

ARTICLES

The Limits of Formal Economics in Tort Law

THE PUZZLE OF NEGLIGENCE

Shawn J. Bayern[†]

Legal scholars commonly see tort law as the jewel of the law-and-economics movement. “The law of accidents was one of the first bodies of private law successfully analyzed using formal economic models,”¹ a leading law-and-economics textbook confidently declares. Leading commentators, even when critical of law and economics, commonly suggest that tort law has succumbed to fairly straightforward economic modeling.²

This Article challenges the view that economic modeling has explained and justified tort law by demonstrating that it has failed to explain or justify tort law’s basic negligence rule. A rule of negligence, versus alternatives like strict liability, holds injurers liable in tort only if they have not taken reasonable precautions.³ An economic understanding of negligence, such as the one commonly known as the Hand Formula, defines “reasonable precautions” as those that are

[†] Assistant Professor, Florida State University College of Law; BS, Yale University; JD, University of California, Berkeley School of Law (Boalt Hall). I thank Jamie Boyle, Ariel Porat, Ruben Pope, and Mark Seidenfeld for helpful critical comments. All errors are negligent.

¹ ROBERT COOTER & THOMAS ULEN, *LAW & ECONOMICS* 324 (5th ed. 2008).

² *E.g.*, Leo Katz, *A Look at Tort Law with Criminal Law Blinders*, 76 B.U. L. REV. 307, 308 (1996) (resisting a solely economic view in criminal law but referring to “law and economics’ admittedly remarkable success at explaining tort law”).

³ See DAN B. DOBBS, 1 *THE LAW OF TORTS* § 116, at 275 (2001) (“Negligence is conduct that creates or fails to avoid unreasonable risks of foreseeable harm to others.”).

cost-justified.⁴ In general, the models that economists have constructed to justify economic conceptions of negligence rules seem persuasive on the surface, but on closer analysis they suffer from significant problems that make them either logically problematic or impossible to administer. Ultimately, the economists' formal models do not serve to justify negligence rules.

Many commentators, from a variety of perspectives, have already criticized particular forms of economic reasoning in tort law. Some, like Richard Wright, have argued persuasively that courts have diverged from the reductionist view now dominant in American legal academia, and thus that economic reasoning has not, in fact, explained or captured American tort law.⁵ My goal in this Article, however, is normative rather than descriptive: my aim is not to suggest that law-and-economics arguments have failed to influence courts, but rather that they have failed on their own terms to justify the efficiency of negligence rules.

Separately, it is common to see arguments that injurers do not respond to incentives in the way economists predict because people are simply not rational, are not aware of the law, or have more pressing concerns than distant and relatively weak financial incentives.⁶ I am sympathetic to these critiques and believe they are largely correct, but my goal here is somewhat different: my aim is to advance the discussion of tort law by showing that the leading formal, deductive conceptions of negligence rules fail in fundamental ways, essentially on their own terms, even if we accept for argument's sake most or all of their reductionist assumptions about human behavior. In

⁴ RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 167-71 (7th ed. 2007).

⁵ Richard W. Wright, *Hand, Posner, and the Myth of the "Hand Formula"*, 4 *THEORETICAL INQUIRIES IN LAW* 145 (2003) (showing that even Learned Hand and Richard Posner, two judges who, respectively, developed and promoted the Hand Formula, have not generally applied it in tort cases before them); Richard W. Wright, *Justice and Reasonable Care in Negligence Law*, 47 *AM. J. JURIS* 143, 145 (2003) ("The [leading economic understanding of negligence], although pervasive in the secondary literature and mentioned by a small minority of courts, is almost never used by the courts to decide whether particular conduct was negligent. Instead, the courts employ . . . a number of different criteria . . . based on the principles of justice."); see also Michael D. Green, *Negligence = Economic Efficiency: Doubts*, 75 *TEX. L. REV.* 1605, 1611 (1997) (questioning the role of economics in positive tort law); Benjamin C. Zipursky, *Sleight of Hand*, 48 *WM. & MARY L. REV.* 1999, 2002, 2026 (2007) (offering a similar rejection of the Hand Formula as a positive account of negligence law's standard of care and listing other commentary that has done so).

⁶ See Gary T. Schwartz, *Reality in the Economic Analysis of Tort Law: Does Tort Law Really Deter?*, 42 *UCLA L. REV.* 377 (1994) (surveying the state of such arguments).

other words, my goal is to take the leading formal economic conceptions of negligence rules off the table as serious contenders for the sole principled basis for tort law, *even if* humans were generally capable of behaving as simple slaves to rationality.

It is important to say that my goal is *not* to show that economic reasoning has provided no insight into tort law or that a proper understanding of tort law would ignore economic analysis. Indeed, I think economic reasoning has been, and can continue to be, helpful. To understand and justify tort rules, however—even on instrumental grounds—courts and commentators need to look beyond model-based deductions.

Similarly, my argument here is not against negligence rules, which may be supported by sound reasons of morality and policy. My argument is just that these rules cannot be derived in ways that leading legal economists appear to purport to derive them.

In Part I, I review the basic economic features of negligence and strict-liability rules and show why early attempts at analyzing tort law economically provided virtually no help in choosing between those two rules—and indeed how, in fairness, they often were not even meant to. For instance, although it is occasionally thought that the early law-and-economics scholars developed economic analyses that favored negligence rules over strict-liability rules,⁷ even those early scholars often admitted that theoretical microeconomics could not help judges or policymakers choose between negligence and strict liability.⁸ As an example, one early straightforward

⁷ See Gary T. Schwartz, *Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice*, 75 TEX. L. REV. 1801, 1819-20 (1997) (“From an economic perspective the Hand formula makes excellent sense. The formula can be seen as designed to encourage efficient investments in safety and risk reduction; as such, it has served as a cornerstone for economic analysis.”); Frank J. Vandall, *Judge Posner’s Negligence-Efficiency Theory: A Critique*, 35 EMORY L.J. 383, 404 (1986) (referring to “[e]conomists’ preferences for negligence”).

⁸ See Mark M. Hager, *The Emperor’s Clothes Are Not Efficient: Posner’s Jurisprudence of Class*, 41 AM. U. L. REV. 7, 44 (1991) (“Although Posner has attempted to defend the efficient character of the negligence standard, his arguments closely read have concentrated on the modest claim that negligence is *no less efficient* than strict liability.”); Vandall, *supra* note 7, at 404 (“Economists’ preference for negligence is undermined by the fundamental admission that ‘after two decades of writing on the subject, we still cannot say with certainty whether strict liability or negligence is the most efficient rule in most areas of accident law.’” (quoting Henry Hansmann, *The Current State of Law-and-Economics Scholarship*, 33 J. LEGAL EDUC. 217, 226 (1983))); J. Hoult Verkerke, *Notice Liability in Employment Discrimination Law*, 81 VA. L. REV. 273, 312 (1995) (“According to the conventional economic analysis,

understanding of negligence rules was that they are efficient because they give self-interested injurers incentives to take efficient precautions.⁹ But even if this proposition is correct, as a defense of negligence rules over strict-liability rules it would be entirely hollow: strict-liability rules give self-interested injurers similar incentives to take efficient precautions, as the pioneers of law and economics well understood.

In Part II, against this backdrop, I explain and then refute what is currently the leading theoretical economic understanding of negligence rules in tort law—a view that top analysts like Robert Cooter and Steven Shavell have helpfully adopted, clarified, and defended.¹⁰ This view has two components: First, it suggests that negligence rules efficiently create incentives for *bilateral precaution*—that is, for both injurers and victims to behave carefully in cases where the actions of both can decrease the likelihood of injuries. (For some economists, nearly all cases of injury are cases of bilateral precaution. As an example, if a car hits a pedestrian on a sidewalk, it might seem like the car's driver was the only one who could have prevented the accident, but economists point out that the pedestrian might have decreased the likelihood of an accident by staying at home.)¹¹ Negligence rules are said to promote bilateral precaution, loosely speaking, by serving as a threat of liability to both injurers and victims simultaneously: both parties are said to know that if they don't behave carefully, they may be made to bear the costs of an accident that occurs. Second, the leading economic view sets forth a formal understanding of who should bear the costs of an accident when both injurers and victims behave carefully. Economists answer this question by turning to arguments

the choice between strict liability and negligence ordinarily does not affect precautions.”).

⁹ Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 33 (1972) [hereinafter Posner, *A Theory of Negligence*] (“If . . . the benefits in accident avoidance exceed the costs of prevention, society is better off if those costs are incurred and the accident averted, and so in this case the [injurer] is made liable, in the expectation that self-interest will lead it to adopt the precautions in order to avoid a greater cost in tort judgments.”).

¹⁰ See generally COOTER & ULEN, *supra* note 1; Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1 (1980) [hereinafter Shavell, *Strict Liability*].

¹¹ See, e.g., COOTER & ULEN, *supra* note 1, at 338 (suggesting that in the case where a “moving car hits [a] parked car,” the victim might “park [the] car in [a] safer space,” and that pedestrians hit by a car might “walk more safely”). Of course, Cooter and Ulen are not suggesting that a victim *should* necessarily do so, just that it is possible. See *infra* text accompanying notes 38-39 for more on this distinction.

about *activity levels*.¹² An activity-levels argument is one that is sensitive to the possibility that potential injurers and victims might engage in *too much* of an activity (like driving), even if they engage in it safely. For instance, if I know that as a driver I can avoid all liability for accidents simply by driving safely, I might therefore drive *inefficiently often*, thereby leading to more accidents than would be efficient.¹³

For several reasons, neither of these components of the modern theoretical understanding of negligence can successfully justify negligence rules over their alternatives. In short, the defense of negligence as promoting bilateral precaution fails because it depends on an exceedingly fragile formal model that cannot adapt even to features of the world that, by the model's own terms, it ought to address. For instance, it cannot accommodate a rational party's probabilistic assessment that another party may act irrationally or with incomplete information, nor can it address a variety of other limitations. As a result, the model does not serve either as a general explanation or a general justification for negligence rules.

The use of activity-levels arguments is flawed too. As a way to assign liability among innocent parties, activity-levels arguments are ordinarily unhelpful because of problems that courts would face if they tried to administer a tort regime based on them, and also because they create several important kinds of inefficiencies on their own. These inefficiencies relate to a sort of conceptualism that has colonized tort law under the banner of economic modeling: commentators tacitly or explicitly observe that the goal of their models is to reduce the net costs of precautions and accidents, but in fact an instrumentally optimal tort law in the real world would be sensitive to other costs, such as incorrect pricing of activities.

As a result, we reach the perhaps surprising conclusion that formal economics has failed to explain or justify, after decades of attempts, the basic features of negligence law. Again, this does not mean economists have provided no insight into the law of torts; indeed, many of their conceptions of negligence rules have been useful in clarifying some of the problems that tort law aims to address. Nonetheless, for a coherent basis of tort law, this Article suggests that we

¹² For the seminal analysis of activity levels in tort law, see Shavell, *Strict Liability*, *supra* note 10.

¹³ *See id.* at 2-3.

probably need to look beyond the models that even the best law-and-economics scholars, like Shavell, Cooter, and Ulen, have provided.¹⁴

I. BASIC ECONOMIC FEATURES OF NEGLIGENCE AND STRICT LIABILITY

This Part reviews the basic economic landscape in which tort law operates, both to provide enough background to understand modern economic conceptions of tort law and to clarify why the simpler economic understanding of tort rules that preceded the presently prevalent models was not sufficient to justify the efficiency of negligence rules over strict-liability rules.

An example that has become conventional in tort law, as a result of its use by Ronald Coase, involves a railroad that passes by a field and potentially causes damage to it because of sparks from its engine.¹⁵ I borrow Robert Cooter's formulation of the example for ease of exposition, because in Part II I will take up Cooter's arguments (among others) at length:

Suppose that Xavier operates a railroad train that emits sparks that sometimes set fire to Yvonne's cornfield. Xavier can reduce the harm to the corn by installing spark arresters, by running the trains more slowly, or by running fewer trains. In a like manner, Yvonne can reduce the harm by planting her corn farther from the tracks, by planting cabbage instead of corn, or by leaving the fields fallow.¹⁶

In this example, there is potential harm that results, in at least some senses,¹⁷ from the interaction between Xavier and Yvonne's business activities. A rule of *no liability* would make Yvonne, the owner of the cornfield, bear the entire cost of the harmful interaction by preventing her from recovering any

¹⁴ See generally COOTER & ULEN, *supra* note 1; Shavell, *Strict Liability*, *supra* note 10.

¹⁵ Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 29-31 (1960).

¹⁶ Robert Cooter, *Unity in Tort, Contract, and Property: The Model of Precaution*, 73 CAL. L. REV. 1, 5 (1985).

¹⁷ According to what Mark Kelman has called "the most basic Coasean insight," no cases of tortious harm occur without both parties in some sense causing the harm, because, at a minimum, the harm can't occur without both of their existence. Mark Kelman, *The Necessary Myth of Objective Causation Judgments in Liberal Political Theory*, 63 CHI.-KENT L. REV. 579, 579 (1987) ("[G]iven that . . . injury cannot have occurred unless the plaintiff . . . existed, . . . [i]t will never be the case that injury could occur without the plaintiff, such that the defendant is entirely causally responsible.").

money from Xavier in the event of damage to her cornfields. A rule of *strict liability*, by contrast, would make Xavier bear the entire cost of the interaction by requiring that he pay Yvonne for any harm she incurs as a result of the railroad's sparks. A rule of *negligence*, in contrast to both other rules, would make Xavier responsible for paying Yvonne only if some social judgment disapproves of Xavier's actions in running his railroad next to Yvonne's cornfields in a way that potentially hurts Yvonne.

