic freedom is for, to wit, the power to speak controversially about the

---

\(^8\) A confession: The one-sentence letter was the first such letter I have ever written and circulated, and because the process was so tedious, I hope never to have to do such a thing again.

\(^9\) See infra note 31 (discussing Burden of Proof and television producers).

\(^10\) But see infra note 31 (discussing my dog).

\(^11\) Devins follows the familiar view in stating that "Republican witnesses at the House impeachment hearings testified that the President's conduct was impeachable whereas Democratic witnesses testified that it was not." Devins, supra note 2, at note 67. In a sense, this statement is accurate; but my hunch is that a majority of witnesses, on both sides, did not consider themselves "Republican witnesses" or "Democratic witnesses." What a strange way to conduct a hearing on the impeachment question—as if, before the debate even became serious, all Republicans were for impeachment and all Democrats against it.

Perhaps the greatest peculiarity of the hearings was that a majority of the witnesses said that the allegations in question pointed to legitimate grounds for impeachment, thus giving the impression that teachers of constitutional law were evenly divided on the question—when in fact the overwhelming majority of constitutional law teachers believed that impeachment was constitutionally illegitimate or at least inappropriate. The selection of witnesses gave a grossly distorted picture of academic opinion.
truth as one sees it, free of risks of political reprisal. I agree with Devins that the principle of academic freedom is violated if those who enjoy it are not acting in good faith or are not pursuing truth (not a doubtful empirical claim about consequences but a sensible claim about principle, which is what appears to me to underlie Devins's essay). But nothing in this point argues against public statements by professors.

Devins's concerns appear narrower. It seems to me that they fall in two categories. These categories are related but best treated separately.

His first point has to do with the possibly limited expertise of many or some of those who signed the relevant letters. Devins says that "it is doubtful that many had thought seriously about the constitutional standards governing impeachment."\(^{12}\) He appears to believe that at least by implication, some or many law professors held themselves out as specialists or experts when, in fact, they lacked knowledge about impeachment that would qualify them as such.

For some of the signatories, perhaps this is true. But I think that Devins's judgment is too harsh. Impeachment is hardly an obscure or invisible issue in constitutional law, and in the wake of the Watergate controversy, many law professors developed genuine, if fairly general, views on the appropriate meaning of the phrase "high crimes or misdemeanors." Certainly most teachers of constitutional law know something about the governing legal standards; they know enough to know, for example, that smoking marijuana or speeding would not ordinarily count as a "high crime or misdemeanor." From there they could reason by analogy to the view that, at least as a general rule, a President cannot be impeached unless he has been charged with large-scale abuse of the powers that he has by virtue of being President.

Those law professors who signed the longer letter but who do not teach constitutional law probably believed that they knew enough—from training and from substantive conversations with colleagues—to have a reasonably informed opinion about the threshold question of whether the charges against President Clinton made out an impeachable offense. Law professors who do not teach constitutional law have informed views about many constitutional issues—for example, about whether racial segregation is generally unconstitutional, whether quotas can make for an acceptable affirmative action program, and

---

\(^{12}\) *Id.* at text accompanying note 21.
whether the Constitution protects the right to use contraceptives. Many law professors believe that with respect to the charges against President Clinton, impeachment falls in the same category. The signatories likely thought, in good faith, that they knew enough about the constitutional provision to conclude that an impeachable offense had not been made out. It is hard to see why there is anything untoward here. I agree with Devins that people should not sign petitions when they are unable to defend the relevant position publicly; but I would give the signatories the benefit of the doubt on this point.

Devins's second point has to do with the motivations of those who participated in the relevant events. He says that "[m]any of the law professor and historian signatories were animated by partisanship and self-interest . . . ." This is certainly possible—and I know very little about the historians involved—but it is very hard for Devins or anyone else to have access to the motivations of strangers. Nor is it clear what he means by "self-interest" and "partisanship." Academics "who want to see their names in print" surely have much better ways to achieve that goal; it is hard to believe that publication was a significant motivating factor here. And contrary to Devins's suggestion, "celebrity" hardly comes from signing such a letter. A few law professors may care about celebrity, but most law professors care about the law and about ideas.

The allegation of "partisanship" raises many questions. Law professors who signed this letter were offering a judgment about the meaning of the Constitution, a judgment that they believed it to be true. In what sense was their motivation "partisan"? A law professor who testifies that a proposed statute would violate the Constitution

13 Judge Posner seems to me unfair in criticizing law professors for not publicly deploring President Clinton's conduct, which, plausibly, included perjury and obstruction of justice. See Richard Posner, An Affair of State 240-45 (1999) (lamenting that "leaders of the legal profession might have been expected to emphasize the importance of the rule of law in general and of telling the truth . . . These expectations would have been disappointed"). I know that many of those who opposed impeachment on constitutional grounds believed that the relevant conduct was deplorable, and certainly many of us emphasized, publicly and privately, the seriousness of perjury and obstruction of justice. In any case, nearly everyone, including President Clinton's closest friends and allies, as well as members of both parties, publicly deplored his conduct, and an outpouring of statements to this effect by law professors would have been largely pointless.

14 Devins, supra note 2, at text accompanying note 31.
15 Id. at text accompanying note 32.
16 Id.
17 Id. at text accompanying note 31.
18 Id. at text accompanying note 32.
should be taken to be addressing the constitutional issue and need not have a political motivation; what is different about a law professor who signs a public statement about what the Constitution means? Perhaps Devins's claim is that the professors who signed this letter were motivated not by a considered judgment about the Constitution, but by their extra-legal political convictions (in favor, for example, of government regulation and the welfare state, or in favor of President Clinton generally). But I do not know how Devins can be confident of this uncharitable judgment, and in any case, the anti-impeachment conclusion would hold if President Reagan, or a future Republican President, were subject to an impeachment inquiry on the basis of similar allegations.19 I very much doubt that the signatories would change their view if the accused President had been Republican. Perhaps Devins disagrees, but he offers little basis for any such disagreement.

Devins seems to think that partisanship is established by his claim that the academy is "overwhelmingly left-liberal, overwhelmingly Democratic."20 But he does not note that many people not normally characterized as "left-liberal" signed one or more of the impeachment letters.21 For what it is worth, the University of Chicago Law School is

19 A possible point in support of Devins: If a group of law professors despised a certain President, they might not go to the effort to circulate and sign a letter against impeachment, even if they believed that the grounds for impeachment were invalid. Perhaps some law professors did not sign the Clinton letters even though they agreed with them, simply because they did not like President Clinton; perhaps some law professors would not sign a similar letter on behalf of, for example, President Reagan, even if they agreed with it.

20 Devins, supra note 2, at text accompanying note 33.

21 These include Douglas Laycock, Stephen Macedo, and Andrzej Rapaczynski. I am sure that there are many more and am happy to say that I do not know the political views of the vast majority of signatories. Of course Devins is allowed to characterize people as he chooses, but for what it is worth, I would not describe myself as a "left-liberal," see Devins, supra note 2, at text accompanying notes 33-39, or indeed as a Democrat, see id. at note 36 and accompanying text.

Devins is misleading to suggest that Judge Posner described the "law professor anti-impeachment letter as 'a form of herd behavior...by the animal that likes to see its name in print.'" Devins, supra note 2, at text accompanying note 54 (quoting RICHARD A. POSNER, AN AFFAIR OF STATE 242 (1999)). Actually Judge Posner describes this as the possible opinion of an "unkind critic," thus distancing himself from the charge, which is not his own view. RICHARD A. POSNER, AN AFFAIR OF STATE 242 (1999). (I owe this correction to Judge Posner.)

I also doubt that "many legal academics see Ken Starr...as their nemesis." Devins, supra note 2, at text accompanying notes 36-37. I do know that many of those (including me) who opposed Clinton's impeachment believe that Starr was a distinguished judge and a distinguished Solicitor General and have not the slightest doubt that he is an entirely honorable person. A number of us who strongly opposed im-
not usually characterized as "left-liberal," and while I cannot speak officially or on behalf of the faculty, it seemed to me that at least a solid majority of the faculty, including many strong critics of President Clinton's policies, were opposed to impeachment on the ground that it would be illegitimate or inappropriate from the constitutional point of view. Indeed, my former colleague Michael McConnell, hardly a left-liberal, went so far as to write a letter to Congressman Hyde, sharply critical of President Clinton and suggesting that impeachment would not be unconstitutional—but pleading with Hyde not to support impeachment of the President in the midst of such partisan divisions. Was McConnell's letter inappropriate? Unacceptably partisan? Illegitimate because (unlike the two widely distributed letters) it did not directly address a purely legal question?

The impeachment of President Clinton produced, on all sides, more than enough challenges to other people's motivations. I am not sure why Devins does not, with respect to law professors and historians on all sides, indulge a principle of charity—suggesting that in cases of doubt, we should assume that those with whom we disagree are acting in good faith and on the basis of evidence that they honestly believe to be sufficient.

---

See Letter from Michael W. McConnell, Presidential Professor, The University of Utah College of Law, to Henry J. Hyde, Jr., Chairman, Committee on the Judiciary Committee of the House of Representatives 1-2 (Dec. 12, 1998) (on file with the University of Pennsylvania Law Review) (indicating that while "the Constitution allows for the possibility that an elected President may so abuse the public trust that he can and should be removed," still "a President should [not] be impeached or removed from office solely on the votes of his political rivals").

This seems to me to infect RICHARD A. POSNER, AN AFFAIR OF STATE (1999) as well, where analysis of the constitutional issue seems to me remarkably casual and off-hand. See id. at 94-109 (discussing the history and scope of impeachment through a "structural and pragmatic... approach to constitutional meaning").

I share many of Devins's concerns about the political campaign against Judge Robert Bork. But I do not think that the law professors who opposed Bork's confirmation did so because they "saw Bork as a threat to their status and influence." Devins, supra note 2, at text accompanying notes 51-53. I think that they did so on principle, that is, because they believed that Judge Bork's view about constitutional interpretation would be a threat to constitutional ideals, properly understood. They may have been wrong on the merits, and I agree with Devins that petitions of the sort circulated against Judge Bork should be rare. (In the interest of full disclosure, I should say that I testified quite critically about Judge Bork's views on separation of powers questions, but signed no petitions and took no public position on the question of whether he should be confirmed.)
II. Generalities

There are much broader questions in the background. Devins thinks that academic participation in the Clinton impeachment is part of a larger trend and a more general problem, one that threatens academic freedom itself. Thus he writes that “[t]oday, academics seek fame through talk show appearances, op-ed pieces, and trade press books.”²⁵ Devins also says that “the traditional image of the academic has given way to ‘postmodernism, multiculturalism, and political correctness,’”²⁶ and hence that “there is good reason to doubt whether academics still think of themselves as truth seekers.”²⁷ His suggestion appears to be that the involvement of law professors in the Clinton impeachment signals a larger trend toward highly partisan, mostly left-wing activity, in which law professors do not care about truth but instead push some kind of political agenda. Devins suggests that if professors are engaging in partisan politics, academic freedom is at risk. I share many of his concerns, and I agree that the principle of academic freedom is undermined if professors do not care about truth. But there seem to me to be three problems with his claims here.

The first problem is that there is no single way to be a law professor. Most of us, most of the time, are devoted to teaching and research; few of us “seek fame” or anything else through talk shows, op-ed pieces, and trade press books. But some professors sometimes do express their views on talk shows, and a few more write op-eds and publish with trade presses. Why is this so terrible, or indeed terrible at all? (Can anyone really object that Gerald Gunther published his excellent book on Learned Hand²⁸ with Knopf rather than Stanford University Press?) One of the purposes of academic freedom is to permit professors to speak publicly without fear of reprisal, and those who write op-eds or publish with a trade press are doing what academic freedom is designed to permit them to do. My hunch is that when law professors write op-eds or go on television, it is to express the truth as they see it, not to “achieve their fifteen minutes of fame.”²⁹ Some people spend all of their time on teaching and academic projects; others testify before Congress and write an occasional amicus brief; still others write op-eds and say what they think on radio or tele-

²⁵ Id. at text accompanying note 55.
²⁶ Id. at text accompanying note 75 (quoting David M. Rabban, Can Academic Freedom Survive Postmodernism?, 86 Cal. L. Rev. 1377, 1378 (1998)).
²⁷ Id.
²⁹ Devins, supra note 2, at text accompanying note 55.
vision shows. Of course some of us may not think so well of colleagues who spend a lot of time on talk shows, especially in light of the (ridiculous, hilarious) restrictions of the format, and above all if this comes at the expense of their scholarship and teaching. But since there are many ways to be a good law professor, and so long as teaching and research are not cheapened or neglected, I do not quite see Devins’s objection here.

The second problem with Devins’s claims is that they overlook the fact that it is perfectly responsible, maybe even a civic duty, for law professors to participate in public affairs, at least some of the time, by showing how what they know bears on public issues. (I do not believe that Devins disagrees with this point.) If, for example, a teacher of constitutional law thinks that a proposed law is unconstitutional, there should be no taboo on her saying so, even through an op-ed piece. If a teacher of criminal law believes that laws forbidding gang loitering for example: In one interview on CNN, I was required to try to stand on a stack of extremely unsteady boxes, not as a scientific experiment or medical test or cruelty of some kind, but so as to ensure a view of the Capitol building on the television screen. Most of the time, it seemed likely that I would fall off. (This interview was live.) And apparently all law professors on television are, by some firmly enforced unwritten rule, required to try to hold a small piece of equipment in one ear, so as to be able to communicate with other people on the show, including the host. (It isn’t easy to hold things in your ear.) Most of the time, the small piece of equipment does not fall out; but some of the time, it does. (This is especially troublesome when you are trying to balance yourself on a stack of extremely unsteady boxes.) See Burden of Proof (CNN television broadcast, Nov. 9, 1998), available in LEXIS transcript no. 98110900V12 (discussing the meaning of “other High Crimes and Misdemeanors,” perjury, and Bills of Attainder).

50 For example: In one interview on CNN, I was required to try to stand on a stack of extremely unsteady boxes, not as a scientific experiment or medical test or cruelty of some kind, but so as to ensure a view of the Capitol building on the television screen. Most of the time, it seemed likely that I would fall off. (This interview was live.) And apparently all law professors on television are, by some firmly enforced unwritten rule, required to try to hold a small piece of equipment in one ear, so as to be able to communicate with other people on the show, including the host. (It isn’t easy to hold things in your ear.) Most of the time, the small piece of equipment does not fall out; but some of the time, it does. (This is especially troublesome when you are trying to balance yourself on a stack of extremely unsteady boxes.) See Burden of Proof (CNN television broadcast, Nov. 9, 1998), available in LEXIS transcript no. 98110900V12 (discussing the meaning of “other High Crimes and Misdemeanors,” perjury, and Bills of Attainder).

51 A personal anecdote: I appeared on several shows to discuss impeachment issues, but the format seemed so confining, and so conducive to ludicrous oversimplification, and driving there and back seemed so time consuming, that after a while I decided not to do it anymore. I told a producer of CNN’s Burden of Proof that I would no longer discuss impeachment, and would appear on the program if and only if my new dog Perry, a (photogenic) Rhodesian Ridgeback puppy, could appear with me. They called my bluff. The show aired on December 31, 1998, and dealt with legal issues relating to dogs. My dog did become at least somewhat famous; in the middle of the show, the producer broke in to tell me that CNN was being “flooded” with calls and compliments about Perry. Interested readers (can there possibly be any?), those who think I’m joking, and Rhodesian Ridgeback enthusiasts can find the transcript at CNN Burden of Proof (CNN television broadcast, Dec. 31, 1998), available in LEXIS, transcript no. 98123100V12. [Ed. Note: Perry’s photo is available at <http://student-www.uchicago.edu/orgs/law-phoenix/2bear.html>].

A small media notation: The producers who invite people to appear on their shows seem to have the same style and character as that of the on-air hosts of those shows (is this an antecedent job qualification, or does it develop over time?). Thus the most hilarious invitation came from the producer of the show Hardball, whose invitation appeared in exactly the same staccato voice as that of Chris Matthews, the host of the show: We-want-you-on-Hardball-know-about-it?
will or will not deter crime, and has reasons for the belief, it is important for the public to know about this. This is an ordinary part of civic discourse, and a nation is better off with it than without it. Indeed, this kind of academic engagement is one of the things that academic freedom is designed to allow.

Of course silence about public issues can be legitimate too. Whether or not professors should speak publicly depends at least partly on the justice or injustice, and on the legality or illegality, of the practices at issue. Since silence is a choice and not a neutral position, professors, like other citizens, are properly criticized for failing to speak out on certain questions, at least if they have relevant expertise. Teachers of constitutional law would have been properly criticized, I believe, if they had remained silent during the unconstitutional impeachment of President Clinton.

If it is legitimate for one professor to speak publicly, it is not illegitimate for a group of them to do so, at least on rare occasions. Surely it is relevant that law professors are by their very profession concerned with highly public issues. They are unlike teachers of French literature, for their daily life is likely to breed expertise on issues that will, at one point or another, have public salience. I emphatically agree with Devins’s concluding suggestion—that academics should not sign letters that they could not defend publicly. But this is a test that could, I believe, be easily passed by those who signed the anti-impeachment letters, and indeed I doubt that this test would be flunked by many academics with respect to letters they have signed in the past.

The final problem is that it is unclear what Devins means to claim with his invocation of “postmodernism, multiculturalism, and political correctness.” Very few of those involved in the Clinton impeachment letters have any sympathy with postmodernism, a movement that (in my view fortunately) has had only a sporadic influence on the academic study of law—far less of an impact than, for example, that of the economic analysis of law. “Multiculturalism” can be understood in many different ways, and certainly no form of “multiculturalism” now dominates law schools. “Political correctness,” if understood as a social taboo on the expression of unpopular opinions, is certainly to be deplored. But so understood, it is hardly limited to left-leaning aca-

---

32 Devins, supra note 2, at text accompanying note 109 (advocating this test as a means of ensuring that academics do the necessary work before registering opinions).
33 Id. at text accompanying note 75 (internal quotations omitted).
34 A qualification: Some unpopular opinions are properly subject to taboo; con-
demics, and here too, I am not sure how the complaint about political correctness bears on the points at hand. Instead of complaining about what he appears to see as left-wing domination of law schools, Devins might defend a substantive position that he believes is unjustly devalued and that deserves a better hearing.

I agree with Devins’s concerns about skepticism with respect to truth, as do many others; but as a class, law professors are hardly skeptical of the idea of truth, and I do not see why Devins believes that there is, within the law schools, much “discomfort with truth-seeking.” Law professors founded their overwhelming opposition to the Clinton impeachment on a commitment to truth, not on skepticism about truth. Nor is Devins entirely clear about the distinction between “truth” and “political motivation.” Suppose that someone really believes that a proposed law is unconstitutional. Suppose someone believes that some proposition about the Constitution is true. Is there any problem with writing, or signing, a letter to that effect? What makes that action inconsistent with academic freedom?

It is time to conclude. Academics have a duty to pursue truth, and for most academics, most of the time, participation in partisan causes can be destructive to that endeavor. But legal academics sometimes know things that are relevant to public life, and when this is so, they are entitled, and sometimes obligated, to say what they know. They may be wrong on the merits, but that is another matter.

---


36 Devins, supra note 2, at text accompanying note 80. Of course there are exceptions, discussed and addressed in FARBER & SHERRY, supra note 35. (Of course I do not mean to endorse, by these brief comments, everything in Farber and Sherry’s controversial book.)
COMMENTS

BIG LEAGUE PERESTROIKA? THE IMPLICATIONS OF FRASER V. MAJOR LEAGUE SOCCER

EDWARD MATHIAS†

INTRODUCTION

The filing of an antitrust suit by Major League Soccer ("MLS") players against MLS¹ was viewed as a rite of passage for the new league.² After all, every established major professional sports league has been sued for alleged violations of the antitrust laws for practices relating to league rules concerning everything from franchise relocation to the wages of practice squad players.³ The importance of Fraser v. MLS for the future of professional sports leagues, however, transcends the continuing legality of the MLS regulations challenged in the suit. Fraser is momentous because it is the first antitrust challenge

† B.A. 1992, Amherst College; M.S. 1997, University of Massachusetts; J.D. Candidate 2000, University of Pennsylvania. I would like to express my appreciation to Professors Glenn Wong, Lisa Pike Masteralexis, and Edward Rock. I would also like to thank Kirstin Thorne, Bill and Jane Mathias, and Thomas Rapisarda for their support.

¹ The court adjudicating the lawsuit has already issued one ruling on pre-trial motions made by the litigants. The motions concerned the legality of MLS's transfer fee arrangement with Federacion Internacional de Football Asociacion ("FIFA"), the sport's international governing body. The arrangement requires any soccer league seeking to procure the services of a MLS player to pay MLS a fee, even if the player's MLS contract has expired. See Fraser v. Major League Soccer, L.L.C., 7 F. Supp. 2d 73, 79 (D. Mass. 1998) (denying parties' cross-motions for summary judgment regarding MLS's transfer fee rule). The opinion did not address the single entity issue, the pivotal legal issue in the pending litigation and the focus of this Comment.


to a "single entity league," a league that is organized as a single corpo-
ration rather than as a group of individually owned teams.  

MLS's single entity structure was designed to insulate the league
from one form of antitrust liability under the Sherman Act. The
league is structured as a single corporation, which wholly owns all of
the teams that compete in the league. "The Sherman Act contains a
'basic distinction between concerted and independent action." Most
antitrust challenges to the established leagues have been brought un-
der section 1 of the Sherman Act ("section 1"), which only applies to
c concerted action between two economic actors: "It does not reach
conduct that is 'wholly unilateral.'" The actions of a single corpo-
ration are only regulated by section 2 of the Sherman Act ("section 2"),
which prohibits monopolization or attempted monopolization of
trade. It is MLS's legal position that, as a single corporation, it can-
not "combine, contract, or conspire" with itself, and therefore its in-
ternal league practices are not actionable under section 1. Should the
court(s) accept MLS's argument, single entity leagues will have a sig-
nificant advantage in their labor relations relative to other leagues.
This advantage will encourage newly forming leagues to follow MLS's
example and organize as single entities. Perhaps even more signifi-
cantly, an MLS legal victory may induce more established leagues to
reorganize themselves as single entities.

Aside from facilitating the formation of new professional sports
leagues, there are no compelling policy reasons for treating MLS as a
single entity. Although it may be easy to sympathize with the players'
desire to earn what might seem to be their true market value, the
players as individual economic entities have little market value. It is
only in the context of a competitive league, provided by MLS investors
risking millions of dollars in losses, that the players' skills become

---

133 (discussing "MLS's trailblazing single-entity structure").
(discussing the reasons MLS structured itself as a single corporation).
Monsanto Co. v. Spray-Rite Serv. Corp., 465 U.S. 752, 761 (1984)).
7 Id. at 768 (quoting Albrecht v. Herald Co., 390 U.S. 145, 149 (1968)).
shall monopolize, or attempt to monopolize, or combine or conspire with any other
person or persons, to monopolize any part of the trade or commerce among the sev-
eral States... shall be deemed guilty of a felony...").
9 The players would not be able to bring a section 1 claim against the league, thus
removing an important bargaining chip. See discussion infra Part III (discussing the
effect of a dismissal of a section 1 claim in light of the low likelihood of success on a
section 2 claim).
valuable. Furthermore, although MLS labor restraints may hold down salaries of players already in the league, a profitable league is likely to expand and thus provide employment opportunities for more players.

Despite the lack of policy concerns, the players’ suit is significant because it will force the courts to consider how economic coadventurers who retain some minimally disparate economic interests should be treated under the antitrust laws. The courts’ response to this question may have a considerable effect on both traditionally organized leagues and non-sports joint ventures that require cooperation among economic competitors.

This Comment explores the implications of a victory for MLS in its current litigation with its players. Part I examines the reasons a professional sports league should be concerned with antitrust law, and the history of the single entity question as it relates to the more “traditional” league model.\(^\text{10}\) Part II compares the various single entity models in existence and analyzes the potential arguments of the *Fraser* litigants. It concludes, on the strength of its legal arguments, that MLS should be considered a single entity for section 1 purposes, thus rendering intraleague rules immune from section 1 scrutiny. Part III discusses the viability of a claim under section 2 of the Sherman Act against a single entity league, which may limit the extent to which MLS and other single entity leagues are shielded from antitrust scrutiny. Part IV explores the possibility of established, traditionally organized leagues restructuring themselves as single entities and suggests one possible plan to accomplish such a reorganization.