A chief feature of the economic analysis of tort law is the economic conception of negligence rules—that is, a reduction of such rules to a cost-benefit test. This particular formulation, whose popularity is commonly traced to an opinion by Judge Learned Hand,¹⁸ *United States v. Carroll Towing*,¹⁹ would decide Xavier's liability based on the costs and benefits of the precautions he might have taken (per the example, "installing spark arresters, . . . running the trains more slowly, or . . . running fewer trains").²⁰ As an example, if the potential sparks from the railroad are estimated to cause fires that lead to \$4000 worth of expected damage to Yvonne's cornfield, Xavier would be liable if *any* of the precautions he might take are expected to cost less than \$4000. If not, he won't be liable.²¹

Analyzed from society's perspective (rather than either Xavier or Yvonne's private perspectives), economists point out that we want Xavier to take precautions only if their social cost is less than their social benefit. That is, even though it seems harsh to say so, there is some sense in which we *want* Yvonne's cornfield to burn down if it is too expensive to prevent fires. Of course, in an ideal world we wouldn't want to see Yvonne suffer this harm, but given the real-world choice between cornfield fires and the precautions necessary to prevent them, from an efficiency perspective we prefer whichever is cheaper.²² Or, as

¹⁸ E.g., Posner, *A Theory of Negligence*, *supra* note 9, at 32 ("The essential clue . . . [to the economic analysis of tort law] is provided by Judge Learned Hand's famous formulation of the negligence standard . . .").

¹⁹ 159 F.2d 169 (2d Cir. 1947).

²⁰ Cooter, *supra* note 16, at 4.

²¹ There are several significant problems with even this simple formulation of the negligence standard. For a more complete discussion, see *infra* Part II.D.

²² Nothing, certainly, mandates that we analyze this scenario only from the perspective of efficiency. For instance, we might think it is unfair to let Xavier cause fires for which he doesn't have to pay. But the instrumental economic view does capture an important insight, which is that Xavier and Yvonne are in some sense symmetric: if the law made Xavier pay for his fires' harm to Yvonne, then Yvonne's

Richard Posner, one of the chief proponents of this efficiency-oriented view, puts it:

If the cost of safety measures or of curtailment—whichever cost is lower—exceeds the benefit in accident avoidance to be gained by incurring that cost, society would be better off, in economic terms, to forgo accident prevention.²³

Consider how no-liability, strict-liability, and negligence rules fare under this view of social efficiency. If Xavier is *never* liable for fires regardless of how much damage they cause, and if he is selfish, he will have insufficient incentives to take precautions against fire. For instance, it could cost him only \$40 to install spark arresters, but he might not want to incur this expense even if fires are expected to cause many thousands of dollars of harm to Yvonne's fields. There may, of course, be many reasons Xavier would in practice bear a small expense even if the law does not require him to do so. He might, for example, have internalized moral norms, be afraid of feeling guilty for hurting Yvonne, empathize with Yvonne or have positive feelings for her, be concerned about adverse publicity, fear retaliatory action by Yvonne, or even consume corn personally and worry that the price of corn will rise if he repeatedly burns down Yvonne's fields. But if he is both selfish and interested only in maximizing his business profits, it is accurate to say that he will not have proper incentives to take efficient precautions against potentially catastrophic losses.

Consider, next, a rule of negligence. The chief feature of the early economic analysis of negligence rules was that, if these rules are conceived economically (as under Judge Hand's formula), they give Xavier incentives to take efficient precautions.²⁴ An economic conception of negligence under the Hand Formula essentially tracks the social-cost analysis described earlier: Xavier is called negligent (or unreasonable) if he didn't take socially efficient precautions, and he's called nonnegligent (or reasonable) otherwise. For example, if spark arresters could have prevented \$4000 fires at a cost of \$40, and if Xavier does not install spark arresters, he is said to be

choice to grow cornfields will have caused harm to Xavier. And it may not be fair, always, to require Xavier to suffer this harm.

In any event, because my purpose in this Article is to show that the economic accounts of negligence fail on roughly their own terms, I do not dwell, here, on the noneconomic problems with viewing all legal rules too instrumentally.

²³ Posner, *A Theory of Negligence*, *supra* note 9, at 32.

²⁴ *See id.*

negligent for not doing so. But if the spark arresters (and all the other precautions) cost more than \$4000, he is not negligent.

Recall that from the perspective of social efficiency, we want Xavier to take efficient precautions and do not want him to take inefficient ones. Because a negligence rule gives Xavier incentives to take efficient precautions and not to take inefficient ones, the early economic analysts of law presented it as an efficient rule. For instance, as Posner wrote:

If . . . the benefits in accident avoidance exceed the costs of prevention, society is better off if those costs are incurred and the accident averted, and so [injurers are] made liable, in the expectation that self-interest will lead [them] to adopt the precautions in order to avoid a greater cost in tort judgments.²⁵

From the way the early legal economists presented negligence rules, it was possible to infer that their instrumental analysis supported those rules, and only those rules. For instance, early in the economic analysis of law, Posner wrote as follows:

A rule making [an] enterprise liable for the accidents that occur [when precautions are more expensive than accidents] cannot be justified on the ground that it will induce the enterprise to increase the safety of its operations.²⁶

Posner seems to be saying that negligence rules are sufficient to ensure that potential injurers like Xavier behave efficiently, and as a result, rules like strict liability are unnecessary.

But even early on, economic analysts of law recognized that negligence rules and strict-liability rules provided similar incentives to injurers.²⁷ To see why this is so, recall that a negligence rule gives Xavier an incentive to spend \$40, but not \$8000, on spark arresters that prevent \$4000 fires. But a strict-liability rule does as well. Under a rule of strict liability, Xavier will be liable for *all* cornfield fires that his railroads cause, regardless of their costs and the costs of various precautions. So Xavier will have to pay both (1) the costs of any precautions he takes and (2) the costs of the fires. Given that he faces both these costs, he will have incentives to take efficient precautions. For instance, if spark arresters cost \$40 but fires

²⁵ *Id.* at 33.

²⁶ *Id.* at 32-33.

²⁷ *See id.*

cost \$4000, he will need to pay only \$40 if he installs the spark arresters but \$4000 if he does not. And if spark arresters cost \$8000 but fires still cost \$4000, he will need to pay \$8000 if he installs the spark arresters but only \$4000 otherwise. As a result, he has incentives to take precautions only when they are efficient.

In other words, it is indeed true that a rule of strict liability (compared to negligence) “cannot be justified on the ground that it will induce [injurers] to increase the safety of [their] operations.”²⁸ But it is *also* true that negligence itself cannot be justified in that way (compared to strict liability). The two rules are just as good at providing efficient incentives to injurers, at least when they are applied to the schematic representation of injuries with which we have been dealing.

The early economists, of course, recognized this symmetry. Posner saw it clearly as early as 1973, when he wrote: “Economic theory provides no basis, in general, for preferring strict liability to negligence, or negligence to strict liability, provided that some version of a contributory negligence defense is recognized.”²⁹ This was true, as Posner also recognized fairly early, even after considering the effects of negligence rules and strict-liability rules on the costs of adjudication.³⁰

Accordingly, the early economic analysis of law appears to justify the proposition that *either* strict-liability or negligence rules are better than rules of no liability. Of course, this proposition is uncontroversial; few were seriously arguing in the 1960s and 1970s for a complete absence of tort liability (or for some other standard far less than negligence liability).³¹ But the early economic analysis, perhaps, spelled out clear efficiency-related reasons for this.

Beyond this basic observation, however, that early analysis offered less than commentators sometimes suppose.

²⁸ *Id.*

²⁹ Richard A. Posner, *Strict Liability: A Comment*, 2 J. LEGAL STUD. 205, 221 (1973) [hereinafter Posner, *Strict Liability*].

³⁰ RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 442 (2d ed. 1977) (“No clear-cut prediction of the impact on the aggregate costs of the procedural system of substituting strict for negligence liability emerges from our analysis.”).

³¹ To be sure, some commentators have suggested that historically, tort law imposed liability less often than it does today because of the prevalence of “no duty” rules and other roadblocks to recovery. See, e.g., Robert L. Rabin, *The Historical Development of the Fault Principle: A Reinterpretation*, 15 GA. L. REV. 925, 928-44 (1981).

Posner and others believed,³² and generally still believe,³³ that this early form of economic analysis of tort law demands, from the point of view of social efficiency, that we conceive of the negligence standard in terms of the Hand Formula. To make this point clear, it is important to distinguish two possible normative conclusions from the kind of economic analysis I have discussed in this Part.

First, the analysis purports to address the choice between no liability, negligence, and strict liability. As I have noted, the analysis is correct, at least on its own terms, if it means to suggest that rules of negligence *or* strict liability are superior to a complete absence of tort liability (at least to the extent that the tort regime aims to govern the conduct of injurers).

Second, the analysis purports to give *form* to negligence rules, by observing that under the Hand Formula, negligence rules give injurers incentives to take efficient precaution. But this purported conclusion is only partially correct. It is true that a standard of negligence that requires *less* than the Hand Formula would be inefficient, at least in the schematic situations we have described involving Xavier and Yvonne. For example, if spark arresters cost \$3000, fires cost \$4000, and the legal standard for negligence liability calls Xavier reasonable (or nonnegligent) based on an arbitrary decision that nobody needs to install spark arresters on railroads, then Xavier will not have incentives to install (efficient) spark arresters. To put it differently, such a legal standard would be insufficient because it falls below the standard of the Hand Formula.

But a legal standard that exceeded, or perhaps even just tended to exceed, the Hand Formula could well be efficient. If spark arresters cost \$5000 and fires cost \$4000, but the legal standard for negligence liability calls Xavier reasonable only if he spends \$5000 or more on spark arresters, then he would still prefer to pay for the fires than for the (inefficient) spark arresters. More generally, a negligence standard based on broad social judgments (rather than narrower cost-benefit tests) can give injurers efficient incentives if those social judgments tend to require more precaution than a cost-benefit test would suggest, thereby imposing a standard between negligence and strict liability. For instance, consider a social

³² See Posner, *A Theory of Negligence*, *supra* note 9, at 32 (referring to the Hand Formula as “one of the few attempts to give content” to the negligence standard).

³³ See POSNER, *supra* note 4, at 167-71.

judgment that it is wrong, without a significant overriding justification, to cause preventable fires through industrial activity.³⁴ Based on that judgment and the numeric figures I used above, a tort regime that holds Xavier liable for all preventable fires caused by his railroad can be just as efficient as a tort regime based exclusively on an economically conceived negligence rule.³⁵

Accordingly, what the early economic analysis of law justified, even if all its assumptions about human behavior were correct, is less than is often imagined. On its own terms, the early analysis justified only two propositions: (1) that either negligence or strict liability can give injurers efficient incentives and (2) that an efficient negligence standard needs to be *at least* as strict as the Hand Formula, but it could well be stricter.³⁶ In short, the early economic analysis told us why we need a legal standard that is *at least* as strict as negligence liability, but it told us little more than this. Of course, this point was probably not controversial; it is not clear that anyone had seriously been arguing for modern tort rules that were weaker than the Hand Formula. But while the early economic analysis justifies at least that standard, it is important to recognize that it would be a mistake to take the early analysis as even slightly suggesting that *no greater* a standard would be efficient. The early economic analysis is compatible with many possible negligence standards, as long as they provide liability when a Hand Formula analysis would.

³⁴ Cf. Stephen G. Gilles, *Inevitable Accident in Classical English Tort Law*, 43 EMORY L.J. 575, 576-77 (1994) (suggesting that old English law imposed liability for accidents that could have been prevented, rather than imposing liability only for those that *should* have been prevented).

³⁵ My point in the text, more strictly, is that it is *possible* for a tort regime with a higher standard than that of a negligence regime to be efficient. Not all such regimes are necessarily efficient, however. Cooter and Ulen, in their textbook, give one reason a legal regime with a standard higher than that of negligence could be inefficient: it could encourage excess precaution when the legal standard is slightly higher than that of the negligence standard, because injurers would prefer to pay for slightly excess precautions rather than to pay for lower precautions plus the expected cost of accidents. See COOTER & ULEN, *supra* note 1, at 356. Cooter and Ulen's conclusion appears to be too strong, however. They conclude that "[i]n general, [the] injurer's precaution responds exactly to court errors in setting the legal standard under a negligence regime." *Id.* However, if the legal standard of conduct exceeds what they call the "social optimum" standard sufficiently, it will typically be more efficient for injurers to conform to that standard than to the legal standard. This is a technical point, however, and further details concerning it are beyond the scope of this Article.

³⁶ See Posner, *Strict Liability*, *supra* note 29, at 221; Posner, *A Theory of Negligence*, *supra* note 9, at 32; *supra* note 35 and accompanying text.

With this background, we can now consider, in Part II, more sophisticated modern economic conceptions of negligence rules.

II. THE LIMITS OF MODERN FORMAL ECONOMIC CONCEPTIONS OF NEGLIGENCE

In contrast to the early economic analysis discussed in Part I, the arguments at the center of economic discussions of tort law (and therefore a significant part of tort-law discussions in the United States) are substantially more complex. They aim to address a wider range of problems and to provide more specific recommendations concerning when injurers should or should not be liable for the harms they cause.

Robert Cooter and Steven Shavell have rearticulated, in slightly different ways, the leading modern analysis of negligence rules.³⁷ These rules depend on the recognition that many torts cases potentially demand *bilateral precaution*, which means that both injurers and victims (Xavier and Yvonne) can take steps to reduce the likelihood or severity of accidents. Recall that in Cooter's example,

Xavier can reduce the harm to the corn by installing spark arresters, by running the trains more slowly, or by running fewer trains. In a like manner, Yvonne can reduce the harm by planting her corn farther from the tracks, by planting cabbage instead of corn, or by leaving the fields fallow.³⁸

Not all cases involve bilateral precaution, but it is a feature of many cases—more than most students initially suppose when presented with the idea. A pedestrian afraid of being hit by cars can avoid walking on sidewalks; a homeowner afraid of airplanes falling from the sky can live in a location

³⁷ STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS 187-88 (2004); Cooter, *supra* note 16, at 4.

³⁸ Cooter, *supra* note 16, at 5. The economic model that forms the basis of both Cooter's and Shavell's restatements of the reasons that negligence rules are efficient in cases of bilateral precaution appears to originate with John Prather Brown, *Toward an Economic Theory of Liability*, 2 J. LEGAL STUD. 323, 347 (1973). Brown recognized, interestingly, some features of the fragility of his model. For instance, he observed: "The standard of care is critical, for, when it was changed to [a particular alternative formulation], the identity between equilibrium and optimality was destroyed." *Id.* I develop further reasons the model is fragile in Section II.B, *infra*. For ease of exposition, and to ensure that I respond to arguments in the forms in which they remain influential, I address my discussion of bilateral precaution in the text to Cooter's and Shavell's more recent formulations.

For further notes on the history of the economic analysis that informs modern academic understanding of tort law, see SHAVELL, *supra* note 37, at 192-93.

where fewer airplanes pass overhead; and so forth. In saying that a case involves bilateral precaution, there is no inherent moral or normative judgment; for instance, when economists say that pedestrians can walk more safely, this does not mean, on its own, that a pedestrian *should* do so or is *at fault* for not doing so.³⁹ To an economist, normative judgments depend on costs. Accordingly, a pedestrian should, as a general matter, walk more safely if doing so is less expensive, overall, than asking drivers to drive more safely. Whether it is cheaper might, of course, depend on complicated and possibly subjective social calculations.

In this Part, I first, in Subpart A, lay out the modern economic understanding of tort rules by explaining and elaborating Cooter's and Shavell's analysis. This understanding is based on two principles: (1) that negligence rules provide bilateral threats of liability and (2) that activity levels can generally inform liability decisions. In Subpart B, I respond to the argument that negligence rules provide efficient bilateral threats of liability, demonstrating that the economic models that underlie the leading economic understanding are exceedingly fragile and almost impossible to apply. In Subpart C, I respond to the argument that activity levels can serve as a principled way to assign the costs of accidents between two innocent parties, showing that modern activity-levels arguments are both narrow and unadministrable. In Subpart D, I address further problems that apply generally to the Hand Formula and similar attempts to conceive negligence solely using economic models.

A. *The Modern Understanding: Bilateral Liability Threats and Activity Levels*

1. Bilateral Precaution

The term *bilateral precaution* can refer to two slightly different concepts, and it is important to keep them separate. Consider again the case of Xavier (an injurer) and Yvonne (a victim). In the example we have been using, both Xavier and Yvonne can take precautions against railroad fires, and in that

³⁹ British comedian Jimmy Carr has used this distinction as a source of humor. Discussing drunk driving, he says: "I think the people that make the drink-driving ads should be forced to make an advert aimed specifically at pedestrians, simply saying, 'Pedestrians: Watch where you're going; some of us have had a drink.'" DVD: Jimmy Carr: Comedian (Bwark Productions 2007).

sense the *available* precautions are bilateral. But ordinarily, economic analysis of bilateral-precaution cases assumes that the case has an additional property—namely, that the *optimal* mix of precautions to be taken in a given situation includes some measures within Xavier’s control and others within Yvonne’s control.⁴⁰ To say this differently, given the optimal set of precautions that can be taken, the injurer will be able to take at least one of those precautions more cheaply than the victim, and the victim will be able to take at least one of those precautions more cheaply than the injurer.⁴¹ Throughout this Article, when I refer to cases of bilateral precaution, I refer to the second kind of case.

In bilateral-precaution cases, it is easy to see that rules of either strict liability or no liability will, on economic grounds, come up short. This is because rules of strict liability give incentives for injurers to take precautions (but not victims), and rules of no liability give incentives for victims to take precautions (but not injurers). For example, in the case of Xavier the railroader and Yvonne the cornfield owner, a rule of strict liability would place the whole cost of fires on Xavier, leading him to take precautions (spark arresters, slower trains, or fewer trains) against fires if it is efficient for him to do so; by contrast, a rule of no liability would place the whole cost of fires on Yvonne, leading her to take precautions (growing corn further away from the tracks, growing a crop more resistant to fires, or not growing anything) against them if it is efficient for her to do so. But in neither case will *both* Xavier and Yvonne—assuming they are purely rational and self-interested, and assuming that they have no other relevant incentives—take precautions against fires.

The modern formal economic conception of negligence rules in tort law attempts to address this problem in cases of bilateral precaution—i.e., those where the only efficient mix of precautions would come from both Xavier and Yvonne, the injurer and the victim. In short, the modern understanding of

⁴⁰ See, e.g., COOTER & ULEN, *supra* note 1, at 341 (“[W]e consider the case in which *both* the victim and injurer *can* take precaution, and efficiency *requires* both of them to take it. We call this condition the assumption of *bilateral precaution*.”); SHAVELL, *supra* note 37, at 182-83 (“[E]xamples can obviously be constructed in which it is optimal only for injurers to take care or only for victims to take care (or for neither to do so). These possibilities are not the focus [of the bilateral-precaution discussion].”).

⁴¹ See Cooter, *supra* note 16, at 6 n.16 (“Some efficient precautions may cost less when taken by one party or the other. Precaution is bilateral when at least one such precaution for each party exists.”).

negligence rules is that they provide efficient incentives to *both* parties by making both think that they may be liable if they don't live up to their efficient standard of care. As Cooter says:

[T]he paradox [of encouraging both Xavier and Yvonne to behave efficiently] can be resolved by adopting fault [i.e., negligence] rules that assign responsibility for harm according to the fault of the parties. To illustrate, a simple negligence rule requires the victim to be compensated by the injurer if, and only if, the latter is at fault. Under a simple negligence rule, Xavier will satisfy the legal standard in order to avoid liability. Thus, if the legal standard corresponds to the efficient level of precaution, Xavier's precaution will be efficient. Since Yvonne knows that she bears residual responsibility, she internalizes the costs and benefits of precaution; therefore, her incentives are efficient. Thus, if the legal standard of fault corresponds to the efficient level of care, both parties will take efficient precaution.⁴²

Shavell puts it similarly:

As in the unilateral model, if the courts choose due care to equal the socially optimal level [i.e., if negligence is set via the Hand Formula], then injurers will be led to take due care. Victims too will be induced to take the optimal level of care because they will bear their losses if injurers take due care. (Drivers will be led to take due care; and knowing that they will bear their losses, bicyclists [that the drivers might hit] will decide to take appropriate care.)⁴³

In other words, a negligence rule appears to encourage injurers to take whatever precautions are efficient for them to take. But then, because victims will expect injurers to take these precautions and thus avoid liability, the victims will fear that they're going to suffer the costs of accidents themselves. As a result, the victims, too, will take efficient precautions.

For those without economic training, what economists mean when they refer to some precautions that victims can take may seem counterintuitive. Why is it a "precaution" for Yvonne, for example, to avoid growing anything in her fields at all? The answer is that by not growing corn, she prevents any social waste that comes from investing in corn. A fire to an empty field might well cause no "harm." By not planting anything in her field, Yvonne is able to remove the possibility that fires caused by trains will cause her to waste money in growing corn that will simply be burned down. Of course, by not planting anything, Yvonne presumably suffers some loss

⁴² Cooter, *supra* note 16, at 6-7.

⁴³ SHAVELL, *supra* note 37, at 185-86.

because her field is not being put to productive use. But to an economist, this loss is exactly the same kind of loss that Xavier himself would suffer by having to install spark arresters (or to run fewer trains). In other words, Xavier and Yvonne interact to cause a social loss, even if Xavier is the one whose sparks cause fires in Yvonne's fields. Stripped of all moral dimensions and other kinds of social judgments, both parties simply face potential costs from two sources: (1) planning in advance of an accident in order to reduce the expected harm from it, and (2) either harm from the accident (in Yvonne's case) or a requirement to pay for harm from the accident (in Xavier's case).⁴⁴

An example of the economists' overall argument about negligence rules' effects on bilateral precaution may be in order. Suppose that in the case of Xavier and Yvonne, two precautions are said to be optimal: as the cheapest mix of accident-avoidance and harm-avoidance, imagine that it is efficient (1) for Xavier to install spark arresters at a cost of \$1200 and (2) for Yvonne to avoid planting corn within ten feet of the railroad track, and instead to install nonflammable rubber in that space at a cost of \$800 (which includes both the cost of the rubber and the cost of the forgone corn). Why would such a mix be optimal? Just for the sake of the hypothetical, suppose that fires are hugely expensive, causing \$40,000 worth of damages. Now, also suppose that if Yvonne didn't leave a buffer of ten feet, Xavier would have to install super-safe spark arresters at a cost of \$8000 in order to prevent fires. But suppose, conversely, that if Xavier didn't install any spark arresters, Yvonne would have to leave a buffer of fifty feet, at a cost of \$6000 (as a result of the greater amount of forgone

⁴⁴ Brown's formalization of the case he describes is a clear and helpful aid to the intuitions that underlie economists' understanding of the bilateral tort model:

Consider a small device, a black box, which is attached to some otherwise useful object such as a railway crossing, an airplane, or a sidewalk. The only function of the device is to emit a bill for a large amount of money from time to time, so we shall call it a liability generator. . . .

On the liability generator are two controls, X and Y. . . . Increasing either or both increases the probability that the accident will be avoided Examples of what will be meant here by controls are built-in safety devices and careful driving in the railway crossing case, defect-free radar and careful flying in the airplane case, and shoveling snow and careful walking in the sidewalk case.

Brown, *supra* note 38, at 324. The description shows the economic symmetries between victims and injurers in a case of bilateral precaution.

corn), to prevent the huge costs of fires. Given this mix, the particular efficient state of affairs is for both Xavier and Yvonne to spend some money on precautions.

According to Cooter and Shavell, under a negligence regime both Xavier and Yvonne can be encouraged to take these precautions, in this case. This is because, if the legal standards for negligence are set correctly, they believe both Xavier and Yvonne will be afraid of suffering the \$40,000 harm from fires if they don't take the relatively cheap—and more importantly, optimal—precautions available to them. So they both will do so, and together they will create a social surplus (over other scenarios in which money is wasted either on precautions or on fires).

It sounds attractive enough. And on the surface, the models appear to work out the way that Cooter and Shavell suggest. But beneath the surface, the models face significant technical problems that prevent them from being applied determinatively in at least many kinds of tort cases. I will describe those problems in Subpart B, *infra*. But first, it will be helpful to explain the other central features of the modern understanding of liability rules in tort law.

2. Residual Liability

Even if negligence rules could be implemented in a way that encourages efficient bilateral precaution, there remains an incompleteness in the model: what happens when both parties live up to their efficient standard of care? Just as under the early, straightforward economic models, which suggested that both negligence rules and strict-liability rules could encourage efficient precaution, the bilateral-precaution model is consistent with assigning the cost of accidents either to injurers or to victims when both act reasonably.

To see why this is so, consider that an ordinary rule of negligence is symmetric with a rule of “strict liability with a defense of contributory negligence.”⁴⁵ If a rule of simple negligence encourages an injurer to take precautions, and then encourages a victim to take precautions because the injurer avoids liability and leaves the victim holding the bill, then a rule of strict liability combined with contributory negligence can do something very similar, but opposite in one important

⁴⁵ COOTER & ULEN, *supra* note 1, at 346; *see also* SHAVELL, *supra* note 37, at 184-87.

way. Specifically, it can serve as a threat of liability to both parties, causing both of them to behave efficiently, but then leave the injurer (instead of the victim) responsible for any harms that occur.⁴⁶

So, for instance, in our last example, Xavier and Yvonne might both be encouraged to take small precautions. But if a large spark occurs from the railroad and causes a fire anyway, Yvonne will suffer the harm under a negligence rule: both parties met their standard, and Yvonne cannot claim in court that Xavier was negligent. Under a strict-liability regime that incorporates a defense of contributory negligence, by contrast, Xavier will have to pay for Yvonne's harm in such a case.

Given this symmetry, how can we distinguish among the potential rules? Indeed, there are not just two possible rules. In addition to (1) negligence and (2) strict liability with a defense of contributory negligence, other possibilities that lead to similar results (because they all achieve bilateral liability threats in theory) are (3) negligence with a defense of contributory negligence, and (4) comparative negligence.⁴⁷ Perhaps surprisingly, the economic analysis of all these rules—including comparative negligence—is essentially the same, at least in the basic cases I have outlined here.⁴⁸ The supposed incentives are the same; the only difference is who ends up with the cost of accidents that occur when everyone behaved efficiently (nonnegligently).