I. ANTITRUST LAW, PROFESSIONAL SPORTS LEAGUES, AND THE SINGLE ENTITY QUESTION

A. Antitrust Law

Antitrust law regulates the conduct of economic actors. Defendants who violate the Sherman Act are liable for treble damages.\(^\text{11}\) Therefore, Sherman Act violators may incur enormous liability for their anticompetitive behavior. Section 1 of the Sherman Act prohibits “[e]very contract, combination . . . or conspiracy, in the restraint of

\(^{10}\) The National Football League, National Basketball Association, National Hockey League, and Major League Baseball are all organized similarly. See discussion infra Part I.C (discussing the traditional professional sports league model).

The actions of a single economic entity are not subject to section 1 scrutiny because section 1 requires an agreement between at least two independent economic actors, commonly called "concerted action," to satisfy the statute's "contract, combination...or conspiracy" requirement. Concerted action is assessed under section 1 under two "complementary categories of antitrust analysis." Those restraints of trade that have no competitive benefits are declared illegal per se. Such restraints include price fixing, group boycotts, and horizontal market division. More commonly, restraints are assessed under the rule of reason. Restraints subject to the rule of reason violate section 1 only if the anticompetitive effects of a particular agreement outweigh its procompetitive effects. The competitive effects of these agreements "can only be evaluated by analyzing the facts peculiar to the business, the history of the restraint, and the reasons why it was imposed."

Section 2 of the Sherman Act applies to single firm conduct as well as concerted action. A firm violates section 2 only when it holds monopoly power in a particular market and engages in behavior that constitutes an abuse of that monopoly power. The requirement that a section 2 defendant possess monopoly power makes a section 2 claim much more difficult to pursue than an action under section 1, which has no such requirement.

B. Why Are Professional Sports Leagues Concerned with Antitrust Law?

The Supreme Court noted in National Collegiate Athletic Ass'n v. Board of Regents that sports leagues and teams require a high level of cooperation among competitors for their product to even exist. Uniform rules of play, restrictions on scheduling, and myriad other regulations are necessary for college and professional teams to compete on a reasonably level playing field. When sports teams within a league

---

15 See Arizona v. Maricopa County Med. Soc'y, 457 U.S. 332, 344 (1982) (deeming per se treatment appropriate "[o]nce experience with a particular kind of restraint enables the Court to predict with confidence that the rule of reason will condemn it").
16 National Soc'y of Prof'l Eng'rs, 435 U.S. at 692.
17 468 U.S. 85, 101 (1984) ("When a [hypothetical] league of professional lacrosse teams is formed, it would be pointless to declare their cooperation illegal on the ground that there are no other professional lacrosse teams." (quoting ROBERT H. BORK, THE ANTITRUST PARADOX 278 (1978))).
agree to certain restrictions, however, they necessarily exclude other potential competitors, suppliers, and distributors of their product.\textsuperscript{18} In other words, the restrictions are agreements in restraint of trade, and, at least facially, violate section 1 of the Sherman Act.\textsuperscript{19}

Just as agreements regarding the number of players who may participate on the field of play are necessary to guarantee a level of competitive balance on the field, leagues have attempted to ensure competitive balance by implementing a variety of restrictions that also interfere with players' ability to market their services. The player draft, free agency restrictions, salary caps, and revenue sharing agreements all restrict, to varying degrees, the wages players may earn. The league practices at issue in \textit{Fraser} are in many ways typical of the type of restraints challenged by players under antitrust law. The standard MLS player contract gives the league the unilateral right to renew the contract, rather than allowing the player to sell his services to the highest bidder.\textsuperscript{20} MLS maintains a salary cap that sets the maximum amount any team may spend on player salaries. This cap is somewhat redundant, however, as a single league official is responsible for negotiating all player contracts and thus has complete control over each team's total salary.\textsuperscript{21} MLS also complies with the transfer fee system created by Federacion Internationale de Football Associacion ("FIFA"), soccer's world governing body, which requires other leagues to pay MLS for the rights to a player, even after the player's contract has expired.\textsuperscript{22}

Players' unions regard the threat of an antitrust suit as a significant bargaining tool in negotiations with the leagues.\textsuperscript{23} However, a re-

\textsuperscript{18} For example, limiting the number of games that each team plays in a season limits the supply of games that can be broadcast to the public by television networks. Allowing teams to draft particular players limits those players' ability to market themselves in the labor supply market.

\textsuperscript{19} The Sherman Act, however, has been interpreted to proscribe only \textit{unreasonable} restraints of trade. \textit{See} Standard Oil Co. v. United States, 221 U.S. 1, 60 (1911) ("[I]t was intended that the standard of reason... be the measure used for the purpose of determining whether in a given case a particular act had or had not brought about the wrong against which the statute provided."); \textit{see also} State Oil Co. v. Khan, 522 U.S. 3, 10 (1997) ("Although the Sherman Act, by its terms, prohibits every agreement 'in restraint of trade,' this Court has long recognized that Congress intended to outlaw only unreasonable restraints." (citation omitted)).


\textsuperscript{21} \textit{See id.} (discussing MLS's salary policies).

\textsuperscript{22} \textit{See id.} (discussing the FIFA transfer fee system).

\textsuperscript{23} \textit{See} WEILER & ROBERTS, \textit{supra} note 5, at 204 (noting that other players' unions' inability to achieve their goals through the collective bargaining process led them to pursue, in the alternative, antitrust challenges against their respective leagues).
recent Supreme Court decision has limited the ability of players represented by unions to sue their respective leagues for antitrust violations. In *Brown v. Pro Football, Inc.*, the Court held that agreements between a multi-employer bargaining unit and a labor union are exempt from antitrust challenge until the "agreement among employers [is] sufficiently distant in time and in circumstances from the collective bargaining process." It is unclear exactly how long employees must wait after renouncing the collective bargaining process before courts will decide that an agreement meets *Brown*'s "distant in time and circumstance" standard. Nevertheless, an antitrust suit does remain an option for players lacking the bargaining leverage to achieve acceptable hours, wages, and working conditions through the collective bargaining process.

C. The Traditional Professional Sports League Model

Four sports leagues have traditionally dominated the national market for team sports: the National Football League ("NFL"), Major League Baseball ("MLB"), the National Basketball Association ("NBA"), and the National Hockey League ("NHL"). Each of these leagues exists as the product of a contractual agreement among its independently-owned member clubs, which compete against one another in their respective sports.

The member clubs, or franchises, vary widely in how they are organized. A franchise may be organized as a partnership, a privately held corporation, or a publicly traded corporation, among other forms. The leagues themselves are actually unincorporated, non-profit associations governed by an elected commissioner and an executive committee. The contract among the franchises sets forth the procedures by which league-wide rules are adopted and enforced, and

---

25 See Weiler & Roberts, supra note 5, at 468 (discussing "the four historically 'major' sports").
26 See id. (analyzing the structure of the traditional leagues).
27 See Gregor Lentze, The Legal Concept of Professional Sports Leagues: The Commissioner and an Alternative Approach from a Corporate Perspective, 6 Marq. Sports L.J. 65, 68-69 (1995) (describing the organization of the four major leagues). As a formal matter, MLB is structured somewhat differently. The organization is divided into the American and National Leagues, both of which are completely independent leagues, each with its own executive power. As a practical matter, however, the American and National Leagues' power is delegated to the Commissioner's office and an Executive Council, similar to the other leagues. See id. at 69.
delegates powers to the central league office.\textsuperscript{28}

D. Single Entity Theory and the Traditional Sports League Model: Existing Case Law

The traditionally organized professional sports leagues have long argued that they are in fact single entities, and that their practices are, therefore, not subject to section 1 challenges that various league rules tend to attract.\textsuperscript{29} The leagues "contend that a professional sports league is a unique business, containing an unusual but necessary mixture of interparticipant competition and cooperation not found in any other kind of partnership or joint venture."\textsuperscript{30} Furthermore, they argue that their teams are not competitors in a traditional business sense, but instead are integral parts of a single entity. Although the member clubs may be considered competitors by the public, which focuses on the athletic competition (each league's product), in reality, the actual outcomes of the athletic contests are irrelevant to the business of the league. In the leagues' view, although the teams may compete against one another for a player's services, this competition, like that on the field, is strictly controlled by league rules that are enforced by the central league office, and therefore should be considered an internal business matter.\textsuperscript{31}

The traditional leagues' position has received limited judicial approval. In one early case, \textit{San Francisco Seals, Ltd. v. NHL}, a federal district court did hold that a professional hockey team could not sue the league in which it competed under section 1 of the Sherman Act.\textsuperscript{32} The \textit{Seals} court stated that section 1 required "at least two independent business entities" and that in the production of professional hockey games before live audiences in the United States and Canada, "plaintiff and defendants are not competitors in the economic sense."\textsuperscript{33}

\begin{flushright}
\textsuperscript{28} See \textit{id.} at 68-69 (explaining the authority of the leagues' governing bodies).
\textsuperscript{29} See \textit{supra} Part I.B (discussing why sports league rules give rise to antitrust claims).
\textsuperscript{31} See \textit{id.} at 29, 31-32 (describing the traditional arguments raised by proponents of the single entity defense, and emphasizing the "unique" nature of the intraleague competition).
\textsuperscript{32} 379 F. Supp. 966 (C.D. Cal. 1974).
\textsuperscript{33} \textit{Id.} at 969.
\end{flushright}
1. Judicial Rejection of the \textit{Seals} Approach: Pre-\textit{Copperweld} Cases

The single entity defense was rejected by the Second Circuit in \textit{North American Soccer League v. NFL}, in which the North American Soccer League ("NASL") challenged the NFL's policy of prohibiting football team owners from owning teams in other sports leagues under section 1 of the Sherman Act.\footnote{670 F.2d 1249, 1256-57 (2d Cir. 1982).} The court concluded that characterizing the NFL as a single entity would create an antitrust "loophole" that could allow the league to adopt restraints to protect individual franchise owners.\footnote{See id. at 1257 ("To tolerate such a loophole would permit league members to escape antitrust responsibility for any restraint entered into by them that would benefit their league or enhance their ability to compete even though the benefit would be outweighed by its anticompetitive effects.").} The 	extit{NASL} court also emphasized the economic independence of the NFL's franchises:

[\textit{E}ach member [club] is a separately owned, discrete legal entity which does not share its expenses, capital expenditures or profits with other members . . . . [I]n spite of the sharing of some revenues, the financial performance of each team, while related to that of the others, does not, because of the variables in revenues and costs as between member teams, necessarily rise or fall with that of the others.\footnote{See Los Angeles Mem'l Coliseum Comm'n v. NFL, 726 F.2d 1381, 1390 (9th Cir. 1984) ("Raiders") ("[A] large portion of League revenue, approximately 90%, is divided equally among the teams . . . .").}]

Arguably, the Second Circuit's rationale values form over substance. Although each team is organized as a separate, independently-owned entity, the NFL teams could also be viewed as acting as a single corporate board of directors in promulgating the cross-ownership ban.\footnote{The Seventh Circuit implicitly adopted this view in \textit{Chicago Professional Sports Ltd. Partnership v. NBA}, 95 F.3d 593, 597-98 (7th Cir. 1996) ("WGN").} Additionally, the court understates the economic interdependence of the franchises in stating that the teams share "some" revenues. In fact, when 	extit{NASL} was decided, NFL teams shared 90\% of League revenues.\footnote{North Am. Soccer League, 670 F.2d at 1252.}
Soon after the NASL decision, the Ninth Circuit rejected the single-entity defense in a suit brought by the Los Angeles Raiders and the Los Angeles Memorial Coliseum, Los Angeles Memorial Coliseum Commission v. NFL ("Raiders"). The suit challenged the NFL's efforts to block the Raiders' move from Oakland to Los Angeles. The Ninth Circuit affirmed the trial court, holding that, as a matter of law, the NFL was not a single entity. The court based its decision on three main reasons:

Initially, the district court recognized the logical extension of the single-entity argument was to make the League incapable of violating Sherman Act § 1 in every other subject restriction—yet courts have held the League violated § 1 in other areas. Secondly, other organizations have been found to violate § 1 though their product was "just as unitary... and requires the same kind of cooperation from the organization's members." Finally, the district court considered the argument to be based upon the false premise that the individual NFL "clubs are not separate business entities whose products have independent value."

First, the court claimed that because other courts had found the League to have violated section 1, it must therefore be subject to section 1. The other decisions cited by the court were the NASL case, and three decisions involving NFL rules concerning player contracts: Smith v. Pro Football, Inc., Mackey v. NFL, and Kapp v. NFL. The Ninth Circuit's reliance on the latter three cases is extremely questionable. The NFL did not proffer a single entity defense in any of those cases, so those courts simply did not consider the issue. Therefore, the precedents on which the Raiders court relied were not particularly persuasive, even though the aforementioned courts could...

---

59 See id. at 1390 (affirming the district court's rejection of the NFL's single entity defense).
49 See id. at 1385 (describing the events preceding the litigation).
41 See id. at 1387 (holding that, based on the undisputed facts, the district court correctly directed a verdict for the plaintiffs).
42 Id. at 1387-88 (citations omitted).
43 See id. at 1388 ("[T]he logical extension of the single-entity argument was to make the League incapable of violating Sherman Act § 1 in every other subject restriction—yet courts have held the League violated § 1 in other areas.").
44 593 F.2d 1173 (D.C. Cir. 1978) (challenging the legality of the NFL's player selection process, commonly called the draft).
45 543 F.2d 606 (8th Cir. 1976) (challenging the NFL's rule allowing the league commissioner to require a club acquiring a free agent to compensate the free agent's former club).
46 586 F.2d 644 (9th Cir. 1978) (challenging, among other things, the legality of the NFL mandate that all players must sign a Standard Players' Contract which binds players to the NFL constitution and bylaws).
have looked outside the parties' arguments to decide the cases on a single entity-based rationale.

In its discussion of precedent, the Raiders court completely ignored Seals, despite its seemingly direct application to the issue. Instead, the Ninth Circuit relegated the Seals opinion to a footnote in its conclusion and rejected the single entity defense, stating that while Seals was "persuasive," "existing precedent [could not] be ignored."\(^\text{47}\) As "existing precedent" in reality consisted of a single Second Circuit opinion, this portion of the Ninth Circuit's reasoning appears to be profoundly unconvincing.

Second, the Raiders court reasoned that the NFL was similar to other businesses shown to be subject to section 1. The court noted the existence of a Ninth Circuit exception to the finding of concerted action where "multiple corporations [are] operated as a single entity" when 'corporate policies are set by one individual or by a parent corporation.'\(^\text{48}\) The NFL was not covered by this exception, however, because "NFL policies are not set by one individual or parent corporation, but by the separate teams acting jointly."\(^\text{49}\) The Seventh Circuit implicitly rejected this logic in Chicago Professional Sports Ltd. Partnership v. NBA by embracing the "corporate board" analogy, which views the clubs as members of a corporate board controlling league policy.\(^\text{50}\) It is not inaccurate to characterize a league as either "separate teams acting jointly" or as a single firm controlled by a board of directors, which is in part why courts willing to give the single entity argument its due consideration have struggled a great deal to reach a conclusion.

In rendering its holding, the Raiders court cited three well known cases in which concerted action was found even though, as is the case with a professional sports league, cooperation was necessary to produce the product:\(^\text{51}\) Associated Press v. United States,\(^\text{52}\) Broadcast Music, Inc. v. Columbia Broadcast System, Inc., ("BMI")\(^\text{53}\) and United States v.

\(^{47}\) Raiders, 726 F.2d at 1390 n.4.
\(^{48}\) Id. at 1388 (quoting General Bus. Sys. v. North Am. Phillips Corp., 699 F.2d 965, 980 (9th Cir. 1983)).
\(^{49}\) Id. at 1389.
\(^{50}\) 95 F.3d 593, 597-98 (7th Cir. 1996) ("WGN").
\(^{51}\) See Raiders, 726 F.2d at 1389 (discussing the relevance of the three cases).
\(^{52}\) 326 U.S. 1, 14-15 (1945) (holding that the cooperative nature of the enterprise did not make the defendant immune to section 1 liability).
\(^{53}\) 441 U.S. 1 (1979) (applying rule of reason analysis to blanket license offered by license holder of copyrighted songs).
Despite the Ninth Circuit's reasoning in *Raiders*, these cases are actually distinguishable because in each, the defendant independent firm had a viable product that did not require cooperation with competitors, but chose to cooperate with competitors to produce a second product. The cooperation was not necessary to the defendants' core business in those cases, only to the additional joint venture. Members of a professional sports league have no alternative but to cooperate with fellow members. The bedding manufacturers in *Sealy*, for example, could have simply continued to manufacture their own mattresses rather than form a joint venture with competitors; cooperation was not absolutely essential to the continued vitality of each individual manufacturer's operations. An NFL team, by contrast, has no product to offer if it does not cooperate with others. It could be argued that the facts of *BMI* are sufficiently analogous because the inability of individual license holders to protect their rights from infringement made the license essentially worthless by itself; only by combining with other license holders did the license hold value. The *BMI* court assumed, nevertheless, without discussion, that the groups of individual license holders were combinations for antitrust purposes. *BMI* was decided five years before *Copperweld*, however, and if the case was reargued in light of *Copperweld*, it is possible that the court might have accepted a single entity argument.

Third, the *Raiders* court refused to accept the NFL's argument that its teams were not "'separate business entities whose products have an independent value.'" The court discounted the cooperation necessary to produce a football game because a team could play outside of the league. The value of such an undertaking, however, especially over the long term, is so small in proportion to the value realized by a team competing in the NFL that it makes the court's argument entirely unconvincing. Although the court's independent

---

54 388 U.S. 350 (1967) (finding that the joint venture among bedding manufacturers violated section 1).
55 As the Seventh Circuit noted in *WGN*, "a league with one team would be like one hand clapping." *WGN*, 95 F.3d at 598-99.
56 See *BMI*, 441 U.S. at 7-9.
57 *Copperweld* Corp. v. Independence Tube Corp., 467 U.S. 752 (1984); see supra text accompanying notes 62-64.
58 *Raiders*, 726 F.2d at 1388 (citation omitted).
59 See id. at 1390 (arguing that teams could play outside of the NFL).
60 The NFL's Washington Redskins were recently sold for a reported $800 million. *See* Adrienne T. Washington, *Bully Turns Benefactor in $800 Million Stroke*, *Wash. Times*, Jan. 12, 1999, at C2 (discussing the sale of the team). If the Redskins had announced before the sale of the franchise that the team would no longer compete in the NFL,
value claim is somewhat understandable given the existence at the time of the United States Football League, which a NFL team could theoretically have joined, the likelihood of such a defection in reality was extremely small. By contrast, the independent value of the products made by the defendants in Associated Press, BMI, and Sealy were not based on a court’s whimsical speculation, but instead on their demonstrated performance in the marketplace. Therefore, such a comparison is inappropriate.

2. Copperweld Corp. v. Independence Tube Corp.

Soon after the Ninth Circuit’s decision in Raiders, the Supreme Court issued a factually distinguishable decision that nonetheless could be viewed as supporting the NFL’s single entity argument. In Copperweld Corp. v. Independence Tube Corp., the Court overturned the intra-enterprise conspiracy doctrine, which provided that agreements within corporations, usually between a parent corporation and its subsidiary, satisfied the section 1 conspiracy requirement. Instead, the Copperweld Court held that a parent corporation and its wholly owned subsidiaries constituted a single firm for antitrust purposes and were thus exempt from section 1. The Court’s approach to the concerted action requirement discounted the importance of an entity’s economic form, and instead instructed courts to look at the economic reality of the business’s structure. To proponents of the professional league’s single entity argument, this new emphasis on reality over form urged a reconsideration of the single entity claim.

3. Post-Copperweld Case Law

In the wake of the Copperweld decision, however, the courts have generally followed Raiders and distinguished Copperweld, rather than finding Copperweld sufficiently analogous to dismiss section 1 claims but would instead arrange exhibition games with whatever teams they could schedule, the team’s value would have plummeted.

The failure of any team in one of the four major professional leagues to defect to another league speaks to the economic irrationality of such a move.

467 U.S. 752, 772-73 (1984) (“Because there is nothing inherently anticompetitive about a corporation’s decision to create a subsidiary, the intra-enterprise conspiracy doctrine ‘imposes grave legal consequences upon organizational distinctions that are of de minimis meaning and effect.’”) (citations omitted).

See Copperweld, 467 U.S. at 776 (“[W]e can only conclude that the coordinated behavior of a parent and its wholly owned subsidiary falls outside the reach of [section 1].”).

See id. at 772 (rejecting the intra-enterprise conspiracy doctrine because it “looks to the form of an enterprise’s structure and ignores the reality”).
against the professional leagues. In *Sullivan v. NFL*, the First Circuit stated that "the critical inquiry is whether the alleged antitrust conspirators have a 'unity of interests' or whether, instead, 'any of the defendants has pursued interests diverse from those of the cooperative itself.'" The *Sullivan* court found that NFL teams compete with one another "for things like fan support, players, coaches, ticket sales, local broadcast revenues, and the sale of team paraphernalia." Due to the existence of this off-field competition, the teams did have "diverse interests," and thus failed the latter part of the "unity of interests" standard. As a result, the *Sullivan* court held that NFL teams were not a single entity for section 1 purposes.

The *Sullivan* court purported to adopt the standard for interpreting *Copperweld* set forth in *City of Mt. Pleasant v. Associated Electric Cooperative, Inc.* ("*Mt. Pleasant*"). In *Mt. Pleasant*, a joint venture that supplied electricity to rural communities was found by the court to be a single entity. The Associate Electric Cooperative featured an elaborate three-tiered ownership structure. It was owned by six generation and transmission cooperatives, which in turn were owned by forty-three distribution companies. These distribution companies' customers—425,000 individual consumers—owned the distribution companies. Therefore, the decision-making power within the Cooperative was, as a formal matter, diffused widely among a variety of separately owned economic entities.

Although the First Circuit in *Sullivan* and the Eighth Circuit in *Mt. Pleasant* agreed on the same legal standard to use, they applied that

---

65 34 F.3d 1091, 1099 (1st Cir. 1994) (quoting City of Mt. Pleasant v. Associated Elec. Cooper., Inc., 838 F.2d 268, 274-77 (8th Cir. 1988)). This approach was criticized by the Seventh Circuit in *WGN*:

Although [the complete unity of interests] phrase appears in *Copperweld*, the Court offered it as a statement of fact about the parent-subsidiary relation, not as a proposition of law about the limits of permissible cooperation. As a proposition of law, it would be silly. Even a single firm contains many competing interests.

Chicago Prof'l Sports Ltd. Partnership v. NBA, 95 F.3d 593, 598 (7th Cir. 1996) ("*WGN*").

66 *Sullivan*, 34 F.3d at 1098.

67 See id. at 1099 ("NFL member clubs compete in several ways off the field, which itself tends to show that the teams pursue diverse interests and thus are not a single enterprise under § 1.").

68 *Id.*

69 *Mt. Pleasant*, 838 F.2d 268 (8th Cir. 1988).

70 *Id.*

71 See id. at 271 (describing the complicated ownership and supply structure of the cooperative).
standard in very different ways. Much like the plaintiff in *Sullivan*, the petitioner in *Mt. Pleasant* argued that the cooperative members had diverse economic interests because they competed for customers, including municipal customers, were forced to arbitrate a dispute concerning rate structures, and disagreed over how to divide the cooperative's profits. The court found that this evidence proved that the cooperative members had diverse economic interests, but "not in the sense necessary to create a fact issue on whether these companies are part of a single enterprise." The *Mt. Pleasant* court's discussion of the intracooperative competition is notable because it focused much more on the overall purpose of the cooperative rather than on certain areas in which cooperative members competed.

Even though the cooperatives may quarrel among themselves on how to divide the spoils of their economic power, it cannot reasonably be said that they are independent sources of that power. Their power depends, and has always depended, on the cooperation among themselves. . . . The disagreements we have described are more like those among the board members of a single enterprise, than those among enterprises which are themselves separate and independent.