Nothing in the analysis of bilateral precautions lets us distinguish, then, between these various possible tort regimes.⁴⁹ A different kind of analysis needs to serve that role, and in the

⁴⁶ The liability threats under such a regime work as follows: the victim will take efficient precautions knowing that if she doesn't, she will be held liable because the injurer will be able to show that the victim was contributorily negligent. But then the injurer will fear liability himself, because the victim has behaved properly and the rule is one of strict liability. So the injurer will take efficient precautions too. The mechanism by which the incentives operate is simply the mirror image of a negligence regime's.

⁴⁷ See COOTER & ULEN, *supra* note 1, at 344-47; SHAVELL, *supra* note 37, at 184-89. Both Cooter and Shavell explain how these rules achieve similar bilateral liability threats, which should not be a surprise: the mechanism is essentially the same under all these regimes.

⁴⁸ For a more complete economic analysis, see Robert D. Cooter & Thomas S. Ulen, *An Economic Case for Comparative Negligence*, 61 N.Y.U. L. REV. 1067, 1070-71 (1986) (arguing that comparative negligence is superior to other negligence-based rules only when parties face particular informational limitations).

⁴⁹ Cf. COOTER & ULEN, *supra* note 1, at 348 ("[T]he . . . model [of bilateral liability threats] provides a policy reason to prefer a negligence rule whenever precaution is bilateral. The simple model does not, however, provide a reason for preferring one form of the negligence rule to another.").

modern economic understanding of tort law, it comes from an analysis of activity levels.

3. Activity Levels

To figure out whether a rule of negligence is more efficient than a rule of strict liability (with a defense of contributory negligence)—which is to say, to figure out who should bear the cost of an accident when both injurers and victims have behaved reasonably (nonnegligently)—economists have turned to an understanding of *activity levels*, as pioneered by Shavell.⁵⁰ The concept is simple, though perhaps unfamiliar to most lawyers. As Shavell originally put it:

By definition, under the negligence rule all that an injurer needs to avoid the possibility of liability is to make sure to exercise due care if he engages in his activity. Consequently *he will not be motivated to consider the effect on accident losses of his choice of whether to engage in his activity or, more generally, of the level at which to engage in his activity*; he will choose his level of activity in accordance only with the personal benefits so derived. But surely any increase in his level of activity will typically raise expected accident losses (holding constant the level of care). Thus he will be led to choose too high a level of activity; the negligence rule is not “efficient.”⁵¹

Consider the activity of driving.⁵² Under a negligence regime, drivers are encouraged to drive safely (because if they don't, they have a greater risk of being held liable for their dangerousness). If they drive safely—that is, if they are confident that they will always be able to drive safely—then they know that they won't be held liable for car accidents. But if their very decision to drive increases the likelihood of accidents—after all, more cars will be on the road if more people drive, there will likely be more congestion, and perhaps some accidents to pedestrians arise even when both drivers and pedestrians behave safely—then they will drive *too much*, even while driving safely.

Of course, it is reasonable to wonder why courts would not simply judge drivers negligent if they drive too often, if indeed excessive driving increases the risks of accidents. For instance, a driver out “on a mere whim”⁵³ could be held liable

⁵⁰ See Shavell, *Strict Liability*, *supra* note 10, at 2.

⁵¹ *Id.*

⁵² This example—now standard in the activity-levels literature—is drawn from Shavell. See *id.* at 2-3.

⁵³ *Id.* at 2.

more readily than an ambulance on an urgent errand, even if both drivers were handling their vehicles with similar care. Or a driver on a trip with virtually no social utility could be judged negligent, when he gets into an accident, merely for being out on the road. But courts do not make determinations like these in practice, and in general it would be difficult for them to do so. I return to problems concerning the distinction between individual choices and activity levels in Part II.D.1, *infra*.

Modern law-and-economics scholars relate an understanding of activity levels to liability judgments in the following way: they argue that, when both the injurer and the victim behave nonnegligently, liability should be assigned to the party whose choices about activity level have a greater chance to reduce accidents efficiently. Thus, for instance, Cooter and Ulen write as follows: “Usually one party’s activity level affects accidents more than the other party’s activity level. Efficiency requires choosing a liability rule so that the party whose activity level most affects accidents bears the residual cost of accidental harm.”⁵⁴ Shavell puts it similarly, although perhaps in a way that accommodates broader considerations: “Strict liability [with a defense of contributory negligence] will result in greater social welfare [than a rule of negligence] if it is more important for society to control injurers’ levels of activity than victims’.”⁵⁵

As an example, consider the kinds of “abnormally dangerous” activities described in the *Restatement (Second) of Torts*.⁵⁶ In destroying buildings with dynamite, it appears (at least on the surface) that those using dynamite influence the likelihood of harm more via their choice of activity level than those who operate stores on nearby streets. To summarize the modern economic approach to tort cases, analyzing this case would work as follows: (1) both store owners and blasters can take precautions against harm from dynamite in a variety of ways, and the case is accordingly one of bilateral precaution; (2) as a result, in terms of basic precautions, any negligence-based rules (including either (a) negligence or (b) strict liability with a defense of contributory negligence) will be optimal; (3) to choose between them, we note that blasters’ activity levels influence accident costs more than store owners’; (4) as a

⁵⁴ COOTER & ULEN, *supra* note 1, at 349.

⁵⁵ SHAVELL, *supra* note 37, at 202.

⁵⁶ RESTATEMENT (SECOND) OF TORTS § 520 (1977).

result, the efficient rule is strict liability with a defense of contributory negligence.

Having laid out this modern economic understanding, my goal now is to demonstrate why it comes up short if its goal is to justify or recommend particular legal rules. In Subpart B, I address the problems with the basic bilateral-precaution model, arguing that it insufficiently justifies negligence rules in the first place. In Subpart C, I demonstrate that even if we assume that negligence rules (including strict liability with a defense of contributory negligence) are efficient, activity-level arguments—at least as understood economically—are of little help in choosing among them. As a result, economics remains of perhaps surprisingly little help in determining when to assign tort liability. In Subpart D, I address a variety of other issues that affect the applicability of the reigning economic models, showing that fundamentally noneconomic social judgment, rather than discrete economic cost-benefit tests, would be needed even if the models otherwise worked as economists intend.

B. The Limits of Models of Bilateral Liability Threats

As I have noted, there are two central features of the modern economic understanding of negligence rules: (1) bilateral threats of liability and (2) activity levels as a mechanism to decide who bears residual liability when all parties behave optimally.⁵⁷ In this Subpart, I describe several fundamental problems with the first of these pillars.

1. A Mathematical Demonstration of the Prevailing Economic Model

To do this, it will be necessary to consider a little more deeply, and mathematically, the formal models at issue. For ease of exposition, I will draw in part, at first, from a particularly clear summary of these views by Cooter and address variations of this model later.⁵⁸

Consider, again, the example of Xavier (a railroader) and Yvonne (a cornfield owner). Formally, the total social cost from fires from Xavier's railroad that burn Yvonne's corn are

⁵⁷ See *supra* Part II.A.

⁵⁸ The material in the first part of this section is, accordingly, based on Cooter. See Cooter, *supra* note 16, at 8-11.

$$SC = x + y + p(x,y)a$$

In this formula, SC is the total social cost, x is the cost of Xavier's precautions (like installing spark arresters), y is the cost of Yvonne's precautions (like growing less corn), $p(x,y)$ is the likelihood of a fire, and a is the cost if there is a fire.⁵⁹

A few features of this formalization are worth specially noting. First, the cost of accidents when they occur, a , is held constant.⁶⁰ This is a reduction, and most reductions threaten the ultimate applicability of formal models; however, this particular reduction is not one I need to challenge for the purposes of this Article, so I accept it in the remaining discussion.

Second, more importantly, the probability of fire-related accidents is expressed as $p(x,y)$. Here, p is a function—a mapping of some values to others. The important feature of the way that the probability of fires is expressed, here, is that it depends on both x and y —that is, on the precautions that both Xavier and Yvonne take. If the probability were expressed simply as $p(x)$ or $p(y)$, it would depend wholly on Xavier's or Yvonne's precautions, respectively, and the case would therefore be one of unilateral precaution. That the probability is expressed as $p(x,y)$ means the case is potentially one of bilateral precaution. (I say "potentially" because the optimal x or y could still be zero.)

To summarize so far, the total social costs (SC) are the sum of Xavier's precautions, Yvonne's precautions, and the expected costs (the probability times the magnitude) of the fire's harms. For the economists who have set forth these models, the goal is simply to reduce SC through tort rules.⁶¹

Ultimately, social costs in this case depend only on x and y . That is, given particular levels of precaution by Xavier and Yvonne, there is an associated social cost (which consists, again, of the costs of those precautions and the expected harm from fires, which itself just depends on the precautions that Xavier and Yvonne take). To economists, accordingly, the goal

⁵⁹ Cf. *id.* at 8. I have simplified Cooter's formula somewhat in ways that do not affect the argument. Specifically, for Cooter, $p(x,y)$ is the likelihood that an accident will *not* occur, whereas in my example, $p(x,y)$ is the likelihood that it *will* occur. Accordingly, Cooter uses $(1 \cdot p(x,y))$ to represent the likelihood of an accident.

⁶⁰ Cf. *id.* at 8 n.24.

⁶¹ See *id.* at 8 ("Efficiency is achieved when social costs are minimized.").

is to find rules that give parties incentives to adopt an optimal pair of values for x and y . Call these optimal values x^* and y^* .⁶²

Now, consider the private costs that Xavier and Yvonne face individually. These private costs will depend on tort law's liability rules, because Xavier and Yvonne can respond to tort law's incentives. The central conclusion of the economists' models is that negligence rules (including, again, rules of strict liability with a defense of contributory negligence)⁶³ create efficient incentives for Xavier to adopt x^* as his level of precaution and for Yvonne to adopt y^* as hers.

Consider first Xavier, the potential injurer. Under a negligence rule, Xavier's costs can be separated into two distinct cases: (1) if he pays for enough precaution to satisfy tort law's standard, then his only cost is that of the precaution, because he won't have to pay for any of the costs of fires; (2) if he does *not* pay for enough precaution, then his cost is that of whatever precautions he does pay for, plus the costs of fires, because tort law will hold him liable for the damages from the fire. In the first case, we can express Xavier's costs simply as x . In the second, Xavier's costs are $x + p(x,y)a$.

And in this formulation, we reach the first central stumbling block of the economic model. Because this is a case of bilateral precaution, Xavier's costs in this case depend in part on the precautions that Yvonne adopts (y). This means that, if all we know is x , there is no way to determine Xavier's costs. If we admit that we do not know what Yvonne's costs might be, there is little more we can say about Xavier's cost. Economists, accordingly, specify one more piece of information. They state, as an example, that "Yvonne's precaution is held constant at the efficient level ($y = y^*$)."⁶⁴

This additional assumption, though it may appear minor, severely undermines the model's applicability to real tort cases. But before I explain its problems, it will first help to understand what it allows the formal model to do. By holding

⁶² Cf. *id.*

⁶³ In the remainder of this section, I will use simple negligence rules as an example of the class of rules that includes strict liability with a defense of contributory negligence, negligence with a defense of contributory negligence, and comparative negligence. This is a simplification without any loss of generality; the economists' arguments at stake in this section, and my responses to them, treat all models in the same way. See *supra* Part II.A.2.

⁶⁴ Cooter, *supra* note 16, at 9. Shavell's formulation is similar, although not identical (for reasons I explain *infra* Part II.C.1): "[I]njurers will exercise optimal care given that victims take due care, because then injurers will be liable for accident losses." SHAVELL, *supra* note 37, at 184-85 (emphasis added).

Yvonne's costs constant at the optimal level, Xavier's costs can now be expressed solely as a function of his own precaution. Assuming that tort law's negligence standard (as applied to Xavier's behavior) is optimal, Xavier's costs are: (1) x if $x \geq x^*$, and (2) $x + q(x)a$ if $x < x^*$, where $q(x)$ simply represents the cost of accidents given Xavier's level of precaution x , assuming Yvonne behaves optimally.

Essentially, Xavier gets to pick between choices 1 and 2, based on the level of x he chooses. Because $x + q(x)a$ is greater than x (because $q(x)a$ is positive), Xavier would prefer to pay only x . The way for him to do this is to choose his level of precaution x to equal x^* , the optimal amount of precaution and the legal negligence standard for him. Consider his choice in the following way: if he chooses less precaution than x^* , he will have to pay for accident costs *plus* whatever precaution he takes; if he choose a level of precaution equal to x^* , he avoids liability and pays only x^* . Therefore, he will (under the model) choose a precaution equal to x^* .⁶⁵

Yvonne's decisions are the mirror image of Xavier's. When considering her costs, the economists tell us similarly to hold Xavier's precautions constant at the optimal level.⁶⁶ Then, she can choose either a cost of y or $y + p(x^*, y)$ depending on whether y is less than y^* (the legal standard for *her*, based on what precautions are socially optimal for her to take). For reasons that track the discussion of Xavier's incentives, the economists conclude that Yvonne will adopt the optimal level of precaution, y^* . Accordingly, negligence rules are said to give both Xavier and Yvonne efficient incentives; the economists expect that Xavier will choose x^* and Yvonne will choose y^* , their respectively optimal levels of precaution.

2. Limits of the Model

As I have suggested, however, the model is flawed, or at least limited in its applicability to many kinds of cases. The internal flaws of the model—as opposed to those that highlight the model's incorrect or incomplete assumptions about human

⁶⁵ For a more mathematical elaboration of this point, see Cooter, *supra* note 16, at 9-10. For an even more formal proof of essentially the same conclusion, see generally the original discussion in Brown, *supra* note 38.