This analysis, when applied to a professional sports league, seems to support a finding of single entity status. The NFL owners do function like "board members of a single enterprise" when they promulgate the league-wide rules that are at issue in antitrust claims against the League. In addition, the League's economic success "depends, and has always depended, on the cooperation among [the member teams]." In light of this language, and more significantly, the overall reluctance of the *Mt. Pleasant* court to accept evidence of intracooperative competition so as to create even an issue of fact, it is clear that the *Sullivan* court interpreted *Copperweld* very differently than the *Mt. Pleasant* court. Furthermore, it seems clear that an Eighth Circuit court dutifully following *Mt. Pleasant* would have no choice but to hold

---

72 *Id.* at 276-77. In addition, the city argued that the cooperative members must have competed in the labor market, see *id.* at 277, a factor the *Sullivan* court found very persuasive in ruling that the NFL failed to meet the unity of interests standard, see *Sullivan*, 34 F.3d at 1098-99.
73 *Mt. Pleasant*, 838 F.2d at 277. The procedural posture of this case further points to the difference between the standards of the First and the Eighth Circuits. The *Mt. Pleasant* petitioners merely had to show that there was a triable issue of fact regarding the diversity of the cooperative members' interests. *Id.*
74 *Id.*
75 *Id.*
76 *Id.*
that a professional sports league is a single entity.

Nevertheless, in an unpublished opinion, a district court in the Eighth Circuit reached the same outcome as the Sullivan court by applying a different legal principle. In reviewing that trial court's decision, the Eighth Circuit in St. Louis Convention & Visitors Commission v. NFL noted that the district court found that the league was estopped from arguing that its teams comprised a single economic entity because of the Ninth Circuit decision in Raiders. The trial court also found that neither Copperweld nor Mt. Pleasant necessitated that it reconsider the league's single entity argument. The district court did dismiss the case during the course of the trial, however, because of insufficient evidence. The Eighth Circuit affirmed the trial court's dismissal, and therefore did not address the NFL's cross-appeal on the single entity question.

Much of the leading case law demonstrates the courts' hostility to the single entity argument, but a Seventh Circuit case, Chicago Professional Sports Ltd. Partnership v. NBA ("WGN"), suggests that the majority view has not necessarily gained a consensus and that the single entity question with regard to traditionally organized sports leagues remains an undecided one. In WGN, the trial court ruled that the NBA was not a single entity because its teams did not have a "complete unity of interest." The Seventh Circuit vacated that holding and remanded the case for further consideration. The Seventh Circuit agreed with the trial court's holding that NBA teams did not have a "complete

---

77 154 F.3d 851, 856 (8th Cir. 1998) (recounting the trial court's findings).
78 See id. ("The district court was not persuaded that two subsequent cases [(Copperweld and Mt. Pleasant)] dealing with the concept of single economic enterprise required a different result.").
79 See id. at 859 ("Since the court concluded that [the plaintiff] had presented no evidence to show that the NFL's rule and the guidelines actually had caused league teams other than the Rams to refrain from competitive bidding on the Trans World Dome lease, it granted the Rule 50 motion.").
80 See id. at 865 n.9 (dismissing the NFL's cross-appeal of the trial court's single entity ruling).
81 95 F.3d 593 (7th Cir. 1996).
82 Id. at 597. The case involved a dispute between the plaintiffs, the Chicago Bulls basketball team and WGN, a cable television station, and the defendant NBA, over how many Bulls games WGN was permitted to broadcast each basketball season. See id. at 595 (explaining that both sides have appealed the trial court's 30 game allowance, with the Bulls and WGN wanting to broadcast 41 games per year, and the NBA seeking to fix the number between 15 to 20). The league argued that the broadcasts infringed on the rights of other cable stations that had contracted with the NBA for rights to televise games nationally.
83 See id. at 601 (vacating the district court's judgment of "all events" and remanding).
unity of interest,” but held that such a determination was not necessarily dispositive of the single entity question. The court did not interpret *Copperweld* to “hold that only conflict-free enterprises may be treated as single entities.” After all, the court noted, “multi-stage” firms such as General Motors or IBM often have conflicts between various departments. According to the *WGN* court, *Copperweld* held that the concerted action requirement exists to scrutinize “[c]onduct that ‘deprives the marketplace of the independent centers of decision-making that competition assumes.’” In *Copperweld*, the parent-subsidiary relationship did not deprive the marketplace and, therefore, was not subject to section 1 scrutiny.

Applying this interpretation, the Seventh Circuit stated that “[w]e see no reason why a sports league cannot be treated as a single firm in this typology.” The court stopped short, however, of declaring the NBA a single entity. In remanding the case, the court stated that the single entity question requires a fact-specific inquiry into the nature of the league. The *WGN* court argued that it is entirely possible for one league (the NBA, for example) to be characterized as a single entity and for another league (the NHL, for example) to be subject to the Rule of Reason. Furthermore, separate aspects of league decision-making may be characterized differently. The *WGN* court stated:

[W]e do not rule out the possibility that an organization such as the NBA is best understood as one firm when selling broadcast rights to a network in competition with a thousand other producers of entertainment, but is best understood as a joint venture when curtailing competition for players who have few other market opportunities.

The standard set forth by the *WGN* court effectively articulates why the single entity question is so difficult, but fails to answer satisfactorily how professional sports leagues ought to be characterized. The “depriving the marketplace of independent centers of decision-making” standard fails to give much guidance to the trial courts that must actually make such determinations. Furthermore, it requires in each in-

---

84 Id. at 598.
85 Id.
86 Id.
87 Id. (quoting *Copperweld Corp. v. Independence Tube Corp.*, 467 U.S. 752, 769 (1984)).
88 Id.
89 See id. at 600 (“Sports are sufficiently diverse that it is essential to investigate their organization and ask *Copperweld*’s functional question one league at a time . . . .”).
90 Id.
stance a fact-intensive review of the facets of a league's operation, as well as comparison with other leagues and other business forms, to reach a conclusion. These requirements, therefore, deny all parties the efficiency benefits of a bright-line rule. An inquiry of this sort has not yet taken place, as the parties in WGN reached a settlement after the Seventh Circuit's decision was rendered. 91

A review of the case law concerning the traditional leagues' single entity argument shows a prevailing hostility to the leagues' claim that they constitute single economic units for antitrust purposes. Courts have consistently followed the result reached by the court in Raiders, notwithstanding the questionable rationale behind that decision. The Seventh Circuit's WGN decision suggests, however, that the rule of Raiders has not been universally accepted. The parameters of the debate will undoubtedly influence the Fraser court's reasoning in affirming or denying MLS's single entity claim.

II. THE NEW PARADIGM: SINGLE ENTITY LEAGUES

The existing case law on the single entity question discussed in Part I.C only directly applies to leagues organized according to the traditional model as described in Part I.A. Partly in response to the courts' generally hostile reaction to the established leagues' single entity arguments, several upstart leagues have organized themselves as single corporate entities. 92 Before evaluating the legal arguments available to the litigants in Fraser v. MLS, it seems appropriate to heed the Seventh Circuit's admonition, 93 and examine the nature and practices of the league at issue, and compare its structure to those of fellow single entity leagues.

A. Benefits of a Single Entity League

Given the history of repeated antitrust challenges to the traditional league practices described in Part I.B, any protection the leagues can muster from antitrust scrutiny would benefit them considerably. If the leagues are viewed as single entities, they could avoid the potential of treble damages as well as the costs of litigation from

92 See infra Part II.A (discussing leagues organized as single entities).
93 See supra notes 81-90 and accompanying text (discussing the Seventh Circuit's decision in WGN).
suits arising under section 1 of the Sherman Act. The Sherman Act provides that successful plaintiffs may recover three times the amount of their actual damages. See 15 U.S.C. § 15 (1994). Potential plaintiffs would be forced to challenge league rules as abuses of monopoly power under section 2 of the Sherman Act, a far more difficult claim to prove. The league would thus enhance its bargaining leverage over its players' union because the union would be unable to threaten the league with a section 1 lawsuit. The corporate structure of the league would offer other advantages as well. The league can increase the value of its sponsorship agreements by ensuring its league-wide sponsors that individual clubs will not enter into sponsorship agreements with competitor firms that dilute the value of the league-wide sponsor's investment. The corporate league can also reduce the number of decisions that require building a consensus among league owners. By contracting to place league decision-making power in the hands of the central league office, the corporate league can prevent franchises from relocating, realign divisions, and otherwise make decisions that serve the purposes of the league as a whole rather than individual owners. Finally, the league can achieve economies of scale through increased purchasing power.

B. Organization of Major League Soccer

MLS was initially designed by Los Angeles attorney Alan Rothen-
berg as a limited-liability Delaware corporation run by a Board of Directors appointed by league investors.\textsuperscript{100} Rothenberg planned that MLS would own and operate all of the teams in the league, assign players and team personnel, and set local ticket and concession prices. League investors would merely own shares in MLS itself.\textsuperscript{101} This structure effectively addressed the reasoning of the \textit{Sullivan} court, which rejected the NFL's single entity defense because NFL teams compete off-field,\textsuperscript{102} by simply eliminating all forms of off-field competition. The only competition possible in this model is in the form of the on-field games themselves. In MLS's view, the games do not represent competition in any economic sense, but rather represent the product of the MLS corporation. It seems highly likely, therefore, that if MLS had retained this structure, the players' section 1 claims in \textit{Fraser} would have been dismissed under \textit{Copperweld}.\textsuperscript{103}

MLS was forced to modify its structure before it began operating, however, when it had difficulty attracting investors. Potential investors were discouraged from investing by the anonymity of their roles and preferred to be more involved in running a team.\textsuperscript{104} Under the new arrangement, MLS retained formal ownership of its franchises but issued a special class of stock to "investor-operators." This special class of stock gave them "almost full operating control" over the management of a particular franchise.\textsuperscript{105} The franchises and the league equally share local revenues (generated by the teams from ticket sales, concessions, signage, etc.), while MLS retains all national television and merchandising revenues. The funds MLS receives are used in part to pay all player salaries. The investor-operators receive dividends on profits from league operations and may also sell their special stock to outside groups, just as in traditional leagues.\textsuperscript{106} In this new structure, MLS teams do compete with each other in some of the same ways noted by the \textit{Sullivan} court. Thus, the new structure is clearly a less "pure" single entity than the MLS Rothenberg originally envisioned, and somewhat more like traditionally organized leagues.\textsuperscript{107}

\textsuperscript{100} See \textsc{Weiler \& Roberts}, supra note 5, at 495-96 (discussing MLS's original design).
\textsuperscript{101} See id. (detailing MLS's original structure).
\textsuperscript{102} See supra text accompanying notes 65-68 (discussing the \textit{Sullivan} decision).
\textsuperscript{103} See \textsc{Weiler \& Roberts}, supra note 5, at 497 (describing the outcome as "almost certain").
\textsuperscript{104} See id. at 496 (explaining why MLS initially had problems attracting investors).
\textsuperscript{105} \textit{Id.}
\textsuperscript{106} See id. at 496 (setting forth the details of MLS's current organization).
\textsuperscript{107} The structures of two other recently formed single entity sports leagues show the extent to which the structures of single entity leagues can differ. The American Bas-
There is another extremely important structural difference between MLS and the traditional leagues, however, which tempers intraleague competition and thus complicates application of the Sullivan court's standard to MLS. MLS allows its investor-operators to run more than one team. In fact, three individual investors manage a total of seven teams, more than half of the twelve teams currently competing in MLS.108 The players and coaches selected by the MLS investor-operators with multiple teams are thus more accurately described as "allocated" to a particular team. Overall, intraleague competition is substantially foreclosed because so many teams share common investor-operators.109 The prospect of the vigorous off-field competition found by the Sullivan court is highly unlikely because of the potential

ketball League ("ABL"), which has since suspended operations and declared bankruptcy, originally consisted of eight teams. See Amy Shipley, ABL Says It Is Bankrupt and Shuts Down, WASH. POST, Dec. 23, 1998, at D1 (reporting the ABL's announcement that it was suspending operations). Both the players and management of each team were employed and allocated to individual teams by the league office. In addition, the players were given a 10% interest in the ABL. See WEILER & ROBERTS, supra note 5, at 497-98 (describing the ABL's corporate structure). The Women's National Basketball Association ("WNBA") could be characterized as a "less pure" single entity than the ABL. The WNBA LLC (limited liability company) is a corporation owned by NBA Development, which in turn is owned by the 29 NBA franchises, each of which has different owners. See id. at 497 (discussing the WNBA's structure). The NBA teams have official control of the WNBA, but a board of eight NBA owners and the NBA commissioner sets WNBA policy. The WNBA, like the ABL, hires both players and coaches, and assigns them to individual teams. The NBA teams who play in the same city as WNBA teams manage the franchises and receive a share of local revenues generated by the team. See id. (contrasting the ABL and the WNBA). Since the WNBA players have chosen to form a union and begin bargaining with WNBA management rather than pursue litigation like the MLS players, it is unlikely that a decision regarding application of section 1 to the WNBA is imminent. See W.H. Stickney Jr., WNBA Athletes Vote to Affiliate with NBA Players' Union, HOUSTON CHRON., Nov. 6, 1998, at 2 (discussing the results of the vote certifying the National Basketball Players Association as the WNBA players' official bargaining representative).

108 See Michael Hiestand, A Family That Plays Together, USA TODAY, Dec. 2, 1998, at 3C (noting this apparent conflict of interest). Robert Kraft and Lamar Hunt, both of whom also own NFL franchises, operate two franchises each. See id. (asserting that three owners control a majority of the franchise's teams). Phillip Anschutz, a part owner of the NHL's Los Angeles Kings, operates a total of three franchises.

109 This structure also raises the danger that a single investor-operator may attempt to load one of her teams with all of her best players. As a result, the power of the MLS Commissioner to block trades (a power each Commissioner also enjoys in the traditional league model) is especially important in maintaining a competitive balance throughout the league. See id. (explaining that the MLS Commissioner is "a final check and balance" on any owner attempting a shady trade between [her] teams"). Even without this safeguard, it is unlikely that the benefits an owner would accrue from creating a "loaded" team would outweigh the losses the owner would suffer from the reduced competitiveness of her other teams.
for direct harm to the operator's other intraleague interests. In addition, MLS still operates two teams, which further weighs against the existence of significant intraleague competition. Thus, the fortunes of a single MLS team are so inextricably linked to the success of all the other teams that any argument emphasizing a diversity of interests among MLS teams seems extremely far-fetched.

C. Fraser v. MLS: The Players' Claims

The league practices at issue in Fraser concern MLS's restraints on its teams' ability to acquire players. In this area of league operations, MLS has retained a system very close to the more pure single entity model originally envisioned by Rothenberg. Once an investor-operator decides which players she would like to acquire, MLS negotiates a contract with the player. The terms of the agreement are subject to the league's salary cap and any restraints on player movement, after the acceptance of which the player becomes an employee of the league, rather than of the "allocated" team. The players allege that MLS's entire system of player restraints violates the antitrust laws. First, they argue, the salary cap, the centralized player allocation system, and the standard reserve clause in every player contract unreasonably restrict the labor market for professional soccer players in the United States by preventing MLS teams from competing against each other for the services of individual players. Second, the standard MLS player contract unlawfully restricts the players' ability to license their names and images. Third, the system of transfer fees promulgated by the sport's International Governing Body, FIFA, and enforced by the United States Soccer Federation and MLS, violates the antitrust laws by restricting players' ability to seek employment in

---

110 See id. ("[T]he idea is...successful businessmen won't abuse their multiple-team ownerships and risk undermining the league's credibility because, ultimately, they're business partners rather than competitors.").


112 See Weiler & Roberts, supra note 5, at 496-97 (discussing MLS's restrictions on player contracts).


114 See id. at 230 (noting the players' claim that "[t]he MLS standard player agreement denies players a fair share of group licensing rights").
other leagues.  

D. Fraser v. MLS: Assessing the Arguments

In assessing the applicability to MLS of the concerted action requirement of section 1, there are several arguments the Fraser court is likely to consider.

1. Copperweld Is Controlling

MLS can argue that Fraser is factually indistinguishable from, and therefore directly controlled by, Copperweld's bright-line rule. As MLS and its shareholders wholly own the franchises, their intracorporate decisions, or league rules, cannot be deemed concerted action for section 1 purposes, and therefore are not subject to section 1 liability. MLS can even argue that its organization is a more pure single entity than the parent corporation and its wholly owned subsidiary in Copperweld. Under this view, MLS's franchises are not subsidiaries of the league, but more like "plants" where the corporation outputs its product.

2. "Form vs. Reality" Analysis

The court may decide, however, that the result in Copperweld is not directly controlling. In that case, the players can argue that the rationale of Copperweld demands a finding of concerted action. Copperweld emphasizes that the concerted action requirement directs courts to look at the economic reality, rather than the form, of the business association. Permitting MLS to escape section 1 liability similarly values form over economic reality. From this perspective, the single entity league has the same purpose, and produces the same basic product, as the traditionally organized leagues, and therefore should receive the same treatment under the antitrust laws. The competition

115 See id. (noting the players' transfer fee claim). The transfer fee claims concern MLS's agreements with another economic entity, FIFA, and thus will not be affected by MLS's single entity argument.

116 See Copperweld Corp. v. Independence Tube Corp., 467 U.S. 752, 756 (1984) (recounting the Copperweld Corporation's transfer of the Regal Company's assets into a subsidiary corporation, Regal Tube). After all, MLS consists of only one corporation, not two, as was the case in Copperweld.

117 In fact, the Court repealed the intra-enterprise conspiracy doctrine in Copperweld precisely because it "looks to the form of an enterprise's structure and ignores the reality." Id. at 772.
for coaches and other team personnel among MLS franchises is just like the off-field intraleague competition in the traditional leagues that the First Circuit emphasized in *Sullivan*.

In addition, the nature of the special class of stock owned by the investor-operators creates incentives for each operator to compete at the expense of fellow shareholders. The investor-operators keep 50% of the ticket sales, local broadcast, and local sponsorship revenue generated by their respective teams, and thus have divergent economic interests. Furthermore, should the investor-operator choose to sell her stock, the stock's value will reflect her team's competitive success. A winning team is likely to encourage greater attendance as well as generate additional revenue by hosting playoff games. Teams with a larger fan base and a healthy balance sheet will appreciate in value as compared to less successful clubs with smaller followings. Although the financial health of the league as a whole may be as, if not more, important than the prospects of an individual franchise, the on-field success or failure of a particular investor-operator's team affects the return on her investment. This economic reality creates additional incentives for the investor-operator to compete at the expense of the rest of the league.

The players' position is problematic, however, in several respects. First, MLS can dispute the players' characterization of "economic reality." The investor-operators that the players would perceive as independent economic actors are merely shareholders in the MLS corporation. It would be unthinkable in another context for a court to find concerted action through the actions of the shareholders of a single corporation. As the Seventh Circuit noted in *WGN*, simply because a large corporation like GM may have employees or shareholders with disparate interests does not mandate characterizing GM as something other than a single entity.

The fact that the interests of shareholders may diverge under certain circumstances is clearly insufficient to warrant a finding of concerted action.

Second, the players' argument is also questionable because courts addressing the single entity question with respect to traditional leagues have found that the form of a sports league (an association of

---

118 *See supra* notes 65-68 and accompanying text (discussing *Sullivan*).

119 *See Singh,* *supra* note 113, at 281 (describing the allocation of MLS revenues).

120 *See* Chicago Prof'l Sports Ltd. Partnership v. NBA, 95 F.3d 593, 598 (7th Cir. 1996) (*WGN*) ("Conflicts are endemic in any multi-stage firm . . . but they do not imply that these large firms must justify all of their acts under the Rule of Reason.").
independently owned businesses) is its economic reality. The fact that the member clubs are independently owned and compete in some respects has been more persuasive than league arguments emphasizing the degree of cooperation undertaken by the member clubs. By this reasoning, Copperweld could actually be distinguished from Fraser (to MLS's benefit) because a sports league's unique qualities dictate that its form does matter, and is in fact the defining aspect of its economic nature. In finding insufficient evidence that the independently owned franchises acted as a single entity, the court rejected arguments concluding that form was irrelevant. Given this precedent, the Fraser court may be required to accept a sports league's form as indicative of its status as a single economic entity. Just as the court in Sullivan refused to look past the divergent interests created by independent club ownership, the Fraser court could find that the lack of independent owners is dispositive evidence of a unitary actor. To find otherwise would seem inconsistent with Sullivan, which, as a First Circuit opinion, is binding precedent for the District of Massachusetts Fraser court.

Finally, the players' position is problematic because the Copperweld court championed the right of business associations to organize themselves as they deem appropriate: "[A] business enterprise should be free to structure itself in ways that serve efficiency of control . . . ." The single entity structure allows MLS numerous efficiencies. MLS's freedom to self-organize would be compromised significantly if courts were free to ignore how a business chose to organize itself, opting instead to characterize a business as something other than its chosen form.

3. Following WGN: Distinguishing MLS’s Operations

As an alternative to the "form vs. economic reality" argument, the MLS players can use the WGN opinion to distinguish certain aspects of MLS's operations from its relations with its players, the practices at is-

---

121 See Sullivan v. NFL, 34 F.3d 1091, 1099 (1st Cir. 1994) (holding that because "NFL member clubs compete in several ways off the field," they could not be viewed as a single entity); Los Angeles Mem'l Coliseum Comm'n v. NFL, 726 F.2d 1381, 1389 (9th Cir. 1984) ("Raiders") (finding that the clubs were "separate business entities").

122 See Sullivan, 34 F.3d at 1099 (finding that NFL teams have divergent interests).


124 See Lebowitz, supra note 97, at 1F (noting the advantages of a single entity league, including controlling player costs, greater leverage in negotiating broadcasting contracts, and coordinated marketing strategies that eliminate the problem of ambush marketing).
sue in *Fraser*. According to the Seventh Circuit, it is perfectly permissible to treat a traditionally organized league as a single entity in some respects and as a joint venture in others.\textsuperscript{125} Under this view, even if MLS is regarded as a single entity in some aspects of its operations, such as national television and marketing deals, and franchise relocation and expansion, it should still be treated as a joint venture in the player market. The nature of a sports league is such that individual teams will always compete with one another in the player market. In order to achieve on-field competitive success, teams must make themselves attractive as potential employers to players both within and outside of the league. Although MLS is technically responsible for negotiating player salaries and signing the players to employment contracts,\textsuperscript{126} it is really the demand of the individual teams for particular players that triggers league negotiations with those players. The league, therefore, merely acts as an agent for the teams by handling contract negotiations. Likewise, in the traditional leagues, the commissioner's office must approve all contracts and trades between teams to ensure that the transactions comply with league rules. This rule, however, does not render the traditional leagues single entities.

The end result of these labor practices, the *Fraser* plaintiffs can argue, is that MLS fails both the First Circuit's and the Eighth Circuit's "unity of interests" standards, articulated in *Sullivan*\textsuperscript{127} and *Mt. Pleasant*,\textsuperscript{128} respectively, as well as the Seventh Circuit's "depriving the marketplace of independent decision-makers" test, articulated in *WGN*.\textsuperscript{129} The competition between teams for players violates the "unity of interests" standard because teams are pursuing disparate interests as they attempt to attract the best players. MLS also fails the Seventh Circuit's test because the artificial market constraints imposed by MLS greatly reduce the autonomy of the individual decision-makers in the players' market. The market for professional soccer players' services in the United States would surely benefit from the removal of these restraints.

MLS can respond to the players' labor market argument by distinguishing the Seventh Circuit's decision in *WGN*. First and foremost, the *WGN* holding applies to traditional sports leagues, and not to

\textsuperscript{125} See *WGN*, 95 F.3d at 600 ("[T]he ability of sports teams to agree on a TV contract need not imply an ability to set wages for players.").

\textsuperscript{126} See supra text accompanying notes 100-107 (discussing MLS's structure).

\textsuperscript{127} *Sullivan*, 34 F.3d at 1099.