⁶⁶ See SHAVELL, *supra* note 37, at 185 (“The specific reasoning [for victims] is analogous to that in the explanation . . . of why injurers will take due care under the negligence rule.”); Cooter, *supra* note 16, at 10 (premising the conclusions for the victim's case on the assumption “that the injurer is nonnegligent”).

behavior—derive from a feature to which I drew attention in the previous section: namely, that Yvonne's behavior is held constant when considering Xavier's, and that Xavier's behavior is held constant when considering Yvonne's.

To put this more succinctly, Xavier's optimal behavior depends on Yvonne's, and Yvonne's depends on Xavier's. Or, more formally, x^* depends on y^* , which depends on x^* . This may seem circular, and in a sense it is, but the internal problem with the economic model isn't precisely that it is *logically* circular. Variables can depend on each other, in this sense, without collapsing a mathematical model. Indeed, this kind of codependence between variables underlies much of game theory: in a game, the actions of one party influence the actions of another, which in turn can influence the actions of the first, until an equilibrium is reached.⁶⁷

The central problem comes instead from what the mirror-image dependence demands, in this particular model's case. The only way to determine what the legal standard for Xavier (x^*) should be is to know what the legal standard for Yvonne (y^*) should be, and vice-versa. Accordingly, before we can set the particular negligence rules that govern *either* Xavier and Yvonne, we have to know what the optimal behavior is for *both* of them. That knowledge must come as a package, and if the economic model is to work, we must use it to set the standards for both parties.

More precisely, for the economic model even to get off the ground, we need to imagine that Xavier and Yvonne can determine such optima *ex ante* and also that they expect that a court analyzing the situation *ex post* will be able to infer the same optima. If Xavier and Yvonne cannot determine the optima themselves, they have no way to plan their behavior accordingly. And if they do not expect courts to be able to determine the proper standards *ex post*, then being purely rational and selfish, they will have no reason to plan their behavior in view of the correct legal standards.⁶⁸

⁶⁷ For a short introduction to game theory and the analysis of equilibria, see COOTER & ULEN, *supra* note 1, at 32-42.

⁶⁸ Shavell clearly outlines this requirement for his argument: “[T]o ascertain the optimal level of due care for just one party, a court must generally determine (if only implicitly) the optimal level of care for the other as well, because the optimal level of care for one party will in principle depend on the other's cost of, and possibilities for, reducing risk.” SHAVELL, *supra* note 37, at 188. “This latter point,” Shavell admits, “makes the comparison of liability rules with respect to their ease of application different from what it might at first seem to be.” *Id.* Shavell also recognizes that

The central difficulty with the formal model of bilateral precaution arises from the impracticality of knowing in advance, with perfect accuracy, what the optimal costs and benefits are for parties like Xavier and Yvonne. Economists, to be sure, do not think parties or courts (or regulators) have access to this kind of perfect knowledge, or that parties in Xavier and Yvonne's positions will have access to perfect information about one another.⁶⁹ But if the model is to be applied to tort cases in practice, the legal economists' arguments implicitly depend on the belief that the model nonetheless provides a useful idealization of the world, and that minor variations from the model's assumptions will only slightly degrade the model's normative power.⁷⁰

That may be true of some models, but it is not true of this one—at least not at the level of generality at which formal deductions about law operate. Recall that the reason a situation involves bilateral precaution in the first place is that the injurer and the victim face different costs in the precautions they might take.⁷¹ (Otherwise, the case could more easily be treated as one of unilateral precaution, in which a single party takes all the care needed to reduce the costs of accidents to an optimal level.) Accordingly, a small change in the precautions Xavier *actually* takes can mandate a very large change in the precautions Yvonne should *efficiently* take, and vice-versa.⁷² As an example, Yvonne's optimal behavior (from both her perspective and an overall social one) might look very different depending on whether Xavier does or doesn't install spark arresters on his railroads, even if those spark arresters are very cheap.

"courts must generally consider the entire tableau of costs and effectiveness of care for the two parties to determine optimal care for either." *Id.* at 188 n.17. One way of understanding my central argument in the text about the model's fragility is that this "entire tableau of costs and effectiveness" need not, in any situation, exhibit any regularity or predictability. *Id.* Minor changes to it (based on, for instance, small changes in what courts expect injurers and victims to do) can radically change courts' beliefs about which precautions are optimal.

⁶⁹ See, e.g., Brown, *supra* note 38, at 343-47 (analyzing, as a variation on the economic model that came to underlie economists' modern understanding of tort law, the effects of informational limitations for injurers and victims).

⁷⁰ Cf. COOTER & ULEN, *supra* note 1, at 347 (suggesting, with specific reference to the bilateral-liability-threat model of tort law, that "[i]t is usually best to build theory from clean results and then handle any messy results as exceptions").

⁷¹ See *supra* notes 40-41 and accompanying text.

⁷² More formally, a small change in x can require a large change in y^* , and a small change in y can require a large change in x^* .

As a result, the central internal problem with the prevailing formal models of negligence is that they are untenably *fragile*: they do not resist minor modification to the parties' behavior. To say this differently, the models are premised on a theoretical perfection, and the slightest variation from this perfection can make their equilibrium collapse entirely, rather than degrade gracefully. Instead of providing an approximation of the real world, they threaten to provide virtually nothing in the real world.

As an example, even if we can narrow x^* down to a relatively small range (say, a rough projected expenditure on a few different kinds of spark arresters), this may not be sufficient to determine what y^* is. In the general case, we need *full* information about x^* in order to specify y^* , and vice-versa. There is little opportunity to reach a second-best result: admitting that we are not sure of x^* means we cannot be sure of y^* , and nothing in the economic model guarantees that this uncertainty will not spiral out of control, so that we are no longer even roughly sure of x^* .⁷³ Even a very good prediction based on aggregate or generally constant behavior is insufficient to ensure that the model reaches what economists call a convergent—that is, a stable—result. Slight changes to one party's precaution can have unpredictable effects on those that the other parties should take.

To be clear, I am not arguing that the model cannot work in any case, no matter how stylized; my criticism is that the model cannot be applied to the general case and that therefore, it cannot justify negligence rules as a general matter. In the general case, even when an injurer can estimate victims' precaution reasonably well and a victim can represent injurers' precaution reasonably well, the model cannot tell us what to do.

Moreover, even if we assume that injurers, victims, and courts have perfect information about what precautions would be socially optimal given the expected probability and harm from accidents, there are additional reasons that the model's fragility is triggered as soon as it is applied to any real case.

⁷³ Of course, it is unfair to expect an economic model to provide stronger conclusions than are justified by the information available to us in the real world. My objection to the model is not that it cannot yield an optimal result given suboptimal information. Rather, my argument is that the model cannot work at all in the general case, in the way it was intended, without perfect information.

For one thing, even if an individual party is perfectly rational and has perfect information, he or she must accommodate the possibility—even in just a probabilistic sense—that other parties will not behave perfectly rationally or have perfect information. In other words, even if both parties turn out to be perfectly rational and fully informed, they would have to account for the possibility that other people are not.

As an example, consider a simple game in which a group of people are asked to choose numbers from zero to one hundred.⁷⁴ The group's numbers will be averaged. The winner of the game is the member of the group who chose a number closest to half the group's average. In a game populated only by fully rational players, the optimal choice would be zero.⁷⁵ But in practice, even a fully rational agent would not choose zero, because he or she would expect error or irrationality in other people's choices.⁷⁶ In other words, it is fully rational to expect irrationality or lack of information in others, at least probabilistically.

The problem that this observation poses for the bilateral-liability-threat model is that it cannot in practice be efficient to dictate efficient legal standards, x^* and y^* , based on the presumption that both parties will expect the other to behave perfectly. Xavier and Yvonne cannot (and should not) plan their behavior in view of that assumption. But once we recognize this, the model unravels; again, without shared knowledge of the optimal pair of values for x^* and y^* , neither can be set in the general case. We can try to guess what precautions Xavier will in fact choose, and set the legal standard for Yvonne accordingly; then, we can try to guess what Yvonne will in fact choose, based on this standard, and set the standard for Xavier accordingly. But then this change in the standard will alter Xavier's behavior, which in turn will

⁷⁴ This example, called a "p-beauty contest game," is drawn from HERVE MOULIN, *GAME THEORY FOR THE SOCIAL SCIENCES* (2d ed. 1982); see also Avinash Dixit, *Restoring Fun to Game Theory*, 36 J. ECON. EDUC. 205 (2005) (discussing this game from a pedagogical perspective).

⁷⁵ If all players in the group chose one hundred, the best choice would have been fifty. Members of the group, knowing this, could all choose fifty. But then, all players could figure out that fifty would be the average, so they would want to choose twenty-five, and so on.

⁷⁶ For experimental results of this game in practice, see Rosemarie Nagel, *Unraveling in Guessing Games: An Experimental Study*, 85 AM. ECON. REV. 1313 (1995) (demonstrating that people do not behave as if they assume everyone else were perfectly rational).

alter Yvonne's behavior, and nothing guarantees that an efficient equilibrium will result in the general case.

There is another, perhaps simpler, way to express this problem and related ones: we can analyze the situation from the perspective of the parties, rather than the policymakers setting x^* and y^* . For example, Cooter is rightly concerned, throughout much of his analysis of legal rules, with what he calls the "paradox of compensation"⁷⁷—the notion that when efficiency requires "double responsibility at the margin"⁷⁸ from multiple parties, there is no single efficient legal rule that provides the right incentives to everyone. This was, in short, the problem we saw with rules of strict liability and no liability earlier: strict liability might give injurers efficient incentives, but it leaves victims free to take no precautions at all, at least in theory, because all their harms are compensated by injurers. Rules providing for no liability do the reverse: they give victims incentives to take precautions, because they bear the costs of harm, but injurers are free to do as they please. In defending the bilateral-liability-threat model of negligence rules, economists have offered it as a solution to this "paradox" of compensation.

But the problem is that a mere *threat* of liability cannot solve this paradox. Even a fully rational Xavier, or Yvonne, will know under a negligence regime that there is some chance they will avoid liability and some chance they will not. For instance, suppose that the legal standard sets x^* (Xavier's efficient precautions) to \$80, which might correspond to a requirement to install spark arresters. The economic model that justifies negligence rules suggests that Xavier will choose \$80 as his level of precaution because he fears liability if he doesn't.⁷⁹ But this liability is not certain under all negligence rules, even if perfect enforcement of the law is assumed. For instance, under a rule of negligence with a defense of contributory negligence (unlike a rule of pure strict liability), Xavier knows that there is some chance that Yvonne will not meet her standard of liability (y^*). As a result, it may not be efficient for Xavier to spend \$80 on precautions in all cases, even if he is fully

⁷⁷ Cooter, *supra* note 16, at 3-4.

⁷⁸ *Id.* at 4.

⁷⁹ See SHAVELL, *supra* note 37, at 184-85; Cooter, *supra* note 16, at 9-10.

rational.⁸⁰ Instead, he may spend less under some conditions because he may expect that he will sometimes be able to avoid liability altogether (because of Yvonne's own negligence).⁸¹ But given this possibility, then—as before—the y^* that courts have chosen for Yvonne may not be socially optimal in a second-best sense, which in turns means that the x^* that courts have chosen for Xavier may not be optimal in that sense, and so on. The model threatens to unravel, again, because of the slightest perturbation.

Moreover, given also the probabilistic nature of the harms in question—that is, fires in Xavier and Yvonne's case are not certain but merely possible—construction of purportedly optimal standards is made even more difficult. As Shavell notes, the economic models in question are meant specifically to address probabilistic harms.⁸² But this means that x^* and y^* may change slightly as a result of new information that comes to light; for instance, if the likelihood of a train-related fire for the next year is estimated to be 1 in 200,000 on March 22, the likelihood may go down to 1 in 240,000 as the result of greater-than-expected ambient humidity (which makes fires less likely) in late March. This kind of minor perturbation in probabilities would not pose a significant problem for a robust model, but given the kinds of

⁸⁰ This result is well understood in the economic literature. See, e.g., John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965 (1984).

⁸¹ In other words, there is no reason to assume that Xavier will assume that the probability of Yvonne's compliance with her legal standard (y^*) is equal to 1. To elaborate the discussion in the text, consider that the formal economic model guarantees that Xavier's total cost when he chooses $x < x^*$ will be at least as great as his cost when he chooses $x = x^*$, because if the total social costs of accidents were lowest at a point smaller than x^* (given a constant y^*), then x^* itself ought to be lowered to that point. But nothing guarantees that Xavier's *expected* costs, in view of the probability of Yvonne's compliance with her legal standard y^* , are not lower when $x < x^*$. These expected costs are essentially a weighted average between x and $(x + p(x,y)a)$ when $x < x^*$ (weighted by Xavier's estimated probability of Yvonne's compliance), and such an average might be smaller than x^* .

Of course, if both parties assume the other will make calculations of this kind, the analysis becomes even more complicated. The particular expected results depend on a variety of case-specific features and cannot be derived in the abstract, and there is no reason to assume it will result in an efficient equilibrium in the general case.

This situation shares some features with a continuous iterated prisoner's dilemma. For an interesting analysis of that phenomenon from a biological perspective, see generally Stephen Le & Robert Boyd, *Evolutionary Dynamics of the Continuous Iterated Prisoner's Dilemma*, 245 J. THEORETICAL BIO. 258 (2007).

⁸² SHAVELL, *supra* note 37, at 177 ("We will assume that accidents and consequent liability arise probabilistically.").

fragility in the model that I have described, it is hard to be confident that a stable justification for purportedly efficient standards (that is, for specific values of x^* and y^*) will exist in many cases. If nothing else, the economic model's dependency on full knowledge of both parties' efficient precautions makes it less likely that either side's efficient precautions can be specified in any given case.

In short, while some idealizations—including economic ones—can serve as useful approximations of the world from which we can later veer, the bilateral-liability-threat model underlying modern economists' view of tort law is untenably fragile if its goal is to justify application of negligence standards to real cases. An idealization cannot justify specific policy propositions in law when the slightest change in information or behavior threatens chaotic results.⁸³ Far from explaining the central feature of Anglo-American tort law, the purported economic justification of the negligence standard provides very little justification for the rule in practice.

3. Alternative Formulations of the Economic Model

My observations and analysis in the prior section addressed a *continuous* version of the model—that is, one that allows precautions to vary to any possible levels, so that a level of precaution might be set to \$74.82, or \$100.64, and so on.⁸⁴ We might alternatively formulate the bilateral-liability-threat model as discrete rather than continuous, which is to say that we might imagine (say) four particular on-or-off precautions that the parties might take. For instance, in a particular situation, we might observe that Xavier has a choice of two spark arresters, one that costs \$80 and reduces the likelihood of fires by fifty percent, and one that costs \$240 and reduces the likelihood of fires by ninety percent. If these are Xavier's

⁸³ I mean “chaotic” in both a lay sense and a technical sense. For more information on chaos theory, which characterizes (among other things) the way in which small changes in the inputs to a system can cause wild swings in its output, see Robert Bishop, *Chaos*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2008), available at <http://plato.stanford.edu/entries/chaos/>.

⁸⁴ Strictly speaking, dollar values are not in practice continuous because they do not extend beyond two decimal places (to cover cents). But because the value of a cent is so small, familiar statements of value in dollars and cents are for most practical purposes better treated as continuous rather than discrete.

only choices, his precaution is said to be *discrete* (or *discontinuous*) rather than *continuous*.⁸⁵

Shavell's most recent statement of the purported economic basis of tort law is framed in largely discrete terms. For instance, though his argument tracks the one I have already laid out, and though it is formalized in the same way,⁸⁶ his particular example involves three possible levels of precaution: "none," "moderate," and "high."⁸⁷

The reason that the difference between discrete and continuous models may be important here is that discreteness, versus continuousness, may save a model from its own fragility. In other words, if Xavier has only three levels of precaution available to him, then minor perturbations in probabilities as he understands them, or in his expectations of Yvonne's conduct, are less likely to be significant enough to cause him to change from an efficient option to a distinct, inefficient one.

While it is possible that the discreteness of available precautions will allow an efficient equilibrium to converge in some cases, there are several reasons that the flaws of the continuous model may still apply in practice to many cases. The world ordinarily offers many options, rather than just two or three. Drivers setting air-conditioning levels in their cars might have only a few levels to choose from (off, low, high, and so on), but in choosing the speed of their cars they face possibilities that are continuous rather than discrete. Pedestrians have enough options in choosing their speed, location, and how often they look at traffic for us to imagine, plausibly, that they face essentially a continuous range of choices about the amount of precaution they take. Railroaders usually won't have only a simple option of spark arresters, but an array of choices in both the *kind* of precautions they choose (spark arresters versus alternative track design versus alternative track location) and the *level* of the precautions they choose (perhaps facing a menu of eighty different spark arresters they might purchase from a variety of suppliers). The same is true of many other decisions, like what kind of

⁸⁵ Cf. COOTER & ULEN, *supra* note 1, at 247 ("Notice that buckling a seat belt is a discontinuous choice (yes-no). For discontinuous precaution, the relative efficiency of different rules depends upon particular facts.").

⁸⁶ See SHAVELL, *supra* note 37, at 179.

⁸⁷ *Id.*

seatbelts an automobile manufacturer should install, what sort of fence to use to surround a swimming pool, and so forth.

In any event, as the legal economists recognize, the success or failure of the formal economic model at stake here depends on the adequacy of the continuous model, rather than a discrete analogue of it, because it is the continuous model that expresses in a general form the conclusions that result from formal proof. As Cooter and Ulen write, “In general, discontinuous variables and cost functions yield messy results about optima, whereas continuous variables and cost functions yield clean results. It is usually best to build theory from clean results and then handle any messy results as exceptions.”⁸⁸ If the bilateral-liability-threat model depends on “messy results” in specific cases, that would, if nothing else, sharply limit its force as a general explanation of tort law’s negligence standard.

C. *The Limits of Activity Levels*

The other pillar of the modern economic analysis of tort law is the view that residual liability—that is, decisions about whether the injurer or the victim should bear the costs of accidents when both have behaved innocently (nonnegligently)—should be determined based on an analysis of activity levels. Specifically, the leading economists’ argument is that residual liability should depend either on “[w]hether injurers’ levels of activity are more important to control than victims”⁸⁹ or on which party’s “activity level most affects accidents.”⁹⁰

The problem with this view is that, while an understanding of activity levels sheds significant insight into the formal economic analysis of tort law, it ordinarily cannot serve as a basis for decisions in real cases. There are several reasons for this; they involve problems related to administrability and to a limitation in the analysis’s scope that makes it difficult for present economic models to promote efficient incentives, even for rational parties.

⁸⁸ COOTER & ULEN, *supra* note 1.

⁸⁹ SHAVELL, *supra* note 37, at 202.

⁹⁰ COOTER & ULEN, *supra* note 1, at 349.

1. Problems of Administrability

At the outset, it is important to consider whether economists have even attempted to offer a clear operational role for activity levels in determining tort rules. Consider, for example, Shavell's most recent explanation of how activity levels should matter in tort law:

Whether injurers' levels of activity are more important to control than victims' will depend on the context. As discussed before, when an activity of injurers (walking dogs of a vicious breed) creates substantial risks despite their exercise of due care, the activity will be desirable to control. This point is not fundamentally altered if account is taken of the activities of victims that expose them to risk. Especially if the victims' activities are just the activities of ordinary life (walking about, going to work), we would not want the activities constrained in favor of injurers' more dangerous activities. Conversely, when an activity of injurers (playing baseball) is not very dangerous if appropriate care is taken, the importance of controlling the activity will not be great; instead, we may see some advantage in reducing certain activities of victims that subject them to particular risks (such as pushing a baby in a stroller across a baseball field while a game is in progress).⁹¹

Though motivated primarily by efficiency, this formulation seems essentially noneconomic in nature, as if economic analysis has led us to a point where other considerations ought to reign, or where the costs and benefits are too complicated for present-day economics to study. If this is the case, there is little more to say except to note my agreement, for my goal here is to show that formal economics is insufficient on its own to determine optimal legal rules.

To say this differently, Shavell's explanation, read broadly, is a sensitive balancing test—so sensitive that it appears to allow essentially noneconomic considerations, or at least considerations very hard to quantify through narrow economic analysis, to influence tort rules. It is not clear, ultimately, that Shavell disagrees with me that formal economic analysis on this point is not especially helpful.

The formulation in Cooter and Ulen's textbook, though more specific, appears to be incomplete, at least on a narrow reading. They describe the role of activity levels as follows:

Usually one party's activity level affects accidents more than the other party's activity level. Efficiency requires choosing a liability

⁹¹ SHAVELL, *supra* note 37, at 202-03.

rule so that the party whose activity level most affects accidents bears the residual costs of accidental harm.⁹²

But, of course, looking only at which party's activity level more directly causes *accidents* violates the general economic observation that the law needs to be sensitive not just to costs of activities but to their benefits. Extra driving might indeed affect the level of accidents more than extra pedestrian activity on sidewalks, but it is at least possible that the benefit drivers get from extra driving is worth this extra cost. On economic terms alone, Cooter and Ulen's conclusion should likely be read more broadly and interpreted to assign liability in ways that reduce the total costs of precaution (including limited activity) plus the total expected costs of accidents. In other words, it is not the party whose activity most increases the likelihood of accidents that needs to bear residual liability, but the party who can restrict activity in ways that reduce social costs optimally.

Even on this broader reading of Cooter and Ulen's (and perhaps Shavell's) conclusion, however, activity-levels arguments would face serious problems as soon as courts or other parties tried to apply them. As I noted when introducing activity levels,⁹³ there is in principle little difference between choices about levels of care and choices about levels of activity, except that courts judge the former but tend not to judge the latter. But there are several reasons, in fact, that it makes sense for courts to avoid making judgments related to activity levels, and these reasons apply regardless of whether it is courts or other policymakers (or commentators) who set or defend particular legal rules.

To begin with, as Shavell's formulation of the role of activity levels seems to recognize,⁹⁴ there are many cases in which *more activity* does not lead to *more accidents*, as long as the activity is conducted safely. Does more walking on the sidewalk necessarily lead to more accidents between cars and pedestrians, assuming the pedestrians are all safe? It seems unlikely: safe pedestrians keep to the sidewalk (and crosswalks), look for oncoming traffic, and in general don't get hit by cars unless those cars veer off the road. Now, it is

⁹² COOTER & ULEN, *supra* note 1, at 349.

⁹³ See *supra* Part II.A.3.

⁹⁴ SHAVELL, *supra* note 37, at 203 (referring to activity that "is not very dangerous if appropriate care is taken").

possible that increased pedestrian activity increases the likelihood that people will get hit by cars, even if the pedestrians are careful, because there might be some cases in which cars veer onto sidewalks but avoid accidents only because those sidewalks were empty. In other words, there might be cases in which empty sidewalks result in harmless veering by cars, whereas full sidewalks result in serious accidents. But in practice, this situation is vague and unlikely to matter: there is enough physical space in the world, and on most sidewalks, that it seems implausible that there is even a measurable increase in the likelihood of an accident between a car and a pedestrian just because there are more pedestrians on the sidewalk.

Many cases are like this on both sides. Despite frequent assumptions that faultless accidents are commonplace, it is not, in fact, even clear that many car accidents result from mere activity, when the activity is safe. Ordinarily in a car accident there is some culprit, perhaps unidentified, who was at least careless: machinery fails because of a defect, a driver was driving unsafely given road conditions, a tire was under- or over-inflated, or something else was done incorrectly. Perhaps some car accidents are truly unavoidable even when everyone involved has behaved reasonably safely, but it is not clear that there are many such accidents or that the possibility of such accidents should dictate tort policy.

A separate problem is that the force of activity-levels arguments depends in part on the proposition that parties expect to be able to avoid liability when they act safely. But in many cases, this proposition assumes too much: many people who engage in potentially dangerous activities do not *know* that they are able to maintain a high level of safety, and in fact many know otherwise. For example, drivers cannot ensure that they are not going to be careless.⁹⁵ If nothing else, people's general knowledge that they cannot avoid carelessness mitigates the force of activity-levels arguments.

But perhaps the most significant problem applying activity levels to real tort cases is that reasoning in view of activity levels requires classificatory judgments that are nearly impossible to make in a principled fashion. For one thing, there

⁹⁵ See Mark Grady, *Res Ipsa Loquitur and Compliance Error*, 142 U. PA. L. REV. 887, 900 (1994) ("It is impossible to drive a car for any period of time without missing a required precaution."); see also MARC FRANKLIN ET AL., TORT LAW & ALTERNATIVES 46 (8th ed. 2006) (discussing Grady's article).

are many situations in which injurers and victims are similarly situated, or in fact engaged in the same activity (or an activity that might as well be the same). For instance, what would an understanding of activity levels suggest for accidents between two automobiles? Both drivers chose to drive. What about for airplane crashes, as between the airline and its passengers? Both the airline and the airplane passengers chose to take the particular flight that crashed.⁹⁶

Even when parties are situated differently in noticeable ways, there is generally not a principled economic method to determine whose activity level makes more of a difference (including both its costs and its benefits) and thus is more worth regulating. What of accidents between cars and pedestrians, where both were behaving safely? If such cases indeed occur frequently enough to worry about, whose activity makes more of a difference, accounting for both costs and benefits, and responds better to incentives? On what basis could a court decide?

Even Shavell's examples of supposedly clearer cases raise many of these problems. When considering "walking dogs of a vicious breed," for example, it is important to ask whether this activity really is dangerous "despite [the] exercise of due care."⁹⁷ And, though pushing a baby in a stroller through the middle of a baseball game is obviously an activity that should be minimized (and indeed not one that would appear to reflect the exercise of due care in a way that even triggers activity-levels arguments, because the activity can be judged unsafe on its own), what should we do about baseball stadiums and the people who live (or build houses) behind them, such that baseballs might break their windows? The levels of activity of both injurers and victims in cases like this appear to be symmetric, and recognizing activity levels does not break the symmetry.

Dynamite cases, and other activities where "a commercial actor has come to the type of location where [some] sort of dangerous thing is not normally done,"⁹⁸ reflect perhaps the strongest case for activity-levels arguments, in view of the

⁹⁶ See, e.g., Robert Cooter & Ariel Porat, *Liability Externalities and Mandatory Choices: Should Doctors Pay Less?*, 1 J. TORT L. (2006), available at <http://www.bepress.com/jtl/vol1/iss1/art2> (observing the same thing, in a related context, about doctors and patients).

⁹⁷ SHAVELL, *supra* note 37, at 203.