\textsuperscript{129} *WGN*, 95 F.3d at 598-99.
leagues organized as single corporations. If, instead, WGN is viewed as directly relevant, MLS can argue that the holding supports its argument. WGN directs courts to scrutinize carefully the nature and form of each challenged aspect of the sports league’s practices. When MLS reorganized itself into a less “pure” single entity, it did not alter its system of negotiating all contracts with players and retaining complete control over assigning players to a particular team. Even though investor-operators may request that the league sign a particular player, the league’s control over the process remains so tight that intraleague competition for players is muted. Although this reality has profoundly anticompetitive effects in the player market, those effects are not relevant for determining whether concerted action exists.

MLS’s player acquisition system is directly analogous to a corporation that solicits advice from its plant managers before hiring its employees. If these plant managers receive potentially lucrative compensation bonuses when their plants outperform the other company plants (as MLS’s investor-operators do if their special class of stock appreciates), the plant managers obviously would lobby the central hiring office to secure the best employees for a particular plant. It is also possible that the skills of a particular prospective employee (like that of a superstar athlete) are so exceptional that the plant managers would be willing to offer the prospective employee significantly more compensation than the company is willing to pay. The company’s refusal to pay the employee more than what the company budgeted, however, does not in itself create an antitrust violation. The prospective employees could argue that the company is “depriving the marketplace of independent centers of decision-making” by not allowing the individual plant managers to pay what they believe the employee is worth. Even so, the courts would likely honor the single corporate form that the company had adopted and refrain from finding concerted action, despite the anticompetitive effect of the company’s practices. MLS’s control over the supply of players entering the league is no different from the company’s control over employees in this corporate example.

The argument against the existence of concerted action is further strengthened by the fact that more than half of MLS’s teams are

---

150 See WGN, 95 F.3d at 600 (“[I]t is essential to investigate [a league’s] organization and ask Copperweld’s functional question one league at a time—and perhaps one facet of a league at a time . . . .”).
151 The anticompetitive effects would of course become relevant if concerted action was found.
owned by investor-operators who own multiple teams.\textsuperscript{132} Aside from the fact that the league would block an investor-operator from doing so, any efforts by a particular investor-operator controlling multiple teams to "buy a championship" for one of her teams would inevitably damage her investment in her other team(s). Thus, for a majority of teams, reckless spending on players would be counterproductive to the interests of the free-spending investor-operator, not just the league as a whole.

The weight of the arguments favors MLS's position. Substituting the single entity sports league for the traditional model changes such dispositive characteristics that it becomes illogical to treat a league like MLS as a collection of economic actors competing in any meaningful economic sense. Therefore, section 1 of the Sherman Act should not apply to MLS.

III. THE POSSIBILITY OF SECTION 2 LIABILITY

A legal victory for MLS on the single entity question would only result in the dismissal of the MLS players' section 1 claim. The players could still pursue a claim under section 2 of the Sherman Act because section 2 does not have a concerted action requirement.

Section 2 prohibits monopolization and attempted monopolization by a single economic actor.\textsuperscript{133} The players could allege that MLS's practices constitute an abuse of the league's monopoly power in the market for soccer players. Generally, players who have brought antitrust claims against leagues have not used section 2 and proving a section 2 violation has, in practice, been far more difficult than making a successful section 1 claim.\textsuperscript{134}

To make a successful section 2 claim, the MLS players would first

\textsuperscript{132} See supra note 108 and accompanying text (noting that several MLS owners operate more than one MLS team).

\textsuperscript{133} See 15 U.S.C. § 2 (1994) ("Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States . . . shall be deemed guilty of a felony . . . ").

\textsuperscript{134} See Jacobs, supra note 30, at 28 n.12 ("[C]hallenges to league activities under section 2 have been infrequent and, for a variety of reasons, unsuccessful."). Section 2 cases generally have featured other leagues or franchises as plaintiffs. See, e.g., United States Football League v. NFL, 842 F.2d 1335, 1341-42 (2d Cir. 1988) (upholding the jury's finding that the defendant was guilty of section 2 antitrust violations but liable for only one dollar); Mid-South Grizzlies v. NFL, 720 F.2d 772, 788 (3d Cir. 1983) (upholding the NFL's refusal to admit a former World Football League franchise into the NFL); American Football League v. NFL, 323 F.2d 124, 134 (4th Cir. 1963) (failing to find that the NFL monopolized the relevant market).
have to show that the relevant product and geographic market is the labor market for elite professional soccer players in the United States. Assuming arguendo that the players successfully characterize this labor market as the relevant product market, they will likely have a much more difficult time proving that the relevant geographic market is the United States. The players would have to prove that if a hypothetical monopolist soccer league in the United States were to impose wage restrictions that depressed wages by a small but significant amount (usually 5-10%), players would have no option but to sign with the monopolist league. This argument ignores the fact that there are many other professional soccer leagues around the world that offer comparable (and in many cases, superior) competition and compensation. Although players may prefer to play in an American league for a variety of reasons, the court may view the availability of close substitutes to an American league as reason to designate the world-wide market as the relevant market for the services of professional soccer players. Because MLS is only one of many top leagues scattered throughout the world, it falls far short of the 50% minimum market share generally required to prove market power. This failure to prove market power would result in dismissal of the section 2 claim against MLS.

135 See United States v. Grinnell Corp., 384 U.S. 563, 570-71 (1966) (“The offense of monopoly under § 2 of the Sherman Act has two elements: (1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident.”).
136 The NFL’s repeated attempts to argue that it “faces market competition” from other professional sports leagues and other forms of entertainment have been rejected. WEILER & ROBERTS, supra note 5, at 603. It is also worth noting that there is considerable scholarly debate as to whether the antitrust laws (either section 1 or section 2) apply to labor markets at all. See id. at 152-53 (noting that because “monopoly” power held by the purchaser may merely induce a wealth transfer, and actually enhance consumer welfare, a buyer’s control of the labor market may not be subject to antitrust scrutiny).
137 See United States v. Engelhard Corp., 126 F.3d 1302, 1304 (11th Cir. 1997) (describing, but not adopting, the Department of Justice’s test for determining the relevant product market).
138 See Domed Stadium Hotel, Inc. v. Holiday Inns, Inc., 732 F.2d 480, 489 (5th Cir. 1984) (“Supreme Court cases . . . suggest that absent special circumstances, a defendant must have a market share of at least fifty percent before he can be guilty of monopolization.”).
139 Even if the players’ proffered market definition was accepted by the court, they still might have difficulty proving that MLS abused its market power. The standard for assessing what constitutes an abuse of monopoly power is the subject of considerable disagreement. Some circuits have adopted an extremely pro-defendant test that requires the plaintiff to prove that the sole purpose of the conduct at issue was to harm
The problems of proving both the existence of market power and the abuse of that power highlight some of the challenges involved in pursuing a section 2 claim, especially against MLS. These difficulties only further accentuate the importance of the single entity question—should MLS prevail on that issue, it likely will be extremely well insulated from any antitrust challenges by its players.  

IV. A PLAN TO REORGANIZE THE TRADITIONAL LEAGUES

New sports leagues need not be the only leagues to enjoy the benefits of the single entity structure. Should MLS succeed in arguing that it is a single entity for antitrust purposes, as this Comment predicts, the traditional leagues should reorganize themselves along the MLS (or some similar) corporate model.  

The reorganization plan set forth in this section modifies a plan proposed by Jeffrey A. Rosen-
that in The Football Answer to the Baseball Problem: Can Revenue Sharing Work\textsuperscript{42} in which he attempts to solve an entirely different problem plaguing one of the traditional leagues: the large disparities in the economic strength of Major League Baseball's franchises.\textsuperscript{145}

A. Big League Perestroika

To begin, I propose that each of the traditional leagues change their status from an unincorporated association to a limited liability corporation. The "corporate league" would then buy its franchises from the owners for the franchises' value, as determined by an independent appraiser.\textsuperscript{144} Any owner who prefers not to continue under the new corporate system would receive cash for her franchise. Participating owners, like MLS investor-operators, would receive two


\textsuperscript{144} As described in Rosenthal's plan, the franchise values could be adjusted to account for anticipated changes in particular league rules. For example, baseball might choose to adopt increased revenue sharing measures, which Rosenthal deemed essential to the game's economic stability. See id. at 423-28 (discussing the need for revenue sharing).
kinds of stock in return for ownership rights to their respective teams. One class of stock would provide for the traditional voting rights within the league that all franchise owners currently enjoy, as well as a share of the league's profits. The other class of stock would grant control of the day-to-day operations of the individual franchises to their previous owners and allow the owners to sell this stock to outside bidders. The franchises themselves could be owned directly by the incorporated league or exist as subsidiaries wholly owned by the parent league corporation, similar to the defendants' arrangement in Copperweld.145

The advantages of the corporate league would be manifold. First and foremost, the league would not face section 1 liability. Because almost every aspect of traditional league operations has been challenged under section 1, leagues necessarily are constrained in adopting new rules and entering into new arrangements by the constant threat (if not the certainty) of a section 1 challenge. By removing section 1 liability, the leagues would not have to overcome such barriers, and would also avoid the expense of litigating many lawsuits. Although it is certainly true that claims could still be brought under section 2, the difficulty of making such a claim successful will likely deter many suits. The corporate league would also enjoy increased bargaining leverage over its players, increased sponsorship opportunities, more efficient decision-making, and improved economies of scale.146

B. Potential Problems with the Proposal

The proposed reorganization is admittedly a radical measure that would fundamentally change the basic structure of each league. The reorganization would probably raise some significant, but not insurmountable, problems. The advantages of the corporate league are so sizeable, however, that a close examination of the problems facing reorganization should not overshadow the long term benefits that the corporate league would enjoy. As one source noted, "[i]f [NFL] Commissioner Paul Tagliabue could convert the NFL to a single entity, he'd do it tomorrow."147

Some team owners may be strongly opposed to such a plan because they derive value from the status their control of a team brings in the community and amongst other wealthy people. Not surpris-

146 See supra Part II.A (discussing the benefits of single entity leagues).
147 Lebowitz, supra note 97, at 1F.
ingly, MLS had great difficulties attracting investors under its original plan, which did not provide for the investors to be associated with a particular team. The plan essentially would have turned the investors into anonymous stockholders. MLS quickly realized its mistake and created a second class of investor-operator stock, which allows an owner to exercise control over a particular team, and thus achieve the public visibility that was absent in the league’s original plan. Owners in other leagues may also fear that they will lose the ability to run their teams without interference, as well as the community recognition that has made sports teams such valuable commodities. By issuing the second class of investor-operator stock, however, the reorganized leagues can ensure that the owners retain the value they derive from the status value of team ownership. Owners would also appreciate the fact that the corporate league would be more profitable due to the economic advantages gained through the change in structure. Greater profitability translates into both higher dividends and higher valuations of the individual teams, which should cause the second class of stock held by the owners to appreciate.

On the other hand, the traditional model may provide certain tax advantages over the single entity model since owners can deduct the operating losses of their individual franchises. The projected loss of tax benefits, however, can be figured into the original appraised price of the franchise. Furthermore, other owners whose franchises regularly recognized annual operating profits may realize tax savings under the single entity plan. Thus, it is likely that any tax considerations can be handled as an internal league matter, with some owners using their gains to compensate others for their losses. Moreover, the dwindling breed of owners who have held franchises for a long time, do not have sizeable holdings in other businesses, and are concerned chiefly with estate taxes also should not be negatively affected since the sale of the teams to the league in exchange for stock would have no effect on estate taxes. Thus, tax concerns should not prove to be a significant obstacle to league reorganization.

The players’ union, fearing the loss of a section 1 claim as a bargaining chip with the league, may file an unfair labor practice charge

---

148 See WEILER & ROBERTS, supra note 5, at 495 (discussing the problems with the original plan).
149 See id. at 496.
150 See supra Part II.A (discussing the advantages of the corporate structure of the league).
151 See I.R.C. § 165(a) (1999) (allowing a deduction for “any loss sustained during the taxable year”).
claiming that the reorganization must be negotiated with the union. Although employers are obligated to bargain with unions over changes "with respect to 'wages, hours, and other terms and conditions of employment,'" certain aspects of a firm's business are reserved to the discretion of the firm's management and do not require bargaining with unions. This includes the manner by which a firm chooses to organize itself. Therefore, a league would not have to bargain with the union over its decision to reorganize itself.

Additionally, the players or another party, such as the federal government, may also attempt to block the reorganization through antitrust law. The single entity question will probably not become a settled matter of law following the disposition of Fraser v. MLS, unless the Supreme Court renders a judgment on the matter. Until that time, a reorganized league would have to accept a degree of uncertainty that it may not be deemed a single entity for section 1 purposes.

Concerns about a section 2 claim may limit some of the benefits reorganization may provide. Section 2 jurisprudence, especially regarding labor claims against employers, is so unsettled that the leagues may be reluctant to risk incurring treble damages by adopting new restraints in the player market. For section 2 purposes, the traditional leagues are unlike MLS in that each one is the preeminent league of its kind in the world. The NFL, for example, has no serious competition from other football leagues in terms of both the quality of play and the level of compensation for its athletes. NFL players pursuing a section 2 claim should be able to define the relevant market as that for elite professional football players in the United States, where the NFL has a complete monopoly. But even if the players are successful in showing a single entity league has market power, the

---


153 See NLRB v. International Harvester Co., 618 F.2d 85, 87 (9th Cir. 1980) ("Management decisions that fundamentally alter the direction of an enterprise ... generally are not considered decisions concerning terms and conditions of employment and are not mandatory subjects of bargaining." (citing Fibreboard Paper Prods. Corp., 379 U.S. at 223 (Stewart, J., concurring))).

154 See supra notes 139, 141 (discussing the circuit split concerning section 2 standards).

155 The other major leagues—the NHL, the NBA, and MLB—might be able to argue that international leagues are viable substitutes in the market for each league's players, but such an argument is unlikely to be successful given the sizable gap in standards of play and compensation between these North American-based leagues and other leagues around the world.
task of proving abuse of market power still remains a difficult one. \(^{156}\)

Finally, the provisions of section 7 of the Clayton Act might be used to attack a league's reorganization in a suit brought by the federal government or the players. Section 7 provides that "[n]o person . . . shall acquire the whole or any part of the assets of another person engaged also in commerce . . . where . . . the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly."\(^{157}\) A union making a section 7 claim would attempt to characterize the league's action as a merger, rather than a reorganization, which substantially lessens competition in the player market. The union's merger claims would be weighed against the familiar (and heretofore unsuccessful) league claims that it is a single entity, but also, perhaps more significantly, the "Copperweld" Court's emphasis on the freedom of business enterprises to organize themselves as they see fit.\(^{158}\) In comparison to the traditional league model, the single entity structure features significant efficiencies which, according to "Copperweld," a business enterprise should be able to realize, free from government interference.\(^{159}\) A section 7 challenge, therefore, is by no means an insurmountable obstacle for traditional leagues seeking to enjoy the advantages of a single entity structure.

**CONCLUSION**

The advent of the single entity sports league may prove to be a vitally important event for the future of professional sports leagues. Although the demise of the American Basketball League\(^{160}\) shows that single entity leagues are by no means exempt from the cruel economic realities of competition in the sports and entertainment marketplace, the single entity form does provide a structure in which nascent sports leagues may flourish more readily with proper marketing of its product. Surely a single entity league will be able to avoid some

---

\(^{156}\) See supra note 199 (describing the debate over the appropriate standard for determining abuse of monopoly power).


\(^{158}\) See Copperweld Corp. v. Independence Tube Corp., 467 U.S. 752, 773 (1984) ("[A] business enterprise should be free to structure itself in ways that serve efficiency of control, economy of operations, and other factors dictated by business judgment without increasing its exposure to antitrust liability.").

\(^{159}\) See id. at 772-73 (asserting that a corporation should be able to restructure and reorganize its operations in order to best serve the corporation's interests, without raising assumptions of illegitimate motives).

\(^{160}\) See supra note 107 (describing the structure and subsequent bankruptcy of the American Basketball League).
of the problems confronted by upstart leagues such as the North American Soccer League and the United States Football League. The single entity structure provides a solution to this problem and thus should make the creation of new sports ventures possible. As a general policy matter, the single entity form is therefore decidedly pro-competitive in its effects, and actually encourages the creation of the labor markets it may subsequently constrain.

The reorganization of the traditional leagues proposed in this Comment remains strictly a long term possibility. Of course, a favorable outcome for MLS in Fraser is necessary. Nevertheless, the leagues should begin exploring the possibility of reorganization. The continual threat of section 1 claims has hampered leagues' efforts to maximize their bargaining leverage with players' unions and exposed the league to protracted and burdensome litigation with players, owners, and other leagues. The removal of the section 1 threat and the potential that the single entity model holds for controlling player costs provide powerful incentives for leagues to follow MLS's lead and adopt single entity structures.

---

161 See Lebowitz, supra note 97, at 1F (contrasting MLS with other upstart leagues, including the NASL and the USFL).

162 See id. ("The NASL situation wasn't much different than the financial disparities between franchises that killed the United States Football League . . . ").
INSIDER TRADING: THE "POSSESSION VERSUS USE" DEBATE

KAREN SCHOEN

INTRODUCTION

Addressing a question that has long been a subject of debate among legal commentators, the Ninth and Eleventh Circuits recently held that "knowing possession" is not sufficient to impose insider trading liability under section 10(b) of the Securities Exchange Act of 1934 and Rule 10b-5 promulgated thereunder. At issue in this debate is whether section 10(b) and Rule 10b-5 require a causal connection between the material, nonpublic information and the insider’s trading—that is, a showing that the insider actually used the information—or whether knowing possession of such information at the time of the trade is sufficient for liability to attach.

Few courts have addressed this issue directly. Indeed, in many cases it has not been necessary for the courts to address the issue: often it is clear that the insider used the information. As Allan Horwich

† B.A. 1994, Yale University; J.D. Candidate 2000 and M.B.A. Candidate 2000, University of Pennsylvania. I would like to express my appreciation to Professor William C. Tyson for valuable comments on a previous draft, and to the editors of the University of Pennsylvania Law Review for all of their assistance.

1 See Allan Horwich, "Use Test" in Insider Trading Cases Update, CORP. LEGAL TIMES, Nov. 1998, at 61 ("In the past year two Courts of Appeal have resolved a question which had long been the subject of debate among commentators . . . ").

2 See United States v. Smith, 155 F.3d 1051, 1069 (9th Cir. 1998), cert. denied, 119 S. Ct. 804 (1999) ("[W]e reject the government's proffered 'knowing possession' standard for insider trading violations . . . "); SEC v. Adler, 137 F.3d 1325, 1337 (11th Cir. 1998) ("[M]ere knowing possession . . . is not a per se violation.").

3 See Adler, 137 F.3d at 1334 ("Surprisingly, few courts have directly addressed whether § 10(b) [and] Rule 10b-5 . . . require a causal connection between the material nonpublic information and the insider's trading or whether knowing possession of material nonpublic information while trading is sufficient for liability."). A few courts have addressed the issue indirectly or in dicta. See, e.g., United States v. Telcher, 987 F.2d 112, 120-21 (2d Cir. 1993) (advocating a knowing possession standard but asserting that it was unnecessary to rule on the issue); Investors Management Co., 44 S.E.C. 633, 644 (1971) (appearing to advocate a "use" standard).

4 See DONALD C. LANGEVOORT, INSIDER TRADING: REGULATION, ENFORCEMENT, AND PREVENTION § 3.04, at 3-22 (1999) ("In the typical case, there is no question that the insider traded in order to take advantage of material nondisclosed information."); see also Harvey L. Pitt & Karl A. Groskaufmanis, Actual Use of Inside Information at Issue,
has noted, however, "[t]he scarcity of pertinent cases does not... mean that the question is merely of abstract interest. ... [T]he issue frequently arises in counseling [corporate executives] who may obtain inside information pending completion of a transaction or in the midst of a pre-established trading program."5

This Comment will examine the "possession versus use" debate, assessing the arguments that have been made in support of the two sides and concluding that "knowing possession" is the more appropriate standard. Because one cannot develop a sensible rule without first understanding the rule's purpose, Part I will begin with an overview of insider trading, including the rationale underlying its prohibition. Part II will then present the "possession versus use" debate, exploring the origins of the debate, as well as the recent cases addressing the issue. Next, Part III will examine more closely the arguments advanced in support of the "use" standard. Advocates of the "use" standard have based their arguments largely on Supreme Court precedent. As discussed in Part III, however, this reliance on prior case law is misplaced because the question of "possession versus use" was not at issue in the cases relied upon. Rather than relying on prior case law, this Comment suggests that one should instead consider the reasons underlying the insider trading prohibition, as well as the difficulty of proof associated with prosecution, to determine the appropriate standard. These considerations form the basis for the arguments in support of the "knowing possession" standard. Part IV will discuss these arguments, concluding that "possession" is the more appropriate test for insider trading liability.

I. OVERVIEW OF INSIDER TRADING

"Insider trading" is a term of art that refers to unlawful trading in securities by persons who possess material nonpublic information about the company whose shares are traded or the market for its

---

5 Allan Horwich, Possession Versus Use: Is There a Causation Element in the Prohibition on Insider Trading?, 52 BUS. LAW. 1235, 1236 (1997). Horwich presents several interesting hypothetical situations in which the issue might arise. Consider, for example, an executive of a public company who, "nearing retirement, begins a regular program to liquidate her large holding of company stock in order to diversify her portfolio, instructing her broker to sell 1000 shares on the first of each month. During the program, the executive learns nonpublic negative material information about the company." Id. at 1235.
shares." Although the federal securities laws do not expressly proscribe insider trading, section 10(b) of the Securities Exchange Act of 1934\(^7\) and Rule 10b-5\(^8\) promulgated thereunder by the Securities and

---

\(^6\) Langevoot, supra note 4, § 1.01, at 1-6. This definition of insider trading, of course, presupposes the answer to the "possession versus use" debate. Compare Ralph C. Ferrara et al., Ferrara on Insider Trading and the Wall § 2.01[1], at 2-2 to 2-3 (1998) (explaining that insider trading occurs "when a corporate "insider" ... trades the securities of the corporation while in possession of material, nonpublic information" (emphasis added)), with 2 Alan R. Bromberg & Lewis D. Lowenfels, Bromberg & Lowenfels on Securities Fraud & Commodities Fraud § 7.4(100), at 7:101 (2d ed. 1998) ("[I]nsider trading is the use of ... material nonpublic information ... to buy securities (if the information is favorable) or sell them (if the information is unfavorable)." (emphasis added)). Unless quoting or describing the views of others, this Comment will employ "possession" language when discussing the nature of insider trading violations.

As many commentators have noted, the term "insider trading" is a misnomer. For example, liability for insider trading has not been limited to the class of persons traditionally considered "insiders." See Langevoot, supra note 4, § 1.01, at 1-6 ("The prohibition against insider trading applies to a larger class of persons than those traditionally considered corporate insiders."); Committee on Federal Regulation of Securities, Report of the Task Force on Regulation of Insider Trading, 41 Bus. Law. 223, 224 (1985) [hereinafter Task Force Report] ("The cases in which the Commission and private litigants have sought to impose liability under existing law have not been limited to insiders, at least as that term is commonly understood ...."); see also infra note 12 and accompanying text (explaining that the term "insider" refers to a broader class of persons than those traditionally considered corporate insiders). Furthermore, a person may be liable for insider trading even if he did not actually trade. See Task Force Report, supra, at 224 (explaining that "not all defendants have been traders—those who 'tip' others who then trade" have also been found guilty of insider trading); see also infra note 13 and accompanying text (discussing tipper liability).

\(^7\) Section 10(b) provides:

It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce or of the mails, or of any facility of any national securities exchange—

(b) To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors.


\(^8\) Rule 10b-5 provides:

It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce, or of the mails or of any facility of any national securities exchange,

(a) To employ any device, scheme, or artifice to defraud,

(b) To make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading, or

(c) To engage in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person,
Exchange Commission ("SEC") have been interpreted as prohibiting insider trading. In interpreting these provisions, the SEC and the

9 The original purpose of section 10(b) was not necessarily to prohibit insider trading. See Elizabeth Szocyj, The Law and Insider Trading: In Search of a Level Playing Field 5 (1993) ("Contrary to popular belief, Section 10(b) . . . which today is the most commonly used section to prosecute insider trading, was not originally designed to serve this purpose."); id. at 6 ("Nowhere in the legislation or in related discussions at the time is there an implication that Section 10(b) should be interpreted to include or could be applied against insider trading."). For a comprehensive discussion of the original purpose of section 10(b), see Steve Thel, The Original Conception of Section 10(b) of the Securities Exchange Act, 42 STAN. L. REV. 385 (1990). Nor was the central purpose of Rule 10b-5 to prohibit insider trading; indeed, the rule was initially drafted to prevent insiders from making fraudulent statements to shareholders about the corporation so that they could buy shares more cheaply. See Milton Freeman, Discussion at the Conference on Codification of the Federal Securities Laws (Nov. 18-19, 1966), in 22 BUS. LAW. 793, 922-23 (1967) (discussing the adoption of Rule 10b-5). In 1943, Milton Freeman, an SEC attorney, received news about a company president who was buying up shares of stock of his own company, simultaneously telling shareholders that the company was faltering, though in fact the prospects of the company were extremely favorable. See id. at 922 (describing a phone call in which Freeman learned that "the president of some company . . . is . . . buying up the stock of his company . . . and . . . telling [shareholders] that the company is doing very badly, whereas, in fact, the earnings are going to be quadrupled"). Attempting to address this problem, Freeman drafted Rule 10b-5. See id. (explaining that Rule 10b-5 "was intended to give the Commission power to deal with this problem").

Section 16(b), 15 U.S.C. § 78p(b) (1994), is the only provision that expressly procribes trading by insiders. See, e.g., Ferrara et al., supra note 6, § 1.02[1][a], at 1-3 ("The only provision of the securities statutes that expressly regulates insider trading is Section 16(b) of the Exchange Act, which prohibits 'short swing' profits by officers, directors, and the direct or indirect beneficial owners of more than 10% of any class of equity securities." (footnote omitted)). "Short-swing" trading refers to "any purchase and sale, or any sale and purchase" within a period of less than six months. 15 U.S.C. § 78p(b). Many have asserted that section 16 was designed to prevent insider trading. See 15 U.S.C. § 78p(b) (explaining that the purpose of the provision is to "prevent[ ] the unfair use of information which may have been obtained by such beneficial owner, director, or officer by reason of his relationship to the issuer"). But section 16 imposes liability on insiders who engage in short-swing trading, regardless of whether they were in possession of material, nonpublic information at the time of the trades. And insiders can avoid liability under section 16(b)—even if they are trading while in possession of material, nonpublic information—as long as they do not enter into offsetting transactions within a period of less than six months (i.e., a sale following a purchase or a purchase following a sale). Some commentators have thus proposed alternative explanations of the purpose of section 16(b). See Karl Shumpei Okamoto, Rereading Section 16(b) of the Securities Exchange Act, 27 GA. L. REV. 183, 185 (1992) ("The purpose of Section 16(b) is not the deterrence of trading on non-public information. . . . Its purpose is to deter insiders from trading to artificially affect market prices—to deter insiders from sending false signals when in fact there is no inside information at all."); Steve Thel, The Genius of Section 16: Regulating the Management of Publicly Held Companies, 42 HASTINGS L.J. 393, 453 (1991) (asserting that the purpose of section 16 was to prevent the manipulation of corporate opportunities).
federal courts have developed two distinct theories of liability: the "traditional" or "classical" theory, and the "misappropriation" theory of insider trading.

A. Theories of Liability: The Classical Theory and the Misappropriation Theory

Under the classical theory of insider trading, a corporate insider may not trade in the securities of his corporation while in possession of material, nonpublic information. Such trading violates section 10(b) and Rule 10b-5 because of the relationship of trust and confidence that exists between the shareholders of a corporation and corporate insiders; because of this relationship, "insiders who have obtained confidential information by reason of their position with that corporation" may not take "unfair advantage of . . . uninformed . . . stockholders." The term "insider" refers to a broader class of persons than those traditionally considered corporate insiders and includes "temporary insiders" such as attorneys, investment bankers, consultants, and accountants.

Liability, however, is not limited to those who actually trade; nor is it limited to insiders. An insider who "tips" material, nonpublic information violates section 10(b) and Rule 10b-5 if the tip breaches a fiduciary duty and the tippee (the recipient of such a tip) trades, even if the insider does not trade. The tippee may also be liable for trad-
ing if the tipper breached his fiduciary duty by conveying the information.\textsuperscript{14}

Liability under the classical theory requires a breach of an insider's duty owed to the shareholders of the corporation in whose securities the insider (or a tippee) traded.\textsuperscript{15} The classical theory thus does not impose liability where an "individual[] 'outside' a corporation purchase[s] the corporation's securities while in possession of material, nonpublic information that was obtained in a manner that did not involve an insider's breach of duty."\textsuperscript{16} Such trading has been described as "outsider trading," since the traders are outsiders of the corporation in whose securities they traded.\textsuperscript{17}

Recognizing this limitation of the classical theory, the SEC and the courts developed the misappropriation theory. The misappropriation theory imposes liability where an individual "misappropriates confidential information for securities trading purposes, in breach of a duty owed to the source of the information."\textsuperscript{18} Rather than predicated on a fiduciary relationship between the insider and the corporation's shareholders, "the misappropriation theory premises liability on a fiduciary-turned-trader's deception of those who entrusted him with access to confidential information."\textsuperscript{19} As in the case

---

\textsuperscript{14} See Dirks, 463 U.S. at 660 ("[A] tippee assumes a fiduciary duty to the shareholders of a corporation not to trade on material nonpublic information only when the insider has breached his fiduciary duty to the shareholders by disclosing the information to the tippee and the tippee knows or should know that there has been a breach.").

\textsuperscript{15} See Chiarella v. United States, 445 U.S. 222, 227-35 (1980) (explaining that there can be no fraud—and therefore no liability under section 10(b) and Rule 10b-5—absent a duty to disclose owed to the stockholders).

\textsuperscript{16} FERRARA ET AL., supra note 6, § 2.01[1], at 2-3.

\textsuperscript{17} See Jill E. Fisch, Start Making Sense: An Analysis and Proposal for Insider Trading Regulation, 26 GA. L. REV. 179, 198 (1991) ("Trading of this type is frequently described as 'outsider trading,' so denominated because the traders are outsiders with respect to the issuer of the securities they trade.").


\textsuperscript{19} Id. A good example of the misappropriation theory is found in O'Hagan. In that case the defendant O'Hagan was a partner in a law firm representing Grand Met in its planned tender offer of Pillsbury (though O'Hagan himself did not do any work for Grand Met). Aware of the planned tender offer, O'Hagan traded in Pillsbury options and common stock, reaping a profit of more than $4.3 million. See id. at 647-48 (describing O'Hagan's conduct).
of the classical theory, both tippers and tippees may be liable for insider trading under the misappropriation theory.20

B. Reasons for the Insider Trading Prohibition

The traditional basis for the insider trading prohibition is the belief that insider trading violates our notion of "fair play" and threatens the integrity of the capital markets.21 As the American Bar Association

20 See FERRARA ET AL., supra note 6, § 2.01[1], at 2-4 ("On similar reasoning, trading based on a 'misappropriated' tip generally triggers liability for both the tippee and the tipper.").

21 See Task Force Report, supra note 6, at 227 (asserting that the "two ... traditional bases for prohibitions against insider trading" are the "'fair play' and 'integrity of the markets' arguments"). Commentators have offered several additional arguments for the prohibitions against insider trading. See generally LOUIS LOSS & JOEL SELIGMAN, FUNDAMENTALS OF SECURITIES REGULATION 762 (3d ed. 1995) (asserting that because insider-traders can profit on negative information, insider trading may create incentives for insiders to act against the best interests of the corporation); Task Force Report, supra note 6, at 228 (asserting that insider trading discourages prompt disclosure of information since corporate insiders have an "incentive to delay [disclosure] long enough to speculate on a stock market profit"); Nicholas L. Georgakopoulos, Insider Trading as a Transactional Cost: A Market Microstructure Justification and Optimization of Insider Trading Regulation, 26 CONN. L. REV. 1, 17-46 (1993) (advocating the prohibition of insider trading by arguing that the prohibition reduces transaction costs); Roy A. Schotland, Unsafe at Any Price: A Reply to Manne, Insider Trading and the Stock Market, 53 VA. L. REV. 1425, 1452 (1967) (arguing, among other things, that insider trading subjects insiders to "a conflict of interest that may affect their judgment not only in the timing of disclosure, but also in the timing of the underlying events themselves [and] may cause such events to be created"); Bevis Longstreth, Halting Insider Trading, N.Y. TIMES, Apr. 12, 1984, at A27 (summarizing the arguments against insider trading and rebutting the arguments in favor of insider trading).

Some commentators, on the other hand, have argued that insider trading should not be prohibited. See generally HENRY G. MANNE, INSIDER TRADING AND THE STOCK MARKET (1966) (attacking the wisdom of the prohibition against insider trading and asserting that insider trading is a useful means of compensating entrepreneurs); Dennis W. Carlton & Daniel R. Fischel, The Regulation of Insider Trading, 35 STAN. L. REV. 857, 868 (1983) (arguing that insider trading promotes market efficiency since such trading is a means of communicating information to the market and thus causes the stock price to move to its "correct" value sooner); Michael P. Dooley, Enforcement of Insider Trading Restrictions, 66 VA. L. REV. 1, 55 (1980) (concluding that "the harm caused by insider trading can be objectively measured and that so measured it does not cause any detectable injury to investors"); Henry G. Manne, Insider Trading and the Law Professors, 23 VAND. L. REV. 547 (1970) (responding to critics of his arguments in support of insider trading).

The arguments that insider trading should be permitted, however, have gained little support in Congress. As Congressman Dingell asserted: "We strongly disagree with those who would seek to permeate our markets with this type of fraud. The arguments in favor of insider trading are rubbish." 130 CONG. REC. 20,969 (1984). And as Senator D'Amato stated: "Some commentators have called insider trading a victimless crime; however, I strongly disagree.... John Fedders, the Director of the Securi-
Task Force on Regulation of Insider Trading observed:

In our society, we traditionally abhor those who refuse to play by the rules, that is, the cheaters and the sneaks. A spitball pitcher, or a card shark with an ace up his sleeve, may win the game but not our respect. And if we know such a person is in the game, chances are we won’t play. These commonsense observations suggest that two of the traditional bases for prohibitions against insider trading are still sound: the ‘fair play’ and ‘integrity of the markets’ arguments.

Underlying the development of the insider trading prohibition is the view that insider trading is fundamentally unfair. As Donald Langevoort has observed, “most people who oppose insider trading seem to believe that quite apart from any harm caused to specific investors, insider trading is simply an unfair exploitation of information that properly belongs to someone else.” The unfairness stems not from the insider-trader’s possession of superior information, but from the trader’s possession of an informational advantage that other investors cannot overcome: other investors cannot acquire the inside information. Clearly, informational advantages exist; some investors will necessarily have superior information as a result of superior intelligence, diligence, or power. And not all such informational disparities and Exchange Commission’s Division of Enforcement, has stated publicly that he believes that those who engage in insider trading are thieves, I concur wholeheartedly with Mr. Fedders.”

An assessment of the various arguments is beyond the scope of this Comment. Congress and the courts have embraced the “fairness” and “integrity of the markets” arguments, and this Comment will proceed under the assumption that such arguments are valid.

See Cady, Roberts & Co., 40 S.E.C. 907, 911-12 (1961) (asserting that the “disclose or abstain” obligation “rests on...the inherent unfairness involved where a party takes advantage of...information knowing it is unavailable to those with whom he is dealing”).

LANGEVOORT, supra note 4, § 1.02[4], at 1-14.

As Victor Brudney commented:

The unfairness is not a function merely of possessing more information -- outsiders may possess more information than other outsiders by reason of their diligence or zeal -- but of the fact that it is an advantage which cannot be competed away since it depends upon a lawful privilege to which an outsider cannot acquire access.


See id. at 360 (“[T]here may nevertheless be systematic inequality of lawful access to information by reason of disparities among individual investors with respect to power, wealth, diligence, or intelligence.”); see also Ronald J. Gilson, The Outside View of Inside Trading, N.Y. TIMES, Feb. 8, 1987, at E23 (“In the public mind, fairness is equated with access. From this perspective, what is unfair about insider trading is that...
ties should be eliminated. When other investors are unable to obtain information, however, no matter how great their resources or diligence, such informational advantages are unfair, and it is those informational advantages that the insider trading prohibition seeks to prevent.

This is the "equality of access to information" theory. See id. at 354-55 ("[This notion is sometimes cast in terms of a rule seeking to effect equal access to material information for persons trading with each other."). It should be distinguished from the "parity-of-information" theory, which suggests that all informational disparities should be eliminated. See id. at 355 ("[V]iewed most broadly, [the equality of access to information theory] does not extend so far as to require actual equality or sharing of information."). As Justice Blackmun explained:

[T]here is a significant conceptual distinction between parity of information and parity of access to material information. The latter gives free rein to certain kinds of informational advantages that the former might foreclose, such as those that result from differences in diligence or acumen. Indeed, by limiting opportunities for profit from manipulation of confidential connections or resort to stealth, equal access helps to ensure that advantages obtained by honest means reap their full reward.

Chiarella v. United States, 445 U.S. 222, 252 n.2 (1980) (Blackmun, J., dissenting); see also Insider Trading: Some Questions and Some Answers, 1 SEC. REG. L.J. 328, 335 (1974) (suggesting that it is "equality of access" to information (as distinguished from "equality of information") among investors that is a goal of the antifraud provisions).

At least one commentator has argued that the insider trading laws should adopt a parity-of-information approach. See Joel Seligman, The Reformulation of Federal Securities Law Concerning Nonpublic Information, 11 J. L. & ECON. 7 (1974) (arguing that a parity-of-information rule "has clear advantages" including the fact that it is a "bright line rule" that would "reassure investors that the securities markets generally prohibit information advantages").

The Supreme Court has explicitly rejected the parity-of-information theory. See Chiarella, 445 U.S. at 233 ("[N]either the Congress nor the Commission ever has adopted a parity-of-information rule."). Chiarella also cast doubt upon the equal access theory, predicating liability on a fiduciary duty to shareholders. See id. at 231-33 (arguing that a duty to "disclose or abstain" exists only where the insider has a fiduciary duty to shareholders). The Supreme Court also appeared to reject the equal access theory in United States v. O'Hagan, asserting that a fiduciary could potentially avoid liability by disclosing to the source of the information his plans to trade. See United States v. O'Hagan, 521 U.S. 642, 655 (1997) ("[F]ull disclosure forecloses liability under the
In addition to—or perhaps because of—this fundamental unfairness, insider trading threatens the capital markets. The strength and stability of the capital markets depend on investor confidence in those markets. Insider trading, however, undermines investors' expectations of honest and fair securities markets. If investors believe that other market participants have an unerodable informational advantage—that the markets are not fair—they will be reluctant to participate in the securities markets:

A rational buyer (or seller) in a market, who knows that the person with whom he is dealing has material information about the value of the product being exchanged which he could not lawfully acquire, will either refrain from dealing with that transactor or demand a risk premium. If the market is thought to be systematically populated with such transactors some investors will refrain from dealing altogether, and others will incur costs to avoid dealing with such transactors or corruptly to overcome their unerodable informational advantages.

misappropriation theory... if the fiduciary discloses to the source that he plans to trade on the nonpublic information, there is no... § 10(b) violation."). This assertion implicitly rejects the equal access theory, for even if the fiduciary disclosed his trading plans to the source of the information, he would still have an informational advantage that could not lawfully be overcome by other investors. But see Chiarella, 445 U.S. at 251 (Blackmun, J., dissenting) (endorsing the equal access theory, asserting that "I would hold that persons having access to confidential material information that is not legally available to others generally are prohibited by Rule 10b-5 from engaging in schemes to exploit their structural informational advantage through trading in affected securities").


See id. ("Insider trading threatens [the capital] markets by undermining the public's expectations of honest and fair securities markets where all participants play by the same rules."); see also LANGEVOORT, supra note 4, § 1.02[4], at 1-15 ("Confidence in the securities marketplace... is diminished if shareholders believe that only those 'in the know' can profit in the stock market.").

Brudney, supra note 25, at 356; see also Insider Trading Sanctions and SEC Enforcement Legislation, Hearing on H.R. 559 Before the Subcomm. on Telecomm., Consumer Protection and Finance of the House Comm. on Energy and Commerce, 98th Cong. 35 (1983) [hereinafter House Hearings] (statement of John M. Fedders, SEC Director of Enforcement) ("[I]f we permit [insider trading] I think it would erode investor confidence in the marketplace and they may very well flee the marketplace and choose other investments, which would hurt the liquidity of the exchanges, the liquidity of the marketplace, and eventually the capital structure of our country."); H.R. REP. NO. 100-910, at 8 (1988), reprinted in 1988 U.S.C.C.A.N. 6043, 6045 ("[T]he small investor will be—and has been—reluctant to invest in the market if he feels it is rigged against him."); Task Force Report, supra note 6, at 227-28 ("Although [we] know[] of no empirical research that directly demonstrates that concerns about integrity affect market activity, both authoritative commentators and common sense tell us that if investors do not anticipate fair treatment, they will avoid investing in securities [and] capital forma-
It is this recognition that underlies the insider trading prohibitions.33 Indeed, as SEC Enforcement Director William R. McLucas asserted, "one of the purposes of bringing an insider trading case, in addition to addressing specific instances of fraud, is to ensure the integrity of the market."34

II. THE "POSSESSION VERSUS USE" DEBATE

Although Congress has condemned insider trading and expressed strong support for the prohibition of such conduct,35 Congress has resisted attempts to expressly define insider trading.36 Instead, the prohibition against insider trading has developed solely from judicial (and administrative) interpretation of the anti-fraud provisions of section 10(b) and Rule 10b-5. The lack of a statutory definition of the insider trading offense has led to some ambiguity regarding the type of conduct that is prohibited.37 One such area of ambiguity is whether the insider’s motivation for trading is relevant to a determination of insider-trading liability under Rule 10b-5. At issue is whether section 10(b) and Rule 10b-5 require a causal connection between the material, nonpublic information and the insider’s trading—that is, a show-
ing that the insider actually used the information in making his decision to trade—or whether knowing possession of such information at the time of the trade is sufficient for liability. The SEC, as well as some legal commentators, have advocated a "knowing possession" standard, arguing that the insider's motivation for trading is largely irrelevant. Other commentators, however, have advocated a "use"

---

58 Although organizations can be subject to insider trading liability under the Insider Trading and Securities Fraud Enforcement Act of 1988 ("ITSEFA"), Pub. L. No. 100-704, 102 Stat. 4677 (codified as amended in scattered sections of 15 U.S.C.), see, e.g., H.R. Rep. No. 100-910, at 7 (1988), reprinted in 1988 U.S.C.C.A.N. 6043, 6044 (explaining that the bill "expands the scope of civil penalties remedies to 'controlling persons' who fail to take adequate steps to prevent insider trading"), the "possession versus use" debate has focused primarily on trading by individuals, rather than by organizations. Although one might argue that the debate should include corporate trading, a discussion of corporate trading is beyond the scope of this Comment.

Organizations frequently trade while in possession of material, nonpublic information—often one department of an organization will trade while another department possesses inside information about the traded securities. See John H. Sturc & Catharine W. Cummer, Possession vs. Use for Insider Trading Liability, INSIGHTS, June 1998, at 3, 7 (explaining that a "common situation is that of an institutional investor with separate departments, one of which may possess material nonpublic information and another which engages in trading"). But such organizations may avoid liability—even under a "knowing possession" standard—by erecting a "Chinese Wall," which is essentially a set of policies and procedures designed to prevent the communication of information between departments. See Marc I. Steinberg & John Fletcher, Compliance Programs for Insider Trading; 47 SMU L. REV. 1783, 1803-04 (1994) ("Chinese Wall procedures consist of policies and procedures designed to control the flow of material, nonpublic information within a multiservice financial firm."). If a Chinese Wall is in place, inside information possessed by individuals on one side of the wall will not be imputed to those on the other side of the wall who are engaged in trading activities. See Horwich, supra note 5, at 1238 ("This [lack of imputation] is understood to free the organization to trade even if trading 'while in possession' of inside information is unlawful."). Although few courts have addressed whether a Chinese Wall can be used by an organization to avoid federal securities liability, the SEC explicitly approved the use of Chinese Walls when it adopted Rule 14e-3. See 17 C.F.R. § 240.14e-3(b) (1999) (providing a safe harbor for organizations implementing Chinese Wall procedures provided it can be shown that the "individual(s) making the investment decision on behalf of the [organization] . . . did not know the material, nonpublic information"); Steinberg & Fletcher, supra, at 1810-12 (discussing judicial and administrative treatment of Chinese Walls).

59 See SEC v. Adler, 137 F.3d 1325, 1332-33 (11th Cir. 1998) (noting that the SEC claimed that "knowing possession of material, nonpublic information while trading is sufficient to establish liability" and that the SEC argued that evidence that the defendant knowingly possessed material, nonpublic information at the time of his trade was sufficient to impose insider trading liability, regardless of whether or not the defendant used the inside information); FERRARA ET AL., supra note 6, § 2.01[5], at 2-14.6 to 2-14.8 (implicitly endorsing the "knowing possession" standard—at least under the classical theory of insider trading liability); LANGEVOORT, supra note 4, § 3.04, at 3-23 to 3-27 (suggesting that "knowing possession" is the appropriate test); 7 LOUIS LOSS & JOEL SELIGMAN, SECURITIES REGULATION 3505 (3d ed. 1991) (advocating the "knowing
standard, arguing that a causal connection between the inside information and the insider’s trading activity is required.40

In many insider trading cases, the difference between these two standards is irrelevant, since it is clear that the insider traded on the basis of the material, nonpublic information.41 The issue can become important, however, in certain situations. Consider, for example, the following situation:

The chief financial officer ["CFO") of a publicly held company has consistently bought shares (in the range of 2,000 to 5,000 shares) in his company during a two- or three-day window period beginning the day after the filing of the company’s quarterly report on Form 10-Q. Based on the advice of house counsel, this timing was intended to protect the CFO from any charges that he possessed non-public information. But in the instant case, the day after the filing of its Form 10-Q, the company (and its senior officers) become aware of a lucrative merger proposal from a well-financed acquirer. Knowing that such a merger was under negotiation, the CFO still buys his more-or-less customary 5,000 shares (and thereby profits handsomely when the nonpublic proposal is eventually announced).42

In this situation, the CFO’s liability for insider trading may depend upon whether the "knowing possession" or the "use" standard is employed. Under the SEC’s "knowing possession" test, the CFO will be liable for insider trading—regardless of the fact that he always buys shares after the company files its quarterly report and regardless of his motivation for his current purchase of shares. Under the "use" test, however, the CFO will escape liability as long as his decision to buy shares was not based on his knowledge of the contemplated merger—

---


41 See LANGEVOORT, supra note 4, § 3.04, at 3-22 (“In the typical case, there is no question that the insider traded in order to take advantage of material nondisclosed information.”).

as long as this information was not a factor in his decision to purchase the shares. In this context, then, the "possession versus use" debate may be quite important.

Attempting to resolve this debate, the Ninth and Eleventh Circuits recently held that "knowing possession" is not sufficient to impose insider trading liability.\(^4\) This Part will discuss the "possession versus use" debate, exploring both the origins of the debate and the recent cases addressing the issue.

A. Origins of the Debate

Federal insider trading case law can be traced back to the SEC's seminal opinion in Cady, Roberts & Co., in which the SEC concluded that insider trading constitutes "fraud or deceit" and thus violates Rule 10b-5.\(^4\) In so doing, the SEC articulated the oft-cited "disclose or ab-
stain rule," asserting that an insider is required to disclose material, nonpublic information or abstain from trading. The Second Circuit embraced this rule in SEC v. Texas Gulf Sulphur Co., holding that "anyone in possession of material inside information must either disclose it to the investing public, or if he is disabled from disclosing it... or he chooses not to do so, must abstain from trading in or recommending the securities concerned while such inside information remains undisclosed."