⁹⁸ Stephen D. Sugarman, *Rethinking Tort Doctrine: Visions of a Restatement (Fourth) of Torts*, 50 UCLA L. REV. 585, 608 (2002).

administrative problems I have just discussed. In these cases, there is perhaps a clear social understanding that an accident can be attributed to the level of one party's, rather than another party's, activities. Still, even these cases involve something of a noneconomic value judgment. For one thing, in theory, victims can still choose where to live in a way that minimizes the costs to them of activities like dynamite blasting, and the proper comparison of costs and benefits, on their own, seems hard to derive without empirical data. What seems to matter in cases of abnormally dangerous activities is that one party did something unexpected or unusual, *changing* a baseline level of activity that was occurring in an area and thereby violating social expectations.⁹⁹ But an analysis that depends on considerations like that is at least somewhat broader than a formal economic one.

2. Problems of Efficiency: Allocative Versus Redistributive Negligence

Even if activity levels could serve as a principled and administrable basis for assigning residual liability, there are powerful reasons that we probably would not want it to—at least without a kind of economic analysis that has not yet been done. This is because the economic study of torts has limited its focus¹⁰⁰ to the optimization of allocative efficiency through incentives for injurers and victims to take precaution. In doing so, it has neglected broader effects of tort rules on social costs.

Consider the ordinary case of what I call *allocative negligence*—that is, negligence that directly creates an allocative inefficiency, in terms of the costs of precaution and the costs of accidents. For example, suppose Xavier has the opportunity to install \$80 spark arresters to prevent \$4000 accidents, and he chooses not to do so. If the law supported his choice, there can be a clear misallocation of resources: spark arresters ought to be installed, but Xavier has no incentive to install them.¹⁰¹

⁹⁹ *See id.*

¹⁰⁰ *E.g.*, RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 69-70 (1st ed. 1972) (giving the classic allocative-efficiency argument for negligence rules).

¹⁰¹ Of course, in a world without transaction costs, Xavier and Yvonne might bargain over the installation of spark arresters. *See generally* Coase, *supra* note 15 (demonstrating that with no transaction costs, assignments of rights in nuisance law do not affect allocative efficiency).

But in casting problems this way, focusing on the particular relationship between injurers and victims, economic tort analysts often ignore a potential problem. I call this the problem of *redistributional negligence*—that is, distorted incentives (ultimately allocative in nature) that arise from tort regimes that give some parties the opportunity to engage in activity that is largely or even primarily redistributive. Of course, such opportunities for opportunistic redistribution may also be unfair, but for the purposes of this Article I need only criticize them on economic grounds, and on those grounds the problem is as follows: the opportunity to externalize costs can lead to incentives to engage in activity that is productive but also redistributive.

To say this somewhat differently, selfish rational actors who can choose between a variety of productive activities will not choose the one that is socially most productive. Instead, they will choose the one that earns *them* the most. However, some activities are more profitable to actors because they externalize costs onto others, not because they are more productive overall. With sufficient capacity to redistribute wealth through externalities, sanctioned by law, activities can function as capitalistic black holes, drawing in resources and effort even if they would be more productively applied elsewhere.

For example, return to the example of Xavier and Yvonne. It is common to see this kind of two-party example in economic analyses of tort law. But consider the following variation of the situation, which both (1) makes particular costs and benefits clear and (2) looks beyond the two activities in question (railroading and cornfield growing).

Xavier has recently graduated from business school. He has little chance to obtain credit, but he has an inheritance that he can use to invest in the business of his choice. His skills and experience make two choices salient: he can set up either a railroad company or a hotel. Both choices would consume his entire inheritance and require his full-time attention. Given all the costs and benefits associated with the opportunity to set up a hotel, Xavier's calculation of the expected value¹⁰² from that

¹⁰² Ordinarily, particularly as a business-school graduate, Xavier would engage in a net-present-value calculation that would consider the lifetime of the opportunity, the discount rate of income streams over that lifetime, and similar considerations. I put aside those details to keep the discussion in the text simple; they have no bearing on questions relevant to my discussion.

opportunity is \$20,000.¹⁰³ To compute the expected value of the railroading opportunity, Xavier considers the following information he has learned: running the railroad would be worth \$40,000 to him, but the expected costs of fires from the railroad to Yvonne's adjacent cornfields are \$37,000. Accordingly, the railroading opportunity is worth only \$3000 to Xavier if tort law makes him liable for these fires; if not, the opportunity is worth \$40,000. Thus, given the figures in this example, the tort regime directly influences Xavier's decision between his two business opportunities. Under a strict-liability regime, Xavier will choose the hotel, because the railroad—though profitable—has a large part of its value set to pay for the harms it causes. Under a negligence regime, however, Xavier can ignore these harms, because his \$40,000 value exceeds the \$37,000 cost to the cornfield and is thus deemed “reasonable.”¹⁰⁴ In that case, we expect that he will choose the railroad over the hotel, even though it is less socially valuable, because it gives him an opportunity to redistribute more wealth to himself through the externalities he causes.¹⁰⁵

An analysis of activities limited to whether a railroader or a corn grower's activity “most affects accidents”¹⁰⁶ (or similar formulations) as between those two parties misses an important feature of the situation in our example: from the perspective of allocative efficiency, assuming we have a choice between railroading and no railroading, we *want* railroading. It creates more value than the fires it causes. But we may want other things *more*, and if capital and other resources are limited, we want a tort system that discourages relevant parties from making decisions based on how much of other people's wealth they can redistribute to themselves by externalizing costs onto others.¹⁰⁷

¹⁰³ We can take this as an opaque figure, although it would be possible to specify details about the costs and benefits of the hotel opportunity and to assess the liabilities that Xavier could face as the owner-operator of a hotel. For the purposes of the example, however, I assume (just for simplicity) that the hotel gives rise to no significant liability or cost externalization, whereas the railroad does. This assumption does not change the force of the argument in the text.

¹⁰⁴ Assuming, again, that he is strictly rational and selfish, does not fear negative publicity, and so on.

¹⁰⁵ Cooter and Porat call this situation a “liability externality,” and agree generally that legal rules “should discourage activities with negative liability externalities.” Cooter & Porat, *supra* note 96, at 1 (emphasis omitted).

¹⁰⁶ COOTER & ULEN, *supra* note 1, at 349.

¹⁰⁷ To say this differently, economists who promote negligence rules for reasons only of narrow microeconomic conceptions of allocative efficiency would subject tort victims to Nozick's “utility monsters,” who derive so much value from some

Of course, if every wealth-producing business opportunity can be pursued—for instance, if capital markets are perfectly efficient and other significant transaction costs are minimal—then we may not care whether Xavier becomes a hotel operator or a railroad operator. In a world with no transaction costs and unlimited resources, whichever opportunity he forgoes (with its attendant wealth-producing and externalizing effects) will be taken up by someone else anyway. But as long as resources are not infinite and there are significant inefficiencies in the ability to nimbly pursue new business opportunities (because of resource limitations, borrowing costs, capital-market inefficiencies, and so on), social efficiency requires such opportunities to be priced as correctly as possible. To say this differently, it is not enough to say that we want to limit activities to levels at which they are still wealth-producing; it may also be vital to overall allocative efficiency for activities to be appropriately priced.

Note that this recognition, alone, doesn't solve the central problem that faces tort law, because it does not specify precisely what it means for activities to be priced appropriately. As commentators have long recognized,¹⁰⁸ there is a symmetry between injurers and victims: strict liability for railroaders, though it prevents railroaders from engaging in redistributive, cost-externalizing activity, might mean that corn growers have too much of an incentive to engage in corn growing, compared to other activities. But for tort law's solutions to this problem to be efficient, they cannot conceptualistically restrict their scope to the costs of accidents and precautions; they need instead to face all social costs and benefits squarely.

To summarize, analysis *internal* to a particular activity, or to the interaction between one activity and another, may well be insufficient to decide even the efficiency (much less the broader social appropriateness) of tort regimes that govern that activity. It may well be necessary to look at the "entire tableau"¹⁰⁹ (to borrow Shavell's phrase) of social costs and

activities that others must suffer in order to satisfy their appetites. See ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 41 (1974). If utility monsters are rewarded, however, then an allocatively inefficient result more broadly obtains: there is too much incentive to become like them, and not enough to engage in other activity.

¹⁰⁸ See, e.g., POSNER, *supra* note 30, at 138-40 (making the point, among others, that strict liability gives injurers but not victims efficient incentives to research new precautions, and vice versa).

¹⁰⁹ SHAVELL, *supra* note 37, at 188 n.17.

benefits—not just for the individual activities in question but for those activities as compared with other activities—in order to decide who should pay for which costs. Each activity, and its costs and benefits, are only part of a broader economic landscape.

Of course, Shavell is ultimately right—if we interpret him broadly enough—that our focus in deciding between strict liability and negligence depends on which activity is “more important to control.”¹¹⁰ But there is no reason to suppose that the question can be decided by the kind of activity-levels arguments that economists have marshaled so far. Instead, deciding between strict liability and negligence seems to demand a significantly broader judgment about which activities should bear the costs of accidents and which should not. Abstract and formal methodologies, unsurprisingly, have little aid to offer those seeking to make that kind of judgment. In any event, it would be a mistake—even on grounds of efficiency alone—to focus only on incentives to take precautions in deciding questions of tort liability, because the prices of activities can have broader incentive effects in society.

D. *Precautions and Activity Levels*

The Hand Formula itself occasionally suffers from classificatory problems similar to those faced by activity-levels arguments.¹¹¹ These are best conceived as problems of time-framing, though I mean something different by this than many other discussions of time-framing in this context.¹¹²

To see these problems, consider again the example of Xavier and Yvonne, the railroader and the cornfield owner. In this example, which we borrowed from Cooter, Xavier was said to have three precautions available to him: installing spark arresters, running trains more slowly, and running fewer trains.

Running fewer trains sounds like an activity-levels concern, in that it would be hard to judge from a single instance whether Xavier had run inefficiently many trains.

¹¹⁰ *Id.* at 202.

¹¹¹ See *supra* text accompanying notes 95-100.

¹¹² Cf. Douglas Husak & Brian P. McLaughlin, *Time-Frames, Voluntary Acts, and Strict Liability*, 12 LAW & PHIL. 95, 96 (1993) (discussing “time-framing” problems raised by Mark Kelman and Larry Alexander that concern how individual acts are to be judged).

Consider, for instance, an accident that occurs on May 13, 2006, at 12:10 a.m. It makes little sense to ask whether Xavier was running too many trains at that particular point in time; the question is not “What should Xavier have done at time *T*, the moment the accident occurred?” because the decision to run more or fewer trains goes beyond that particular time. More precisely, there are multiple time intervals that include time *T*, and there is no clear way to choose among them. What does it mean to say that Xavier should have run fewer trains? Fewer trains on May 13, 2006? In May 2006? In all of 2006 up to that point? Or over a broader period? What if Xavier had planned to run fewer trains later in the year, after the fateful accident happened (unfortunately) to arise? Could he convince a court that this plan was genuine, or would he have an inefficient incentive under an economically informed negligence rule to reschedule his trains evenly throughout the year (even if this weren’t otherwise optimal) in order to be confident that he could demonstrate that he ran fewer trains? Because courts cannot easily answer these questions, running trains is classified as an activity-levels problem, not as one about which courts can judge care or precaution directly.¹¹³

Unfortunately for tests like the Hand Formula, however, the same kind of reasoning applies to questions that are not so readily seen as activity-levels problems. For instance, what does it mean that on May 13, 2006, Xavier didn’t use spark arresters on his trains? Perhaps he had decided that, given the expected wear on spark arresters and the expected cost of fires, it was optimal to use them some but not all of the time.¹¹⁴ Over what period are we to evaluate that question? In some sense, every question about precaution under the Hand Formula can be reframed as a potentially intractable activity-levels problem.¹¹⁵

¹¹³ Cf. *supra* Part II.A.3. Note that decisions about activity levels suffer from a similar problem: in addressing which party’s activity more directly affected accidents, or whatever else, questions of time framing may be decisive.

¹¹⁴ Cf. Grady, *supra* note 95 (distinguishing the adoption of standards of precaution with *compliance* to those standards, and observing that “[i]n most activities, courts require perfect compliance; in others they do not”).

¹¹⁵ Shavell admits that there are other, somewhat similar concerns about the Hand Formula. For instance, “there may be dimensions of injurers’ care (such as the frequency with which drivers look in their rearview mirrors) that courts would not take into account in the determination of negligence because of difficulties in assessing them.” SHAVELL, *supra* note 37, at 189.

III. CONCLUSION

Perhaps surprisingly, then, we reach the conclusion that the prominent economic models still struggle fundamentally with the most basic question in tort law: when should there be liability? Far from having easily succumbed to formal economic analysis, tort law has firmly resisted simple or reductive explanations.

That said, my goal here has not been to minimize the insights of economists or to suggest that they be ignored. Indeed, law-and-economics scholars have successfully explained why the negligence standard is a *minimal* standard of liability. That is, liability must ordinarily be awarded *at least* in those cases where it would be inefficient, from the perspective of allocative efficiency, *not* to do so. I have called these cases of negligence, where precautions are clearly cheaper than the expected costs of the accidents they are designed to prevent, cases of *allocative negligence*, and economic analysis has indeed helped illuminate questions related to these cases.

Of course, such cases have not been especially controversial. The fundamental questions concerning negligence rules in tort law are whether they are better or worse than greater standards, not lesser ones. Perhaps the most basic question in tort law is whether behavior ought to be governed by rules of negligence, rules of strict liability, or perhaps something roughly in between. Economic analysis has helped characterize this question, but on its own, it does not provide an answer.