Courts and commentators, however, have failed to reach a consensus on the correct interpretation of the "disclose or abstain" rule. To some, the rule suggests that mere possession of nonpublic information is sufficient to establish insider trading liability, and the rule has in fact been cited for such a proposition. Others have insisted that the maxim "disclose or abstain" fails to "capture the complexity of the idea that [it is] supposed to represent," and have argued that it "enjoins not all trading, but [only] trading on the basis of material nonpublic information."
Even the SEC appears to have wavered in its interpretation of the "disclose or abstain" rule. Despite the assertion by SEC Enforcement Director William R. McLucas that the Commission has always advocated "knowing possession" as the appropriate rule for insider trading liability,\(^49\) the SEC has been inconsistent in its interpretation of the rule. In *Investors Management Co.*, for example, the SEC effectively endorsed a "use" requirement, asserting that a requisite element for insider trading liability is "that the information be a factor in [the] decision to effect the transaction."\(^50\) The SEC continued to explain: "[W]here a transaction... is effected by the recipient [of material, nonpublic information] prior to its public dissemination, an inference arises that the information was... a factor. The recipient of course may seek to overcome such inference by countervailing evidence."\(^51\) But in *Sterling Drug, Inc.*, the SEC rejected the "use" requirement suggested by *Investors Management* and instead endorsed the "knowing possession" standard, asserting:

The Commission... believes that Rule 10b-5... does not require a showing that an insider sold his securities for the purpose of taking advantage of material non-public information.... If an insider sells his securities while in possession of material adverse non-public information, such an insider is taking advantage of his position to the detriment of...
the public.\footnote{Sterling Drug, Inc., 14 S.E.C. Docket 824, 827 (1978). In \textit{Sterling Drug}, three directors accused of insider trading maintained that they had decided to sell stock in the company prior to obtaining the inside information. One director presented documentary evidence that he had formed an intention to sell prior to obtaining the information and testified that he sold stock because he had retired from his position and wished to diversify his investments. The other two directors testified that they sold stock to satisfy various obligations. See id. at 827 (discussing the directors' testimony regarding their reasons for selling stock).}

In neither of these opinions, however, did the SEC articulate a rationale for its position.\footnote{The SEC did not cite any authority for its assertion in \textit{Sterling Drug}, nor did the SEC reference its opinion in \textit{Investors Management}.} The SEC's seemingly abrupt and unexplained shift in position has served only to exacerbate the ambiguity and has made some reluctant to accord much deference to the Commission's position that "knowing possession" is the appropriate test.\footnote{See infra notes 82-85 and accompanying text (explaining that the Eleventh Circuit has declined to accord much deference to the SEC's position regarding the "possession versus use" debate since the SEC's position has not been consistent over time).}

\section*{B. Recent Cases Addressing the "Possession Versus Use" Debate}

More recently, three federal circuit courts have expressly examined the issue of "possession versus use."

\subsection*{1. The Possession Standard: \textit{United States v. Teicher}}

The first case to discuss carefully the "possession versus use" debate was \textit{United States v. Teicher}.\footnote{\textit{Teicher}, 987 F.2d at 119 (discussing the defendants' arguments). The district court instructed the jury:}

The government need not prove a causal relationship between the misappropriated material nonpublic information and the defendants' trading. That is, the government need not prove that the defendants purchased or sold securities because of the material nonpublic information that they knowingly possessed. It is sufficient if the government proves that the defendants purchased or sold securities while knowingly in possession of the material nonpublic information.

See \textit{infra} notes 82-85 and accompanying text (explaining that the Eleventh Circuit has declined to accord much deference to the SEC's position regarding the "possession versus use" debate since the SEC's position has not been consistent over time).

In \textit{Teicher}, the defendants argued that the district court erroneously instructed the jury that the defendants could be found guilty of insider trading if they possessed material, nonpublic information at the time of their trades, regardless of whether this information was the actual cause of those trades.\footnote{987 F.2d 112 (2d Cir. 1993). See \textit{LANGEVOORT}, supra note 4, § 3.04, at 3-23 ("The first extensive judicial discussion of this issue is found in... \textit{United States v. Teicher}.") The case has also been referred to as the "leading case" in support of the "knowing possession" standard. Horwich, \textit{supra} note 5, at 1250.}

\textbf{THE "POSSSESSION VERSUS USE" DEBATE}
though merely dicta, the Second Circuit expressly rejected the "use" requirement advocated by the defendants and instead endorsed a "knowing possession" standard.\textsuperscript{57} The court began by observing that "Rule 14e-3, which prohibits fraud in connection with tender offers explicitly provides a 'knowing possession' standard."\textsuperscript{58} The court then set forth three factors in support of its position.\textsuperscript{59}

First, the \textit{Teicher} court stated that section 10(b) and Rule 10b-5 require only that a deceptive practice be conducted "in connection with the purchase or sale of a security."\textsuperscript{60} The court noted that the phrase "in connection with" has been construed quite flexibly, suggesting that such an interpretation supports the more flexible and less restric-

\texttt{Id.}

\textsuperscript{57} See \textit{id.} at 120-21 (rejecting the defendants' arguments and asserting that "a number of factors weigh in favor of a 'knowing possession' standard"). The court's adoption of the "knowing possession" standard was dicta since the court found it "unnecessary to determine whether proof of securities fraud requires a causal connection, because any alleged defect in the instruction was harmless beyond doubt." \textit{Id.} at 121.

\textsuperscript{58} \textit{Id.} at 120. Rule 14e-3 provides in relevant part:

(a) If any person has taken a substantial step or steps to commence . . . a tender offer, . . . it shall constitute a fraudulent, deceptive or manipulative act or practice within the meaning of section 14(e) of the Act for any other person who in possession of material information relating to such tender offer which information he knows or has reason to know is nonpublic and which he knows or has reason to know has been acquired directly or indirectly from:

(1) The offering person,

(2) The issuer of the securities sought . . . by such tender offer, or

(3) Any officer, director, partner or employee or any other person acting on behalf of the offering person or such issuer, to purchase or sell . . . such securities . . . unless within a reasonable time prior to any purchase or sale such information and its source are publicly disclosed by press release or otherwise.


The applicability of Rule 14e-3's "knowing possession" standard to the "possession versus use" debate is somewhat questionable. In section 14(e), Congress expressly gave the SEC the power to define fraud: "The Commission shall, for the purposes of this subsection, by rules and regulations define, and prescribe means reasonably designed to prevent, such acts and practices as are fraudulent, deceptive, or manipulative." 15 U.S.C. \S 78n(e) (1998). The SEC's definition of fraud contained in Rule 14e-3 may not be transferable to section 10(b) and Rule 10b-5 fraud.\textsuperscript{59}

\textsuperscript{59} See \textit{Teicher}, 987 F.2d at 120-21. In addition to the factors discussed, the court also asserted that the SEC's interpretation of Rule 10b-5 is "entitled to some consideration." \textit{Id.} at 120. Citing \textit{Sterling Drug, Inc.}, the court noted that the SEC has "consistently endorsed [the view] that a violation of \S 10(b) and Rule 10b-5 occurs when a trade is conducted in 'knowing possession' of material nonpublic information obtained in breach of a fiduciary or similar duty." \textit{Id.} For a discussion of \textit{Sterling Drug, Inc.}, see \textit{supra} note 52 and accompanying text.

\textsuperscript{60} \textit{Teicher}, 987 F.2d at 120 (quoting section 10(b) and Rule 10b-5).
tive "knowing possession" standard.61

Second, the court argued that the "disclose or abstain" rule supports the "knowing possession" standard.62 The court explained that if an insider is in possession of material, nonpublic information and he chooses not to disclose such information, he must abstain from trading.63

Finally, and perhaps most persuasively, the court recommended the "knowing possession" standard because it is simple to apply, requiring only a determination of whether the trader possessed material, nonpublic information.64 The "use" standard, on the other hand, requires factual inquiries into the state of mind and motivations of the trader.65 The court further suggested that it would be unlikely that information possessed by a trader would not in some way influence his trades. As the court explained:

Unlike a loaded weapon which may stand ready but unused, material information can not lay idle in the human brain. The individual with such information may decide to trade upon that information, to alter a previously decided-upon transaction, to continue with a previously planned transaction even though publicly available information would now suggest otherwise, or simply to do nothing.66

The court believed that a "use" standard would pose difficulties of proof for the SEC or the government and would "frustrate attempts to distinguish between legitimate trades and those conducted in connection with inside information."67

61 See id. at 120 ("We have previously stated that the 'in connection with' clause must be 'construed...flexibly to include deceptive practices 'touching' the sale of securities, a relationship which has been described as 'very tenuous......'(quoting United States v. Newman, 664 F.2d 12, 18 (2d Cir. 1981)); see also LANGEVOORT, supra note 4, § 3.04, at 3-23 (explaining that the Teicher court believed that "the possession standard better comports with the view that Rule 10b-5 should be read flexibly, not restrictively").

62 See Teicher, 987 F.2d at 120 ("[A] 'knowing possession' standard comports with the oft-quoted maxim that one with a fiduciary or similar duty to hold material non-public information in confidence must either 'disclose or abstain' with regard to trading.").

63 See id. ("When the...insider...is not in a position to make a public announcement, [he] must abstain.").

64 See id. ("Finally, a 'knowing possession' standard has the attribute of simplicity.").

65 See FERRARA ET AL., supra note 6, § 2.01[5], at 2-14.6 (explaining that the Teicher court believed that "a 'use' test would entail significant factual inquiries into the state of mind and the motivations of the insider trader").

66 Teicher, 987 F.2d at 120.

67 Id. at 121.
2. The Use Standard: *SEC v. Adler* and *United States v. Smith*

Although *Teicher* devoted considerable attention to the "possession versus use" debate, the Second Circuit's discussion was merely dicta.68 The first circuit court to directly confront the issue was the Eleventh Circuit in *SEC v. Adler*.69 In that case, the district court effectively adopted a "use" standard for insider trading, granting summary judgment for one of the defendants because he had a preexisting plan to sell stock.70 On appeal, the SEC argued that the district court erred in granting summary judgment, asserting that the defendant knowingly possessed material, nonpublic information and that such knowing possession was sufficient for liability.71 Viewing the choice between the "possession" and "use" standards as "a difficult and close question of first impression," the Eleventh Circuit concluded that "the use test is the appropriate test."72 The court asserted, however, that an "inference of use . . . arises from the fact that an insider traded while in possession of inside information."73

Acknowledging that the Supreme Court's language in several cases was merely dicta, the *Adler* court nonetheless felt that its "use" test "best comports with the language of § 10(b) and Rule 10b-5, and with Supreme Court precedent."74 The court noted, for example, the "use" language employed throughout the Supreme Court's opinions in *Chiarella v. United States*,75 *Dirks v. SEC*,76 and *United States v. O'Hagan*.77 The court also asserted that the Supreme Court's language in those opinions "repeatedly emphasized [a] focus on fraud and deception,"78 and argued that a "knowing possession" test fails to embody such a focus: "[W]e do not believe that the SEC's knowing possession test would always and inevitably be limited to situations involving

---

68 See supra note 57 (explaining why the court's discussion was dicta).
69 137 F.3d 1325 (11th Cir. 1998).
70 See id. at 1332 (explaining that the district court granted summary judgment for one of the defendants because his "preexisting plan to sell stock rebutted any reasonable inference of scienter").
71 See id. ("The SEC argues that it presented evidence that [the defendant] knowingly possessed material nonpublic information, and thus . . . violated the prohibition against insider trading . . . .").
72 Id. at 1337.
73 Id.
74 Id. at 1338.
78 *Adler*, 137 F.3d at 1338.
fraud."\(^{79}\)

The Adler court found support for its position in other cases as well. Citing cases including *In re Worlds of Wonder Securities Litigation*,\(^ {80}\) the court noted that several other courts "have allowed insiders to introduce evidence of preexisting plans or other 'innocuous' reasons for sales in order to" escape liability.\(^ {81}\)

Furthermore, although the SEC advocated a "knowing possession" standard, the court refused to defer to the SEC's position.\(^ {82}\) The Adler court asserted that the SEC position had not been consistent, noting that in *Investors Management*\(^ {83}\) the SEC appeared to support a "use" test, but that in *Sterling Drug*\(^ {84}\) the SEC adopted a "knowing possession" test without providing any rationale for the change.\(^ {85}\) The court also observed that although the SEC had expressly adopted a "knowing possession" standard in the context of tender offers with Rule 14e-3,\(^ {86}\) the SEC had not amended Rule 10b-5 or promulgated other rules adopting the "knowing possession" test for the more general case of insider trading.\(^ {87}\)

Finally, the court considered the Insider Trading Sanctions Act of 1984. Although acknowledging that Congress used the phrase "in possession of" in the statute, the court nonetheless concluded that the

---

\(^{79}\) Id.

\(^{80}\) 35 F.3d 1407 (9th Cir. 1994) [hereinafter Worlds of Wonder].

\(^{81}\) Adler, 137 F.3d at 1335.

\(^{82}\) See id. at 1339 ("We decline to accord much deference to the SEC position ... ").

\(^{83}\) 44 S.E.C. 633 (1971).

\(^{84}\) 14 S.E.C. Docket 824 (1978).

\(^{85}\) See Adler, 137 F.3d at 1339 (explaining that the SEC's position regarding the "possession versus use" debate has not been consistent over time); see also supra notes 50-52 and accompanying text (discussing the SEC's positions in *Investors Management* and *Sterling Drug*).

\(^{86}\) Rule 14e-3 prohibits trading in securities while in possession of material, non-public information, if such securities are being sought (or will be sought) in a tender offer. See 17 C.F.R. § 240.14e-3 (1999) (providing that "it shall constitute a fraudulent, deceptive or manipulative act or practice ... for [a] person who is in possession of material [nonpublic] information relating to [a] tender offer ... to purchase or sell ... such securities"); supra note 58 and accompanying text (discussing Rule 14e-3).

\(^{87}\) See Adler, 137 F.3d at 1339 ("[A]lthough the SEC has adopted such a rule in the context of tender offers, it has not formally adopted the knowing possession test for insider trading."). Some have questioned the SEC's ability to promulgate such a rule under section 10(b). See Sturc & Cummer, supra note 38, at 6 (asserting that the Adler court's "assumption about the SEC's rulemaking authority is questionable," because "Section 10(b) does not give the SEC the same authority to make rules 'reasonably designed to prevent' any acts and practices which may violate the statute as does Section 14(e), but rather only the authority to proscribe the deceptive devices themselves").
"language of [the Insider Trading Sanctions Act] does not resolve whether possession or use is the proper standard for an insider trading violation."

The court noted that "[t]he strongest argument that has been articulated in support of the knowing possession test is that a strict use test would pose serious difficulties of proof for the SEC." Agreeing that a trader's motivations are often difficult to prove, the court held that an "inference of use... arises from the fact that the insider traded while in knowing possession of material nonpublic information." The_ Adler_ court asserted that such an inference would be sufficient to alleviate the problems of proof.

Shortly after the Eleventh Circuit addressed the "possession versus use" debate, the Ninth Circuit addressed the issue in _United States v. Smith._ The Smith court agreed with much of the reasoning in _Adler_ and adopted the "use" standard. Because Smith involved a criminal prosecution, however, the court stopped short of Adler's holding that an inference of use arises when the insider trades while in possession of material, nonpublic information: "[W]e deal here with a criminal prosecution, not a civil enforcement proceeding, as was the situation in Adler. We are therefore not at liberty, as was the Adler court, to establish an evidentiary presumption that gives rise to an inference of use."

---

83 _Adler,_ 137 F.3d at 1337; _see also_ infra Part II.C (discussing the Insider Trading Sanctions Act and Congress's role in the "possession versus use" debate).

89 _Adler,_ 137 F.3d at 1337.

90 _Id.;_ _see also id._ ("[W]hen an insider trades while in possession of material non-public information, a strong inference arises that such information was used by the insider in trading. The insider can attempt to rebut the inference by adducing evidence that there was no causal connection between the information and the trade ....").

91 _Id._ ("[W]e believe that the SEC's problems [with respect to proof] are sufficiently alleviated by the inference of use ...").

95 _Id._ at 1068 (agreeing with the _Adler_ court that "a 'use' requirement [is] more consistent with the language of § 10(b) and Rule 10b-5, which emphasizes 'manipulation,' 'deception,' and 'fraud'"; _id._ ("Like our colleagues on the Eleventh Circuit, we are concerned that the SEC's 'knowing possession' standard would not be—indeed, could not be—strictly limited to those situations actually involving intentional fraud.").

96 _See id._ at 1069 (holding that Rule 10b-5 requires the government to "demonstrate that the suspected inside[r] trader actually used material nonpublic information in consummating his transaction").

97 _Id._
C. Congressional Role in the Debate

Congress also appeared to weigh in on the "possession versus use" debate when it enacted the Insider Trading Sanctions Act of 1984 ("ITSA"). The major purpose of the ITSA was to increase the sanctions for insider trading by creating a new civil penalty. In creating the new civil penalty, however, Congress explicitly adopted a "possession" standard: the SEC may seek the civil penalty from any person who violates a rule or statutory provision by "purchasing or selling a security while in possession of material, nonpublic information."

When it adopted the ITSA, Congress expressly declined to define insider trading. Nonetheless, during the hearings on the proposed legislation, Congress examined the "possession versus use" debate, entertaining testimony from various witnesses regarding the appropriate standard. Several witnesses urged that Congress employ "use" language, arguing that the "possession" standard was inappropriate and

---


97 See H.R. REP. No. 98-355, at 1 (1984), reprinted in 1984 U.S.C.C.A.N. 2274, 2274 (describing the purpose of the ITSA). Prior to the enactment of the ITSA, the only remedies available to the SEC for insider trading violations were injunctions against future violations and disgorgement orders, requiring defendants to pay the profits gained (or losses avoided). See Russell B. Stevenson, Jr., The Insider Trading Sanctions Act: Some Unfinished Business Ahead, NAT'L J., Oct. 15, 1984, at 18 (discussing the remedies available to the SEC prior to the passage of the ITSA). The ITSA created additional remedies, giving the SEC the authority to seek a civil penalty of up to three times the amount of profit gained or loss avoided from insider trading. The statute also increased the maximum fine for a criminal violation (not necessarily limited to insider trading) from $10,000 to $100,000. See H.R. REP. No. 98-355, at 1 (summarizing the ITSA).


99 See H.R. REP. No. 98-355, at 13-14 (1984), reprinted in 1984 U.S.C.C.A.N. 2274, 2286-87 (explaining that the Committee decided not to define insider trading in the new legislation). Some congressional witnesses suggested that Congress should define insider trading. Congress, however, resisted such pleas, arguing that because the insider trading law was "sufficiently well-developed...to provide adequate guidance," a definition was not needed and that a definition could reduce flexibility and create new ambiguities. Id. at 13; see also 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(610), at 7:160.4 ("Congress and the SEC strongly resisted pleas from the bar to define insider trading.")
could subject defendants to the new penalty even if their motives for trading were innocent. The SEC, on the other hand, advocated the "possession" standard. Advocates of the "use" standard have argued that the ITSA's "possession" language does not resolve the "possession versus use" debate, asserting that the ITSA does not define the insider trading violation and merely sets "a condition for the SEC to seek the civil penalty which is discretionary with the court and which is implicitly dependent on the court finding a violation." They argue that the "possession" standard is a "necessary condition for the civil penalty," but that Congress did not intend for "possession" to be "a sufficient condition for the penalty, much less for the violation." These advocates contend that by employing the "possession" language, Congress chose not to involve itself in the "possession versus use" debate, leaving it to the courts to decide whether insider trading violations require use of material, nonpublic information or whether mere possession is sufficient. Their arguments appear to be supported by the testimony of

---

100 See The Insider Trading Sanctions Act of 1983: Hearing on H.R. 559 Before the Subcomm. on Sec. of the Senate Comm. on Banking, Housing, and Urban Affairs, 98th Cong. 116 (1984) (hereinafter Senate Hearings) (statement of Faith Colish, attorney) (asserting that "it is important to distinguish between possession and use of [inside] information"); id. at 126 (statement of the Securities Industry Association) (explaining that "[t]he legislation should... be couched in terms of trading 'on the basis of' covered information"); House Hearings, supra note 32, at 196-98 (statement of Arnold S. Jacobs, attorney) (asserting that the "use," not the "possession" standard reflects the insider trading case law); id. at 296-97 (statement of the New York State Bar Association Committee on Securities Regulation) (explaining that "the critical conduct is not the purchase or sale of a security 'while in the possession' of material non-public information; rather, it is the purchase or sale of a security 'on the basis of' material non-public information"); see also 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(610), at 7:160.3 ("Congress chose the 'possession' language despite objections from a number of witnesses that it was incorrect or inappropriate."); Stevenson, supra note 97, at 18 ("Several congressional witnesses urged that this [possession] language be amended to limit the penalty's applicability to people who traded 'on the basis of' material non-public information.").

101 See House Hearings, supra note 32, at 48 (statement of Daniel L. Goelzer, SEC General Counsel) (asserting that "possession of material inside information is the test").

102 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(610), at 7:160.3; see also SEC v. Adler, 187 F.3d 1325, 1337 (11th Cir. 1998) (adopting this argument).

103 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(610), at 7:160.4.

104 See Stevenson, supra note 97, at 20 (suggesting that Congress did not employ the "on the basis of" language in the ITSA because it "chose not to enroll itself in resolving definitional nuances"); see also WILLIAM K.S. WANG & MARC I. STEINBERG, INSIDER TRADING § 4.4.5, at 184 (1996) ("Choice of the phrase 'while in possession of' could be either an endorsement of the broader [possession] standard or a refusal to choose between the two standards."); Stuart J. Kaswell, An Insider's View of the Insider Trading and
SEC Director of Enforcement, John M. Fedders. When the Committee was considering the “possession versus use” issue, Fedders explained:

The proposed legislation in my view goes to a remedy. It does not at the present time at all impact the existing case law with regard to insider trading. It is strictly a remedy saying that if a person engages in this insider trading, however defined, that then the amount of disgorgement can be three times the ill-gained profit. And the proposed language that you have before you, presented by the Commission, does not impact the “based on,” “in possession of,” or a “knowing” standard at all.

But Professor Langevoort has argued that, although Congress chose not to define insider trading, it did endorse the “possession” standard. Noting that Congress examined the “possession versus use” debate and heard testimony on both sides of the issue, Langevoort argues that Congress was well aware of the implications of its decision to include the “possession” language.

Securities Fraud Enforcement Act of 1988, 45 BUS. LAW. 145, 158 n.57 (1989) (explaining that the ITSEFA’s “possession” language “did not mean that [Congress] endorsed a ‘possession’ standard” because “[t]his language had appeared in ITSA and [Congress] simply replicated existing language,” and asserting that “this legislation was entirely neutral with respect to the debate over the possession standard”).

House Hearings, supra note 32, at 49 (statement of John M. Fedders, SEC Director of Enforcement). But see id. (“[I]f you put the ‘knowing’ standard in or if you put the ‘based on’ standard in, you would be making our prosecutorial efforts much more difficult.”).

See LANGEVOORT, supra note 4, § 3.04, at 3-24 (“The choice of the ‘while in possession of’ statutory language... can be read as an endorsement of the broader [possession] test for insider trading liability.”).

See Donald C. Langevoort, The Insider Trading Sanctions Act of 1984 and Its Effect on Existing Law, 37 VAND. L. REV. 1273, 1290 (1984) (“[A] discussion in the course of the House hearings suggests that the drafters were aware of precisely what they were doing.”); see also LANGEVOORT, supra note 4, § 3.04, at 3-24 (“[T]he legislative history is clear that this language was chosen to reflect precedent that makes motivation insignificant in determining insider trading liability.”); id. at 3-24 n.8 (“Certainly, Congress had the opportunity to change the law to an ‘on the basis of’ test, but chose instead to endorse what it felt was the preferable prevailing interpretation of the law.”); LOSS & SELIGMAN, supra note 21, at 786 (“Congress, both in the Insider Trading Sanctions Act of 1984 and in the Insider Trading and Securities Fraud Enforcement Act of 1988, opted specifically for the ‘possession’ test.”).

Indeed, although the securities laws do not define insider trading, Congress has described insider trading as “the term used to refer to trading in the securities markets while in possession of ‘material’ information (generally, information that would be important to an investor in making a decision to buy or sell a security) that is not available to the general public.” H.R. REP. NO. 98-355, at 2 (1984) (emphasis added), reprinted in 1984 U.S.C.C.A.N. 2274, 2275; see also 130 CONG. REC. 20,107 (1984) (statement of Senator D’Amato) (“Insider trading is defined loosely as ‘trading while in possession of material nonpublic information.’” (emphasis added)).
Having briefly reviewed the “possession versus use” debate, the next Part will examine more closely and assess the arguments that have been offered in support of the “use” standard, concluding that these arguments improperly rely on prior case law and are based on incorrect interpretations of Supreme Court precedent.

III. ARGUMENTS SUPPORTING THE “USE” STANDARD

Advocates of the “use” standard, including the Adler and Smith courts, have found support for the “use” standard in insider trading case law, especially in Supreme Court precedent. In particular, they have observed that such precedent contains references to the “use” of inside information or to trading “on the basis of” such information. These advocates have also noted the Supreme Court’s emphasis on fraud in its discussions of the insider trading prohibition. But, as will be discussed, this reliance on prior case law—and, more specifically, on Supreme Court precedent—is misplaced.

A. “Use” Language

Advocates of the “use” standard often point to the “use” language employed throughout the Supreme Court opinions addressing insider trading. They argue that this language suggests that the trader must have actually used the information to be liable for insider trading. Reliance on this “use” language, however, is inappropriate. In all of the Supreme Court cases cited by advocates of the “use” standard, it was clear that the defendants had in fact used nonpublic information in consummating their trades. As a result, it was natural for the Court to employ “use” language, and the Court was not careful about distinguishing between “possession” and “use.”

In Chiarella, for example, the Supreme Court asserted that “a duty to disclose under § 10(b) does not arise from the mere possession of nonpublic market information.” And throughout Chiarella, the Supreme Court refers to the “use” of inside information and trading “on the basis of” such information. Similarly, throughout the Dirks opin-

---

103 See, e.g., Adler, 137 F.3d at 1337 (“[W]e believe that Supreme Court dicta and the lower court precedent suggest that the use test is the appropriate test.”). See generally Horwich, supra note 5 (providing extensive examination of insider trading case law in the context of the “possession versus use” debate).
110 See, e.g., id. at 226 (explaining that in Cady, Roberts, the SEC found that a broker-dealer violated § 10(b) by “selling securities on the basis of undisclosed information”
The Supreme Court referred to "trading on" inside information, and asserted that "insiders [are] forbidden ... from personally using undisclosed corporate information to their advantage." More recently, in O'Hagan, the Supreme Court, referring to the "classical theory" of insider trading liability," explained that "[section] 10(b) and Rule 10b-5 are violated when a corporate insider trades in the securities of his corporation on the basis of material, nonpublic information" and, referring to the misappropriation theory, asserted that "the fiduciary's fraud is consummated, not when the fiduciary gains the confidential information, but when, without disclosure to his principal, he uses the information to purchase or sell securities." The Supreme Court's repeated employment of phrases such as "on the basis of"—particularly in the recent O'Hagan opinion—has suggested to many observers that the Court has considered the "possession versus use" debate and has endorsed the "use" standard.
As the Adler and Smith courts acknowledged, however, the Supreme Court's use of such language is merely dicta. The issue of whether possession is sufficient was not before the Court in any of those cases: in all three cases, it was clear that the defendants had in fact used the inside information in consummating their trades. Because the defendants had used the information, it was natural for the Court to employ "use" language; in referring to insider trading with "use" language, the Court was merely describing what the defendants had done in those cases. Further, because the issue of "possession versus use" was not before the Court, it is likely that the Court did not consider carefully its choice of language in that context. Indeed, in

> See Smith, 155 F.3d at 1067 (recognizing that the Supreme Court's language was merely dicta); Adler, 137 F.3d at 1334 ("We acknowledge that the Supreme Court's language in Chiarella, Dirks, and O'Hagan is dicta, because in those cases there was no question that the material nonpublic information was actually used in trading.").

> See O'Hagan, 521 U.S. at 648 (explaining that O'Hagan "us[ed] for his own trading purposes material, nonpublic information regarding Grand Met's planned tender offer [of Pillsbury]"); Dirks, 463 U.S. at 648 (asserting that Dirks "disclosed [material, nonpublic] information to investors who relied on it in trading in the shares of the corporation"); Chiarella v. United States, 445 U.S. 222, 245 (1980) (Burger, C.J., dissenting) (referring to Chiarella's testimony at trial that he "used that information [the names of companies that were to become the object of takeover bids] as a basis for purchasing stock" (internal quotations omitted)).

> See 7 LOSS & SELIGMAN, supra note 39, at 3504-05 ("When there is no question that the inside information was actually used in trading—which is normally the case—it seems natural to speak in terms of not trading 'on' or 'on the basis of' the information without necessarily implying that possession alone would not suffice.").

> There are several lower court decisions in which the court employs both "use" and "possession" language interchangeably. See 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(620), at 7:160.12 ("A number of other decisions speak of use (or its equivalents) but also, and more or less interchangeably, of possession."); see also United States v. Cusimano, 123 F.3d 83, 87 (2d Cir. 1997) (explaining that "under [the traditional theory of insider trading] a corporate insider trades in the securities of his own corporation on the basis of material, non-public information," but that "under the misappropriation theory § 10(b) and Rule 10b-5 are violated whenever a person trades while in knowing possession of material, non-public information that has been gained in violation of a fiduciary duty to its source" (emphasis added)); Fridrich v. Bradford, 542 F.2d 307, 312, 314 (6th Cir. 1976) (noting that "the district court found that all [the defendants] had violated Rule 10b-5 by trading in ... stock while in possession of material inside information," and later describing the issue before the court as "how far the courts are to extend the private civil right of action under Section 10(b) and Rule 10b-5 when the alleged violation is the unlawful use of inside information and the stock involved is
O'Hagan, the Court employed “use” language when referring to Rule 14e-3, even though that rule explicitly adopted a “possession” standard. The Court asserted: “One violates Rule 14e-3(a) if he trades on the basis of material nonpublic information concerning a pending tender offer that he knows or has reason to know has been acquired... from an insider of the offeror or issuer, or someone working on their behalf.” It thus appears that the Court was not paying close attention to its employment of “use” and “possession” language.

When analyzing the language of the Supreme Court, one must also examine carefully the context of the Court’s statements and the point the Court was attempting to make. For example, when the Supreme Court declared that “a duty to disclose under § 10(b) does not arise from the mere possession of nonpublic market information,” the Court was referring to the notion that there is no fraud under section 10(b) and Rule 10b-5 absent a fiduciary duty to disclose information before trading. The Court asserted that such a duty to disclose “arises from a specific relationship between two parties,” such as a “relationship of trust and confidence between the shareholders of a corporation and those insiders who have obtained confidential information by reason of their position with that corporation.” Rather than making a distinction between “possession” and “use” as an element of an insider trading violation, the Court was instead addressing the issue of who could be liable for insider trading—to whom the duty to “abstain or disclose” applied. And, explicitly rejecting the parity-of-information theory, the Court concluded that the “disclose or abstain” rule applies not to everyone who possesses (or uses) material, traded upon an impersonal market” (emphasis added)); SEC v. Antar, 15 F. Supp. 2d 477, 528, 529 (D.N.J. 1998) (noting that “§ 10(b) and Rule 10b-5 are violated when a corporate insider trades in the securities of his corporation on the basis of material, nonpublic information,” and later asserting that “the provisions discussed above [section 10(b) and Rule 10b-5] prohibit a person from selling securities while in possession of material, nonpublic information” (emphasis added) (citations omitted)).

See supra note 58 and accompanying text (discussing Rule 14e-3 and observing that it provides for a “knowing possession” standard).

O'Hagan, 521 U.S. at 669 (emphasis added) (quoting United States v. Chestman, 947 F.2d 551, 557 (1991) (en banc)).

Chiarella, 445 U.S. at 235.

See id. (“When an allegation of fraud is based upon nondisclosure, there can be no fraud absent a duty to speak.”).

Id. at 233.

Id. at 228.

See 2 BROMBERG & LOWENFELS, supra note 6, § 7.4(620), at 7:160.11 (“[P]ossession is considered by the Court in reference to the creation of the disclose or abstain duty, not in juxtaposition to use as an element of violation.”).
nonpublic information, but only to those with a pre-existing fiduciary duty to shareholders.\textsuperscript{127}

Similarly, in considering the Supreme Court's statements in \textit{O’Hagan}, it is important to note that the Court’s discussion took place in the context of the misappropriation theory.\textsuperscript{128} In the context of the “possession versus use” debate, some commentators have drawn a distinction between the classical theory of insider trading and the misappropriation theory.\textsuperscript{129} They argue that the misappropriation theory presupposes the use of inside information: underlying the theory is the notion that “a fiduciary’s undisclosed, self-serving use of a principal’s information to purchase or sell securities, in breach of a duty of loyalty and confidentiality, defrauds the principal of the exclusive use of that information.”\textsuperscript{130} Under this theory, the “misappropriation” oc-

\textsuperscript{127} See \textit{Chiarella}, 445 U.S. at 231-33 (rejecting the parity-of-information theory). For a discussion of the parity-of-information theory, see supra note 29 and accompanying text.

\textsuperscript{128} See \textit{O’Hagan}, 521 U.S. at 649 (explaining that the Court was considering “the propriety of the misappropriation theory under § 10(b) and Rule 10b-5”). For a description of the misappropriation theory, see supra notes 18-20 and accompanying text.

\textsuperscript{129} See, e.g., \textit{FERRARA ET AL.}, supra note 6, § 2.01[5], at 2-14.6 to 2-14.7 (suggesting that “the answer [to the ‘possession versus use’ debate] is best viewed as a function of a number of specific factors, including: whether the case is brought under the misappropriation theory or the classical theory of insider trading liability”); \textit{Horwich}, supra note 5, at 1237-38 (“In order to focus the analysis of whether there is a distinction between (i) affirmatively using material inside information to trade, and (ii) trading while in possession of material inside information but without making deliberate use of the information, the analysis [in] this Article is limited to the classical theory.”); see also supra Part I.A (discussing the classical theory of insider trading and the misappropriation theory). \textit{But see infra} note 133 (arguing that if one accepts the “equal-access” theory, there is no reason to distinguish between the classical theory and the misappropriation theory for purposes of the “possession versus use” debate).

\textsuperscript{130} \textit{O’Hagan}, 521 U.S. at 652. The view that the misappropriation theory presupposes use, however, is not universally held. See \textit{United States v. Cusimano}, 123 F.3d 83, 87 (2d Cir. 1997) (asserting that “under the misappropriation theory § 10(b) and Rule 10b-5 are violated whenever a person trades \textit{while in knowing possession} of material, nonpublic information that has been gained in violation of a fiduciary duty to its source” (emphasis added)); \textit{LANGEVOORT}, supra note 4, § 6.05[3], at 6-32 (referring to the misappropriation theory and describing the prohibited conduct as “trading while in possession” of inside information); \textit{id.} at 6-32 n.16 (“Actually, an open question is whether the misappropriation theory requires motivation or mere possession.

Indeed, \textit{Teicher}, the case in which the Second Circuit explicitly endorsed the “knowing possession” standard, involved the misappropriation theory. See \textit{United States v. Teicher}, 987 F.2d 112, 119 (2d Cir. 1993) (discussing the misappropriation theory and its application to the case). The Supreme Court may have employed “use” language in \textit{O’Hagan} merely because it was clear that the defendant had in fact used the inside information. See supra notes 117-19 and accompanying text (arguing that because the defendant used the information, it was natural for the Court to employ “use” language); \textit{see also} \textit{BROMBERG \\& LOWENFELS}, supra note 6, § 7.5(517)(3), at 7:242.27 to
curs when the fiduciary uses the information without the principal’s knowledge or permission.\footnote{See O’Hagan, 521 U.S. at 656 (“[T]he fiduciary’s fraud is consummated . . . when, without disclosure to his principal, he uses the information to purchase or sell securities.”); Phyllis Diamond, McLucas Hails O’Hagan Ruling, But Says Issues over Reach of Theory Remain, 29 SEC. REG. & L. REP. 1097, 1098 (1997) (quoting SEC Enforcement Director William McLucas, who explained that “the impropriety involved when someone like Mr. O’Hagan is a defendant is not getting the information” and that the misappropriation takes place when the defendant “uses [the information] to trade”); see also Horwich, supra note 5, at 1238 n.13 (“While it is easy to pose examples of an insider possessing nonpublic information but arguably not using it in making a trading decision, it is much more difficult to envision a plausible situation where someone first misappropriates information (an affirmative wrongful act) and then trades without using it.”).}

Even if the Court was consciously and purposefully employing “use” language in O’Hagan,\footnote{See supra note 119 and accompanying text (asserting that because the “possession versus use” issue was not before the Supreme Court, the Court may not have paid close attention to its choice of “possession” or “use” language).} such language is closely linked with its discussion of the misappropriation theory and should not be read as implicating the “possession” standard under the classical theory.\footnote{See Ferrara et al., supra note 6, § 2.01[5], at 2-14.7 to 2-14.8 (suggesting that the O’Hagan Court’s use of the phrase “on the basis of” is intertwined with its analysis of the misappropriation theory and arguing that under the classical theory of insider trading the “knowing possession” standard may still be “the better test”); Diamond, supra note 131, at 1098 (asserting that despite the Court’s “use” language in O’Hagan, “certainly with respect to traditional insiders and certainly in the conventional analysis of tippee liability, we would argue that the standard remains . . . ‘in possession of’”).}
Proponents of the "use" standard have relied heavily on the "use" language employed in many of the Supreme Court cases addressing insider trading. Such reliance, however, is misplaced. In each of those cases, a causal connection clearly existed between the insider's trades and the inside information. Because the defendants had in fact used the information, it was natural for the Court to employ "use" language. Advocates of the "use" standard have recognized that "possession versus use" was not at issue in these cases and thus that the Supreme Court's adoption of "use" language is merely dicta. These advocates, however, have failed to recognize that the Court's choice of language in the cases is unreliable because the Court was not careful about distinguishing between "possession" and "use."

B. Scienter and the Emphasis on Fraud

Advocates of the "use" standard have also found support for their position in the Supreme Court's emphasis on fraud in its discussions of section 10(b) and Rule 10b-5 and in the scienter requirement. Liability cannot attach under section 10(b) and Rule 10b-5 unless it is shown that the defendant acted with scienter. Nevertheless, the courts have not clearly defined the term—particularly in the context of insider trading—the Supreme Court has asserted that scienter "refers to a mental state embracing intent to deceive, manipulate, or defraud."

Inherently an elusive concept, scienter suffers particular problems in insider trading. It has to fit many different versions of the violation, some of them quite peculiar. Some of the standard articulations aren't well suited to insider trading. And, perhaps as a result, courts struggling to find words that make sense in factual contexts use multiple expressions which are not always consis-
Some advocates of the "use" standard have argued that a showing of scienter requires a showing that the defendant actually used the inside information in consummating his trade. These arguments, however, are based on an improper interpretation of the scienter requirement in the context of insider trading.

In Adler, for example, the Eleventh Circuit argued that the defendant's motivation for trading—namely, whether or not he traded on the basis of inside information—was relevant to the issue of scienter. The Adler court based its argument on a footnote from Dirks, in which the Supreme Court asserted:

> [M]otivation is not irrelevant to the issue of scienter. It is not enough that an insider's conduct results in harm to investors; rather, a violation may be found only where there is "intentional or willful conduct designed to deceive or defraud investors by controlling or artificially affecting the price of securities."

Similarly, in Smith, the Ninth Circuit argued that the scienter requirement supports a "use" standard, contending that "an investor who has a preexisting plan to trade, and who carries through with that plan after coming into possession of material nonpublic information, does not intend to defraud or deceive." Again, these arguments are based on an improper interpretation of the scienter requirement in the context of insider trading.

Under the common law concept of fraud, a defendant is said to tent... Moreover, scienter is described in very different terms, sometimes in the same court opinion.

2 Bromberg & Lowenfels, supra note 6, § 7.4(810), at 7:163. Nor is the term scienter defined in the federal securities laws. See id. § 7.4(800), at 7:162 ("[Scienter] appears nowhere in the 1933 or 1934 Securities Acts or their rules.").

137 Adler, 137 F.3d at 1334 (noting that "the majority and the dissent in Dirks disagreed about whether motivation is relevant to proving scienter as an element of an insider trading violation").

138 Dirks, 463 U.S. at 663 n.23 (quoting Hochfelder, 425 U.S. at 199).

139 Smith, 155 F.3d at 1068.
act with scienter if he misstates a material fact, knowing that the statement is false and intending that the recipient of such information rely on it. In the context of insider trading, the scienter requirement has been interpreted to mean that the defendant must have had actual knowledge of material, nonpublic information and also must have known that the information was both material and nonpublic. The reason for the defendant's actions is largely irrelevant, provided he acted knowingly.

Indeed, in interpreting the scienter requirement, one must distinguish between "the lack of good faith required for scienter" and "bad motive," which is not required. Although scienter requires intent, it does not require intent "in the sense of intent to harm or do evil"; rather, the intent required by scienter is "the intent to commit the fraud itself." As William H. Kuehnle has explained, "the graver men of the moral deficiency in fraud is not that the defendant was try-

---

140 See LOSS & SELIGMAN, supra note 21, at 751 (explaining that "the defendant must know of the falsity [of his statement] (this kind of knowledge is called 'scienter'), but make the statement nevertheless for the purpose of inducing the plaintiff to rely on it").

141 See SEC v. MacDonald, 699 F.2d 47, 50 (1st Cir. 1983) (explaining that "the [scienter] requirement is satisfied if at the time defendant purchased stock he had actual knowledge of undisclosed material information; knew it was undisclosed, and knew it was material"); see also ROBERT CHARLES CLARK, CORPORATE LAW § 8.10.3, at 328 (1986) ("The point of [Hochfelder] in this context would seem to be only that the defendant must have known that the information to which he had access while trading was material and nonpublic.").

In Hochfelder, the Supreme Court expressly reserved the question of whether recklessness could constitute scienter, see Hochfelder, 425 U.S. at 194 n.12 ("In certain areas of the law recklessness is considered to be a form of intentional conduct for purposes of imposing liability for some act. We need not address here the question whether, in some circumstances, reckless behavior is sufficient for civil liability under § 10(b) and Rule 10b-5."), and "[v]irtually all lower courts addressing this issue since Hochfelder have concluded that a rule 10b-5 violation may be grounded on 'recklessness' . . . and that knowing, intentional misconduct is not a necessary ingredient of establishing liability." ROBERT W. HAMILTON, CASES AND MATERIALS ON CORPORATIONS 935 (5th ed. 1994).

142 As Langevoort notes: "The law of fraud on which this [duty to disclose or abstain] is based is a clear-cut disclosure duty for the benefit of the protected party, not limited to situations where it is shown that the fiduciary was trying to take advantage of the information." Langevoort, supra note 4, § 3.04, at 3-26; see also RESTATEMENT (SECOND) OF TORTS § 551 (2) (a) (1977) ("One party to a business transaction is under a duty to exercise reasonable care to disclose to the other before the transaction is consummated, (a) matters known to him that the other is entitled to know because of a fiduciary or other similar relation of trust and confidence between them . . .").


144 Id.
ing to harm the victim, but rather that the defendant was willing to mislead the victim." If, for example, an insider knowingly made false statements to an investor about the corporation's stock, he would be guilty of fraud—regardless of his motives for making such misrepresentations. Analogizing to the context of insider trading, a defendant who knowingly trades while in possession of material, nonpublic information acts with the requisite scienter, regardless of his motives for trading. As the SEC argued in Smith, the "requisite intent to defraud is inherent in the act of trading while in possession of inside information." The Adler and Smith courts fail to make this distinction between "intent to harm" and "intent to commit fraud" when they assert that a defendant does not act with the requisite scienter if he has an "innocent" reason for trading.

Much of the confusion surrounding the scienter requirement may be attributed to the frequent establishment of scienter through the use of circumstantial evidence. In reading cases that have discussed the scienter requirement, courts and commentators have confused the reliance on use of information to prove scienter with a requirement that use of information be proved in order to satisfy the scienter element. In some cases, there may not be direct evidence that the defendant had knowledge of (knowingly possessed) the inside information or that the defendant knew such information was material. In such cases, courts may rely on "suspicious" trading as evidence that the defendant actually knew the information, or the court may rely on evidence that the defendant used inside information to prove that the defendant knew such information was material. In these cases, evidence of the defendant's use of information is not independently important; rather, it is relevant merely as a means of establishing that the

---

1999] THE "POSSESSION VERSUS USE" DEBATE 273

145 Id.

146 See id. (discussing a similar hypothetical and citing RESTATEMENT (SECOND) OF TORTS § 526 (1977) as support); see also LOSS & SELIGMAN, supra note 21, at 842 n.314 (citing W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS 741 (W. Page Keeton ed., 5th ed. 1984) and asserting that "the intent to deceive or mislead or convey a false impression that is implicit in the scienter concept goes simply to the matter of belief, or absence of belief, that the representation is true").

147 Smith, 155 F.3d at 1068.

148 See Herman & MacLean v. Huddleston, 459 U.S. 375, 390 n.30 (1983) (noting that "proof of scienter required in fraud cases is often a matter of inference from circumstantial evidence"); see also LANGEVOORT, supra note 4, § 3.04, at 3-22 n.2 ("Scienter can, of course, be established by circumstantial evidence."); LOSS & SELIGMAN, supra note 21, at 837 ("[I]ntent (or scienter) being subjective, it must often 'be inferred from a series of seemingly isolated acts and instances which have been . . . aptly designated as badges of fraud.'" (quoting Nassan v. United States, 126 F.2d 613, 615 (4th Cir. 1942))).
defendant possessed the information and knew that the information was material. In SEC v. MacDonald, for example, the defendant appealed the trial court's finding of scienter, asserting that the SEC had failed to establish that he knew the inside information was material. The court, however, found that the "inside information was a motivating factor in [the defendant's] purchase of... stock" and concluded that such use of the information was evidence that the defendant considered the information to be material.

On the other hand, where there is little or no evidence that the defendants actually possessed material, nonpublic information, the court may rely on evidence of legitimate or innocent reasons for trading to conclude that the defendant did not act with the requisite scienter. In In re Worlds of Wonder Securities Litigation, for example, the plaintiffs alleged that the defendants, insiders who sold some of their shares prior to the company's bankruptcy, violated the insider trading prohibitions. The plaintiffs tried to use evidence of the defendants' trading as evidence that the defendants had access to material, nonpublic information. But, observing that the defendants sold only a small percentage of their holdings and suffered large losses when the

---

149 MacDonald, 699 F.2d at 50 ("Defendant does not contest... the court's finding that he had knowledge of the facts found to be material, and knew such facts to be undisclosed. He does claim, however, that the SEC failed to establish... the necessary third element, knowledge of materiality... ").
150 Id. at 51.
151 See id. ("Obviously, if defendant himself considered the information important in deciding whether to purchase, he knew a reasonable investor would likewise consider it important in deciding whether to sell.").
152 As Langevoort explains:
   In many cases, a defendant argues factually that he was not in possession... of any inside information when he traded. The SEC... has only circumstantial evidence of possession. In this setting, many courts have wisely taken the position that the timing of the trade (e.g., shortly after a telephone conversation with someone who clearly knew the information and had some reason to pass it on) suffices to create an inference of insider trading, but only if such a trade was inconsistent with the defendant's prior pattern of trading activity. As such, if the defendant had some predetermined plan of selling, or was buying or selling only in normal amounts, it would rebut any inference of... possession...
153 Langevoort, supra note 4, § 3.04, at 3-25.
154 See id. ("Plaintiffs assert that since these defendants sold some of their shares... prior to [the company's] financial collapse, they must have known of some information, not already known to the market, that signalled [the company's] imminent doom." (quoting In re Worlds of Wonder Sec. Litig., 814 F. Supp. 850, 871 (N.D. Cal. 1993) [hereinafter Worlds of Wonder II])).
company went bankrupt, the court suggested that the trades did not seem suspicious and asserted that one could not reasonably conclude from such trades alone that the defendants possessed inside information. Because there was no other evidence that the defendants possessed nonpublic information, the court thus concluded that the defendants lacked scienter.

In reading cases such as MacDonald and Worlds of Wonder, courts and commentators have confused the reliance on use of information to prove scienter with a requirement that use of information be proved in order to satisfy the scienter element. Indeed, the Adler court cited Worlds of Wonder in support of the “use” standard, asserting that “[i]n a number of cases, courts have allowed insiders to introduce evidence of preexisting plans or other ‘innocuous’ reasons for sales in order to rebut an inference of scienter.” In particular, the Adler court pointed to the statement in Worlds of Wonder that “[e]ven if the evidence was sufficient to permit an inference that one or more of the defendants had access to inside information, the defendants’ actual trading would conclusively rebut an inference of scienter.” The Adler court further observed that in Worlds of Wonder, “the court emphasized that some of the insiders ‘sold their shares pursuant to a prede-termined plan,’ one insider sold ‘because it faced a pressing need to service a huge debt incurred from overinvesting in real estate,’ and another insider only sold a small percentage of his shares.” The Adler court therefore concluded that scienter requires use—not mere possession—of inside information. But in interpreting Worlds of Wonder, it is important to recall that there was no evidence that the defendants actually possessed any inside information. The court was not

---

155 See id. (asserting that “when one considers the amounts traded, Plaintiffs’ claims appear fantastic” and concluding that “no reasonable jury could find that insider trading had occurred” (quoting Worlds of Wonder, 814 F. Supp. at 872)).

156 See id. (explaining that the “[p]laintiffs cannot point to any bit of information traded on by these defendants that was not already known to the market” and concluding that “no reasonable jury could find . . . that these defendants acted with scienter” (quoting Worlds of Wonder, 814 F. Supp. at 872)).

157 Adler, 137 F.3d at 1335.

158 Id. (quoting Worlds of Wonder, 35 F.3d at 1427-28).

159 Id. (citing Worlds of Wonder, 35 F.3d at 1427-28).

160 See id. (asserting that cases such as Worlds of Wonder “provide support for the proposition that there is no violation of § 10(b) and Rule 10b-5 in the absence of some causal connection between the material nonpublic information and an insider’s trading”).

161 See supra note 156 and accompanying text (discussing the lack of evidence that defendants possessed nonpublic information).
suggesting that use of information is a requirement for scienter. Rather, the defendants’ preexisting trading plans were relevant only as far as they provided evidence that the defendants did not possess inside information.\(^{162}\) Indeed, if there had in fact been independent proof that the defendants possessed material, nonpublic information, the court likely would have found scienter. The *Adler* court’s conclusion that the scienter requirement mandates a “use” standard is thus based on an improper interpretation of scienter and the manner in which it is often established in the context of insider trading.\(^{163}\)

The *Adler* court found further support for its assertion that a defendant’s motivation for trading is relevant in the Supreme Court’s holding that a tipper must personally benefit from his tip to be liable for insider trading.\(^{164}\) The *Adler* court, however, seems to have taken the Supreme Court’s holding out of context. In asserting that an insider must benefit from his tip, the Supreme Court observed that not all disclosures of confidential information constitute a breach of fiduciary duty.\(^{165}\) The Supreme Court noted that an insider who conveys such information may not know that the information is confidential—he may mistakenly believe that the information has already been publicly disclosed—or the insider may not believe that such information is material.\(^{166}\) The requirement of a benefit, then, seems merely to be a means of ensuring that the tipper-defendant knew the information was material and nonpublic.\(^{167}\)

Citing the *O'Hagan* opinion, the *Adler* court also claimed the sup-

\(^{162}\) See Langevoort, supra note 4, § 3.04, at 3-25 n.11 (discussing Worlds of Wonder and the inappropriateness of interpreting that case as favoring the use standard).

\(^{163}\) See id. at 3-26 n.11 ("Citation to these sorts of cases as authority in favor of a use standard . . . is plainly inappropriate."); id. at 3-25 to 3-26 ("These cases . . . have no bearing whatsoever on choosing between the two standards [possession or use]—a point that the courts in Adler and Smith unfortunately missed when they cited them in support of the use test.").

\(^{164}\) Adler, 137 F.3d at 1334 ("[T]he Dirks Court’s holding that [a] tipper must gain some personal advantage in order for [a] tippee to be liable for trading on material nonpublic information, suggests that knowing possession of material nonpublic information at the time of trading may not be enough to establish liability for insider trading."); see also Dirks v. SEC, 463 U.S. 646, 661-62 (1983) (asserting that the insider must benefit from his tip for there to be a breach of fiduciary duty and insider trading liability).

\(^{165}\) See Dirks, 463 U.S. at 661-62 ("All disclosures of confidential corporate information are not inconsistent with the duty insiders owe to shareholders.").

\(^{166}\) See id. at 662 ("Corporate officials may mistakenly think the information already has been disclosed or that it is not material enough to affect the market.").

\(^{167}\) See Langevoort, supra note 4, § 3.04, at 3-26 n.13 (asserting that one cannot look to Dirks’s requirement of a benefit to resolve the “possession versus use” debate).
port of the Supreme Court for the proposition that the "knowing possession test may prohibit actions that are not themselves fraudulent."\footnote{Adler, 137 F.3d at 1338 (citing O'Hagan, 521 U.S. at 673).} In particular, the Adler court noted the following assertion from O'Hagan: "under § 14(e), the Commission may prohibit acts, not themselves fraudulent under the common law or § 10(b), if the prohibition is "reasonably designed to prevent...acts and practices [that] are fraudulent."\footnote{Id. (citing O'Hagan, 521 U.S. at 673).} But the Adler court misinterpreted the Supreme Court's discussion, taking statements it made about section 14(e) and Rule 14e-3 out of their original context. In O'Hagan, the Supreme Court was addressing the defendant's contention that the SEC had exceeded its rulemaking authority under § 14(e) when it adopted Rule 14e-3(a).\footnote{O'Hagan, 521 U.S. at 666-67 (considering whether "the Commission... exceed[ed] its rulemaking authority under § 14(e) when it adopted Rule 14e-3(a) without requiring a showing that the trading at issue entailed a breach of fiduciary duty"). Section 14(e) gave the SEC authority to "define, and prescribe means reasonably designed to prevent, such acts and practices as are fraudulent, deceptive, or manipulative." 15 U.S.C. § 78n(e) (1997).} Although it is true that Rule 14e-3 adopted a "knowing possession standard,"\footnote{See supra note 58 (discussing Rule 14e-3).} this aspect of the rule was not under attack. Rather, O'Hagan contested the validity of Rule 14e-3(a) because it imposes a duty to "disclose or abstain" regardless of whether the trader has a preexisting fiduciary duty to either the shareholders or the source of the information.\footnote{See O'Hagan, 521 U.S. at 669 (explaining that "[i]n the Eighth Circuit's view, [and presumably in the defendant's view,] because Rule 14e-3(a) applies whether or not the trading in question breaches a fiduciary duty, the regulation exceeds the SEC's § 14(e) rulemaking authority"). The "possession" standard suggested for section 10(b) and Rule 10b-5 liability does not eliminate the requirement of a fiduciary duty required under both the classical theory and the misappropriation theory. See supra Part I.A (discussing the two theories of insider trading liability).} As the Supreme Court asserted in Chiarella, there is no fraud absent a fiduciary duty to disclose information before trading.\footnote{Chiarella, 445 U.S. at 235 ("When an allegation of fraud is based upon nondisclosure, there can be no fraud absent a duty to speak."); see also supra notes 122-27 (discussing Chiarella and the fiduciary duty requirement).} Because Rule 14e-3 imposes liability even where there is no such fiduciary duty, effectively prohibiting conduct which is not fraudulent under Chiarella, the defendant in O'Hagan argued that the SEC had exceeded its rulemaking authority to define fraud under section 14(e). The Supreme Court did not resolve whether the SEC exceeded its section 14(e) power to define fraud. Instead, the Court noted that section 14(e) also gave the SEC
"prophylactic" authority to prevent fraud,174 and concluded that Rule 14e-3(a) "qualifies under § 14(e) as a ‘means reasonably designed to prevent’ fraudulent trading on material, nonpublic information in the tender offer context."175 It was only in this context that the Supreme Court held that under section 14(e) the SEC may prohibit actions that are not fraudulent. When considered in the proper context, the Supreme Court's assertion does not support the Adler court's conclusion that the "knowing possession" standard prohibits conduct that is not fraudulent.

Indeed, examined in the proper context, statements by the Supreme Court lend little support to the arguments in favor of the "use" standard; many of the statements relied upon by advocates of the "use" standard have been misinterpreted and taken out of their original context. Although reliance on Supreme Court precedent and other insider trading case law in support of the "use" standard thus appears unwarranted, nothing in the preceding discussion suggests that "knowing possession" is the correct standard.176 The next Part sets forth the reasons for adopting a "possession" standard.

IV. ARGUMENTS SUPPORTING THE "KNOWING POSSESSION" STANDARD

There are two primary reasons for adopting the "possession" standard and imposing insider trading liability for trading while in possession of material, nonpublic information. The first relates to the difficulty of proving actual use. The second relates to the rationale underlying the insider trading prohibition—the concern that inves-

174 See O'Hagan, 521 U.S. at 673 (citing Schreiber v. Burlington N., Inc., 472 U.S. 1, 11 n.11 (1985), in which the Court asserted that "§ 14(e)’s rulemaking authorization gives the Commission ‘latitude,’ . . . to regulate nondeceptive activities as a ‘reasonably designed’ means of preventing manipulative acts, without suggesting any change in the meaning of the term ‘manipulative’ itself").

175 Id. at 672. Note, however, that the SEC seems to have been exercising its fraud-defining authority, rather than its prophylactic authority in Rule 14e-3(a), see 17 C.F.R. § 240.14e-3(a) (1999) (providing that the activity described "shall constitute a fraudulent, deceptive or manipulative act or practice within the meaning of Section 14(e) of the Act"), although the SEC apparently believed it was exercising its prophylactic power in Rule 14e-3(d). See 17 C.F.R. § 240.14e-3(d) (prohibiting certain conduct "as a means reasonably designed to prevent fraudulent, deceptive or manipulative acts or practices within the meaning of section 14(e) of the Act").

176 Indeed, because the "possession versus use" debate has not squarely been at issue in prior cases, one should not interpret the language in earlier case law as endorsing either the "possession" or the "use" standard. See supra Part III.A (discussing the inappropriateness of relying on "use" language in Supreme Court insider trading opinions for support of the "use" standard).
tors perceive the securities markets as fair and honest.\textsuperscript{177}

\section*{A. Difficulty of Establishing Actual Use}

The SEC has advocated the "possession" standard largely because it believes that a "use" standard would make it unnecessarily difficult to prosecute insider trading.\textsuperscript{178} Under a "use" test, the SEC (or the government), as well as private parties, not only would be required to establish that the defendant possessed material, nonpublic information, but also would "have to get into his mind" to determine whether the trade was based on such information.\textsuperscript{179} As Director McLucas explained: "The problem with the 'on the basis of' standard is clear—it makes proof of a violation subject to metaphysical impossibility. . . . When a trader knows the material information, divining that he . . . acted 'on the basis of' as opposed to 'while in possession of' may be impossible."\textsuperscript{180} With the "use" standard, the concern is that insider-traders will "contrive" legitimate reasons for their transactions, offering innocent explanations such as mortgage obligations and college tuition payments.\textsuperscript{181} As the SEC asserted: "Individuals who have actually traded on the basis of inside information frequently attempt to invent arguments that they have traded for other reasons. Under a 'possession' standard, such \textit{post hoc} rationalizations would be irrelevant, and could not be used to impede enforcement of the law."\textsuperscript{182} The \textit{Teicher} court also acknowledged this concern when it endorsed the "knowing possession" standard: "As a matter of policy then, a requirement of a causal connection between the information and the

\textsuperscript{177} See \textit{supra} Part I.B (discussing the reasons for the insider trading prohibition).

\textsuperscript{178} See \textit{House Hearings, supra} note 32, at 49 (statement of John M. Fedders, SEC Director of Enforcement) ("[T]he 'based on' [or 'use'] standard . . . would . . . mak[e] our prosecutorial efforts much more difficult."). The SEC's position, however, has not been consistent. See \textit{supra} notes 49-54 and accompanying text (discussing the SEC's inconsistency in its interpretation of Rule 10b-5).

\textsuperscript{179} \textit{House Hearings, supra} note 32, at 49 (statement of John M. Fedders, SEC Director of Enforcement).

\textsuperscript{180} McLucas & Angotti, \textit{supra} note 34, at 5; see also Diamond, \textit{supra} note 131, at 1098 (quoting McLucas's assertion that the government "cannot metaphysically get into someone's head and discern what factors within their state of mind were directly causal").

\textsuperscript{181} See Diamond, \textit{supra} note 131, at 1098 (explaining that the SEC's concern is that "[t]here's not a defendant that comes in the door that won't have a separate, independent, superseding tort law factor for his buying or selling activity").

trade could frustrate attempts to distinguish between legitimate trades and those conducted in connection with inside information.\footnote{Teicher, 987 F.2d at 121. Other commentators have also agreed with this argument. See, e.g., 7 LOSS & SELIGMAN, supra note 39, at 3505 ("The very difficulty of establishing actual use of inside information points to possession as the test.").}

\section*{B. Perception That Markets are Honest and Fair}

The difficulty of establishing actual use of inside information is not the only reason for adopting a "possession" standard. Considered alone, such an explanation invites the inference that there is nothing wrong with mere possession of material, nonpublic information at the time of a trade, aside from the difficulty of proving actual use of such information. One might infer that the securities laws seek to prevent the use of inside information but, because of difficulties of proof, also prohibit trading while in possession of such information—even though the policy reasons underlying the insider trading prohibition do not really apply to mere possession. This inference is the one that many advocates of the "use" standard apparently have drawn in rejecting the SEC's rationale for the "possession" standard.\footnote{See Horwich, supra note 5, at 1270 (explaining that "choosing the simpler rule is defensible only if it comports equally well with the underlying policy of the law as does the alternative, more complex rule," but that "that is not the case if actual use of the inside information is what underlies the rationale for the prohibition"). Horwich is implicitly assuming that the policy reasons for the insider trading prohibition do not apply to trading while in possession of material, nonpublic information, provided there is no causal connection between the trade and the information. See id. at 1271 (contending that the insider who trades merely while in possession of inside information is "innocent"). As a result, he does not believe that such trading should be prohibited merely because it may be difficult to distinguish between use and possession. See id. at 1270-71 ("The fact that the distinction may be difficult to draw in some cases does not mean that it cannot or should not be drawn wherever possible, so long as the underlying policy for the prohibition can be respected.").}

But the policy reasons underlying the insider trading prohibition do apply to trading while in possession of material, nonpublic information. Recognizing the difficulties of proof associated with a "use" test, the Adler court held that an "inference of use . . . arises from the fact that [an] insider traded while in knowing possession of material nonpublic information."\footnote{Adler, 137 F.3d at 1337; see also id. ("[W]hen an insider trades while in possession of material nonpublic information, a strong inference arises that such information was used by the insider in trading. The insider can attempt to rebut the inference by ad\-dicing evidence that there was no causal connection between the information and the trade . . . .")}

The problem with this approach, however, is its failure to recognize that the rationale for the insider trading prohibition applies
equally to both trading on the basis of, and trading while in possession of, inside information.166

The basic rationale for the insider trading prohibition is the belief that insider trading undermines investors' expectations of honest and fair securities markets and that, as a result, insider trading will discourage investors from participating in the markets.167 The argument is that investors view trading on the basis of material, nonpublic information as fundamentally unfair. The question, then, is whether investors would also question the fairness of markets in which investors routinely trade while in possession of material, nonpublic information—even if such traders claimed that there was no causal connection between such information and the decision to trade.

The answer to this question may be found in an observation by the Teicher court about the nature of information. The Teicher court observed: "Unlike a loaded weapon which may stand ready but unused, material information cannot lay idle in the human brain."168 Even if an insider in possession of material, nonpublic information believes that his trades are motivated solely by legitimate reasons, it is unlikely that the information possessed by a trader would not in some way influence his trades. Investors are likely to believe that such information influenced—whether consciously or subconsciously—the insider's trades, despite the insider's assertions to the contrary. Investors are likely to believe, for example, that an insider with a pre-existing plan to sell shares to pay his daughter's college tuition, may decide to sell 110 shares upon coming into possession of the inside information, instead of the 100 shares he had previously planned to sell. Investors are therefore just as likely to doubt the fairness of markets in which participants trade while in possession of inside information as they are to doubt the fairness of markets in which participants trade on the basis of inside information. It is for this reason that "knowing possession" should be sufficient for insider trading liability.169

166 As the Smith court noted, another problem with this approach is that such a presumption of use cannot be applied in the criminal context. United States v. Smith, 155 F.3d 1051, 1069 (9th Cir. 1998) (asserting that in the context of a criminal prosecution, "[w]e are . . . not at liberty, as was the Adler court, to establish an evidentiary presumption that gives rise to an inference of use"). On the other hand, there is no reason why the standards for criminal and civil liability cannot be different; courts could employ a "use" standard in the criminal context and a "use by presumption" standard in the civil context.

167 See supra Part I.B (discussing the reasons for the insider trading prohibition).

168 Teicher, 987 F.2d at 120.

169 Note that this reasoning applies whether liability is predicated upon the classical theory or the misappropriation theory of insider trading. Investors are just as likely to
Advocates of the use standard suggest that there is nothing inherently unfair about trading merely while in possession of material, nonpublic information, since such trading involves no exploitation or abuse of inside information. That may be true, but, as discussed above, investors are likely to perceive such trading as unfair, believing that a trader possessing inside information cannot escape the influence of such information. And to the extent that the insider trading prohibition is based on the belief that investors will not participate in a market that they perceive to be unfair, it is investors' perceptions that are relevant in determining what conduct should be prohibited.

Some proponents of the "use" standard further suggest that it is unfair to penalize and prevent from trading the insider who engages in a regular and periodic program of selling to liquidate his holdings in the corporation's stock and obtain needed cash. For the insider who, because of cash requirements, must sell regularly and cannot always wait until the information is publicly disclosed, the "possession" test may seem particularly burdensome and unfair. But it need not doubt the fairness of the markets if the trader is a corporate insider trading in the securities of his corporation while in possession of inside information as they are if the trader is an "outsider" such as O'Hagan, see supra notes 16-19 and accompanying text (discussing "outsider trading" and the O'Hagan case), trading in another corporation's securities while in possession of material, nonpublic information obtained from an outside (confidential) source. The "knowing possession" standard should therefore be applied in insider trading cases under both the classical theory and the misappropriation theory of insider trading liability. This conclusion clearly follows if one accepts the concept of equality of access to information (or the "equal-access" theory). See supra notes 29, 133 and accompanying text (discussing the equal-access theory and its implications for the classical theory and misappropriation theory in the context of the "possession versus use" debate). The Supreme Court, however, has rejected the equal-access theory, see supra note 29, and, as a result, courts and commentators may continue to distinguish between the classical theory and the misappropriation theory in the context of the "possession versus use" debate. See supra note 129 and accompanying text (explaining that some commentators have distinguished the classical theory from the misappropriation theory for purposes of the "possession versus use" debate).

See Horwich, supra note 5, at 1270-76 (using various hypotheticals to suggest the unfairness of imposing insider trading liability "when the decision to trade was made without any exploitation of material nonpublic information").

Since the insider has the choice of disclosing the information or abstaining from trading, the insider is prevented from trading only to the extent that he cannot disclose the information. Often, however, insiders do not have the authority to disclose the information. See id. at 1271 n.230 ("[F]ew individual insiders have the discretion or right to choose to disclose confidential material information about their employer.").

See id. at 1273 (discussing the case of an executive with a regular, consistent purchase or sale program).

The situation is a bit different where the insider wishes to buy, rather than sell, stock. Here, there is not the same necessity for engaging in the trade—the insider has
be. Such an insider could, for example, maintain his stock portfolio with an investment advisor who has complete discretion over the nature and timing of transactions. The investment advisor could then execute whatever trades are necessary to provide the insider with the cash he requires. The “knowing possession” standard thus goes further than the “use” standard in ensuring the integrity of the markets, without unnecessarily burdening insiders who wish to trade.

The Smith court, however, was concerned that a “knowing-possession standard would... go a long way toward making insider trading a strict liability crime.” Given the statutorily authorized ten-year prison sentence for insider trading convictions, the court was reluctant to adopt such a standard. But the Smith court’s concern was unnecessary. The “possession” standard does not make insider trading a strict liability crime. First, the scienter requirement ensures that the defendant knowingly (or at least recklessly) possessed inside information and that he knew such information was material. Second, as the Supreme Court noted in O’Hagan, a defendant may not be im-

more “flexibility” to delay his purchases until the information is disclosed to the public. Of course, delaying purchases until disclosure may mean that the insider is purchasing at a higher price. Nonetheless, this situation does not seem unnecessarily unfair to the insider. Delaying the insider’s purchases seems to be a reasonable price to pay for ensuring the integrity of the market. Indeed, as between the ordinary investor and the insider, it is the insider who is best able to bear the risk that the insider will come into possession of inside information when he seeks to trade. See id. at 1271 (recognizing this as a counterargument to the contention that the “possession” standard is unfair).

194 See Alan M. Weinberger, Preventing Insider Trading Violations: A Survey of Corporate Compliance Programs, 18 SEC. REG. L.J. 180, 188 (1990) (explaining that the New York Stock Exchange recommends, among other things, that insiders engage in trading activities through regular investment programs in which the timing of trades is beyond the control of the insider); see also Steinberg & Fletcher, supra note 38, at 1820 (noting that many broker-dealer and investment firms require employees to maintain all of their trading accounts with the firm). Of course, one could still argue that this is unfair, since the insider must give up discretion over his own accounts. As between the ordinary investor and the insider, however, the insider is in the better position to bear the “costs” of inside information. See supra note 193 (making this argument).

195 Such an insider could argue credibly that he did not trade, since it was his investment advisor that made the decision to trade and executed the trade.

196 Smith, 155 F.3d at 1068 n.25.

197 See id. (“In view of the statutorily authorized ten-year prison sentence that may accompany an insider trading conviction, any construction of Rule 10b-5 that de facto eliminates the mens rea requirement should be disfavored.” (citations omitted)); see also 15 U.S.C. § 78ff(a) (1994) (authorizing a ten-year prison sentence).

198 See supra notes 134, 140-41 and accompanying text (discussing the scienter requirement); see also United States v. O’Hagan, 521 U.S. 642, 665 (1997) (explaining that under 15 U.S.C. § 78ff(a), “[t]o establish a criminal violation of Rule 16b-5, the Government must prove that a person ‘willfully’ violated the provision”).
prisoned for violating Rule 10b-5 if he proves that he had no knowledge of the rule.199

Both the rationale underlying the insider trading prohibition, as well as the difficulty of proving an insider’s motivation for trading, thus suggest that courts should adopt the “knowing possession” standard for determining insider trading liability under the federal securities laws.

CONCLUSION

Advocates of the “use” standard have looked primarily to prior case law, particularly Supreme Court precedent, for support of their position. The problem with this approach, however, is that the question of “possession versus use” was not at issue in the cases relied upon, and those opinions were not careful about making a distinction between “possession” and “use.” Indeed, many of the statements relied upon by advocates of the “use” standard either are dicta or have been misinterpreted and taken out of their original context. Rather than relying on case law, then, one should look to the reasons for the insider trading prohibition, as well as to the difficulty of proof associated with prosecution, to determine the proper standard. Such an inquiry suggests that “possession” is the more appropriate standard for imposing insider trading liability. Nonetheless, with the recent opinions by the Ninth and Eleventh Circuits, the question of “possession versus use” seems well on its way to being resolved in favor of the “use” standard. Until the Supreme Court explicitly addresses the question, however, the issue will remain open for debate.

---

199 See 15 U.S.C. § 78ff(a) (“[N]o person shall be subject to imprisonment under this section for the violation of any rule or regulation if he proves that he had no knowledge of such rule or regulation.”). Lack of knowledge of the law, however, is not a defense to the imposition of the statutorily authorized fines of up to $1,000,000. See O’Hagan, 521 U.S. at 666 n.13 (“The statute provides no such defense to imposition of monetary fines.”).