There are several reasons formal economic analysis of negligence rules is insufficient, even for commentators who believe that concerns about efficiency are paramount. For one thing, the prevailing formal defense of the efficiency of negligence standards—what I have called the bilateral-liability-threat argument—rests on a fragile economic model that breaks down rapidly, rather than degrading gracefully, as soon as assumptions are recognized to be slightly imperfect. It is one thing for a model to approximate the real world rather than purport to describe it perfectly. But the bilateral-liability-threat model threatens to unravel, in the general case, upon slight modifications to its assumptions.

The other pillar of the modern economic analysis of tort law is an understanding of activity levels, but arguments based on this understanding have limited force for courts and other

policymakers. The chief reason is that rules based on activity levels ordinarily cannot be applied to real cases; there are problems of framing, classification, information, and relevance that undermine attempts to put the theoretical understanding into practice, even if actors were rational and selfish and other classic assumptions of law-and-economics commentators were true. Moreover, the scope of activity-levels arguments, and perhaps of much economic analysis of tort law generally, has been too narrow. It has limited itself to only a few kinds of costs and benefits. As a result, the analysis is in danger of ignoring the problem that I have called *redistributional negligence*, where activity, although allocatively desirable, is priced incorrectly because of socially inefficient tort rules that nonetheless comport with leading economic models.¹¹⁶

¹¹⁶ This incorrect pricing helps explain, at least broadly, several other kinds of efficiency-related problems that others have observed. *See, e.g.*, David Gilo & Ehud Guttel, *Negligence and Insufficient Activity: The Missing Paradigm in Torts*, 108 MICH. L. REV. 277 (2009) (arguing that negligence standards might not lead only to *too much* safe activity, as traditional activity-levels analysis suggests, but also to *too little* safe activity). It should not be surprising that such problems exist under the prevailing economic models.

The Costs of Abusing Probationary Sentences

OVERINCARCERATION AND THE EROSION OF DUE PROCESS

Andrew Horwitz[†]

I. INTRODUCTION

The American criminal justice system has an apparent addiction to the use of probation as a means for adjudicating vast numbers of cases, particularly misdemeanors. With little discussion of or agreement about the appropriateness or efficacy of this sentencing practice, probation has become by far the most common form of criminal sentencing. While the primary justification given for such heavy reliance on probation is that we simply do not have the resources to incarcerate these offenders, that justification cannot survive serious scrutiny. In the first instance, it relies on the premise that we would, if we could, incarcerate huge numbers of low-level offenders, a proposition that is highly unrealistic. Additionally, however, because probation violators constitute the fastest growing component of an exploding prison population, it may well be that our reliance on probation as a default sentence is not really reducing our incarceration ranks, but simply reorganizing them to incarcerate different offenders.¹ So if the

[†] Distinguished Service Professor of Law and Director of Clinical Programs, Roger Williams University School of Law. B.A. 1983, Haverford College; J.D. 1986, New York University School of Law. I would like to thank the Roger Williams University School of Law for its financial support for this project. I would also like to thank Wendy Andrade, Emily Drosback and Lynn Laweryson for their able research assistance and my wife and children for their love and support.

¹ There are tremendous but largely unexplored public policy ramifications when decisions about whom to incarcerate are made in such a backward and unintentional fashion. For instance, if any interaction with the criminal justice system can easily escalate to incarceration, the system will disproportionately incarcerate those who are most likely to have that interaction, most notably the urban poor and people of color. See, e.g., Jerome Miller, *Do We Really Need More Prisons?*, N.J. RECORD, May 14, 1989, at O1 (noting that in certain parts of the country, seven out of ten young black men can expect to be arrested at least once).

use of probation as a default sentence for those we do not incarcerate cannot be justified on those grounds, why have we continued down this path? And at what cost?

There are two clear and direct consequences of the overuse and abuse of probation as a criminal disposition. The first is that we are losing a tremendous opportunity to use probation for its historically intended purpose: rehabilitation. When the numbers of probationers becomes so large that supervision and the provision of services and support becomes impossible, the reformatory potential of probation is completely lost. The second and more disturbing consequence is the creation of a shadow criminal justice system in which an extraordinary percentage of criminal charges is resolved not through our normal adjudicative process, but rather through a probation violation process that runs roughshod over the constitutional rights of the accused. When a probationer is charged with a new criminal offense, that new offense typically generates a corresponding allegation that the probationer violated the terms of his or her probation. When the new charge is processed as an alleged probation violation, the probationer is entitled only to a limited hearing at which the rules of evidence are relaxed, the right to confrontation is limited, the burden of proof is far lower than the usual beyond a reasonable doubt standard, and the right to a trial by jury is non-existent. The outcome of this violation hearing will often obviate the relevance of any trial on the new charge. A probationer facing such a violation hearing, with its quite limited prospect for a successful outcome, will most often simply admit to the new charge, whether innocent or not.

This Article will explore each of these consequences in depth and provide some ideas for constructive ways to avoid them. In Part II, the Article will describe the evolution of probation from its roots as a condition imposed upon a select population of criminal defendants who seemed likely to benefit from assistance, support, and supervision, to the default sentence imposed upon a majority of defendants with little to no regard for whether probation makes sense for that defendant. As our prison population grows larger and larger, a relatively small amount of our corrections budget is designated for probation departments, resulting in minimal or, in many cases, no supervision or provision of services to an exploding probation population. We have overwhelmed probation departments so that they cannot possibly perform the function that we hope they might. Recognizing this, we have abandoned

any prospect that probation might assist a defendant's rehabilitation and now use it simply as a noose around an offender's neck, waiting for the inevitable violation. Not surprisingly, recidivism and failure to adhere to the technical requirements of probation have created a burgeoning prison population. Having placed an offender on probation, a court often feels compelled to respond to a violation with incarceration in order to maintain credibility. And, in that fashion, probation imposed upon a defendant who never needed programmatic support and supervision, or who needed it but never got it, turns a case that never merited incarceration into an incarceration case. In the process, we have created a cycle out of which many defendants never emerge, preventing them from obtaining jobs and decent housing.

Part III of this Article will detail the probation violation process, explaining how the process has largely taken over the criminal justice system, eradicating some of the constitutional rights and protections that we hold most dear. We have created a second class of citizens—those on probation—for whom the Constitution no longer applies in any meaningful way. The rights and protections inherent in our legal system, developed and refined over the past two centuries, are relegated to the caboose of a train driven by the probation violation process. The end result is often that any interaction whatsoever with the criminal justice system can escalate into incarceration, which means that those most likely to have low-level interactions with the criminal justice system—the urban poor and people of color—will suffer disproportionately.

Part IV will suggest a return to an earlier time when probation was used with a specific purpose in mind, reserved for those defendants who can truly benefit from support and supervision. In order to allow probation to engage in meaningful support and supervision, the number of probationers each probation officer is expected to supervise must be drastically reduced. One means of accomplishing this, of course, would be a significant expansion in funding allowing for the hiring of many more probation officers. In today's economic and political climate, however, the likelihood of greater funding for probation on the scale that would be required is simply unrealistic. Although these concepts are not mutually exclusive, a more feasible and realistic approach would entail a significant reduction in the number of probationers. This reduction can be accomplished by ending the concept of unsupervised probation and by significantly

expanding the use of other alternative sentencing options. If we move beyond the “probation-as-default” approach of the last few decades and return to using probation only when actual support and supervision are merited, we can accomplish several crime control objectives and avoid unnecessary fiscal and human costs.

Finally, Part V of this Article will propose changes to the probation violation process to enhance the fairness and reliability of the criminal justice system. To maintain the integrity of the criminal justice system, we must stop using probation as a means of engaging in an end-run around the system’s mechanisms for protecting the rights of defendants. Except under extraordinary circumstances, the hearing concerning an alleged probation violation predicated on a new criminal charge should not be held before the resolution of the new charge. If the probationer is acquitted of the new charge or the new charge is dismissed, the violation allegation should likewise be dismissed. If the probationer is convicted of or admits to a new charge, he or she can be sentenced on the probation violation accordingly. By sequencing the events in this fashion, we can avoid the use of the probation violation hearing as a substitute for a trial. In so doing, we can restore some of the public’s eroded faith in the fairness and integrity of the system.

II. FROM POSITIVE REHABILITATIVE TOOL TO LOST OPPORTUNITY

A. *The Origins of Probation*

The origins of probation in the United States can be traced to a Boston cobbler named John Augustus, referred to as the “father of probation.” In the 1840s, Augustus intervened in the Massachusetts court system on behalf of thousands of “common drunkards” and “petty criminals.”² The prevailing penal philosophy of the eighteenth century was quite severe, suggesting that the only response to criminal behavior was harsh corporal punishment.³ Reforms during the early

² See 1 NEIL P. COHEN, *THE LAW OF PROBATION AND PAROLE* § 1:3 (2d ed. 1999); PAUL F. CROMWELL, JR. ET AL., *PROBATION AND PAROLE IN THE CRIMINAL JUSTICE SYSTEM* 10 (2d ed. 1985); Wayne A. Logan, *The Importance of Purpose in Probation Decision Making*, 7 *BUFF. CRIM. L. REV.* 171, 174-75 (2003).

³ COHEN, *supra* note 2, § 1:2; CROMWELL, *supra* note 2, at 5.

nineteenth century focused on the replacement of corporal punishment with incarceration.⁴ Augustus's intervention was part of a larger reform movement that questioned the retributive orientation of the criminal justice system and sought a greater focus on the rehabilitation of the offender. Augustus's view was that the purpose of the criminal law should be "to reform criminals and to prevent crime, and not to punish maliciously or from a spirit of revenge."⁵

Early on, Augustus's efforts were roundly criticized as being soft on crime and encouraging criminal behavior.⁶ But due in part to the widespread recognition that prisons were not serving any rehabilitative purpose and in part to Augustus's early successes, the concept of probation became more popular and more widely accepted. In 1878, Massachusetts passed the first probation statute, followed quickly by a number of other states.⁷ By 1925, all forty-eight states and the federal government formally adopted probation by statute.⁸

Although Augustus supervised over two thousand probationers in his eighteen years in the field,⁹ he chose them carefully, recognizing that probation would not be an appropriate disposition for every offender. As he described the process, "Great care was observed, of course, to ascertain whether the prisoners were promising subjects for probation, and to this end it was necessary to take into consideration the previous character of the person, his age, and the influences by which he would in future be likely to be surrounded."¹⁰ In the early part of the twentieth century, the prevailing notions of probation incorporated this selective ideal. A summary of the professional literature in 1960 described probation as "the application of modern, scientific case work to specially selected offenders who are placed by the court under the personal supervision of a probation officer . . . and given treatment

⁴ COHEN, *supra* note 2, § 1:2; CROMWELL, *supra* note 2, at 5.

⁵ CROMWELL, *supra* note 2, at 11 (quoting JOHN AUGUSTUS, A REPORT OF THE LABORS OF JOHN AUGUSTUS 23 (1852)).

⁶ *Id.*; Logan, *supra* note 2, at 176; Joan Petersilia, *Probation in the United States*, 22 CRIME & JUST. 149, 155-56 (1997) [hereinafter Petersilia, *Probation*].

⁷ Logan, *supra* note 2, at 175.

⁸ CROMWELL, *supra* note 2, at 12.

⁹ Logan, *supra* note 2, at 175.

¹⁰ CROMWELL, *supra* note 2, at 10 (quoting JOHN AUGUSTUS, A REPORT OF THE LABORS OF JOHN AUGUSTUS 34 (1852)).

aimed at their complete and permanent social rehabilitation.”¹¹ The expansion of probation coincided with a significant shift in the prevailing philosophy of the criminal justice system away from retribution and in the direction of reform and rehabilitation. In 1949, the United States Supreme Court recognized the magnitude of the attitudinal shift: “Retribution is no longer the dominant objective of the criminal law. Reformation and rehabilitation of offenders have become important goals of criminal jurisprudence.”¹²

The second half of the twentieth century brought with it a number of developments. Perhaps the most notable was the abandonment of the notion that probation was a disposition that should be reserved for specially selected offenders. The newly minted Model Penal Code suggested a “probation-as-default” approach to criminal sentencing, suggesting that all cases should be resolved with probation unless incarceration was absolutely necessary for public protection.¹³ Similarly, the American Bar Association’s Standards for Criminal Justice suggested that “the automatic response in a sentencing situation ought to be probation, unless particular aggravating factors emerge in the case at hand.”¹⁴ What followed was a substantial expansion of the probation population¹⁵ as probation quickly became “the most common form of criminal sentencing in the United States.”¹⁶ Between the 1950s and the 1970s, probation “evolved in relative obscurity” until published reports in the 1970s exposed the massive underfunding of probation departments and criticized the utility of probation as a criminal disposition.¹⁷

B. *Probation in the Modern Era*

In what many view as a watershed event, sociologist Robert Martinson in 1974 published a meta-analysis of over two hundred evaluations of rehabilitative programs, famously

¹¹ Logan, *supra* note 2, at 180 n.42 (alteration in original) (quoting Lewis Diana, *What is Probation?*, 51 J. CRIM. L., CRIMINOLOGY & POL. SCI. 189, 197 (1960)).

¹² Williams v. New York, 337 U.S. 241, 248 (1949).

¹³ See Logan, *supra* note 2, at 181-87 (detailing the creation of the Model Penal Code provisions relating to probation).

¹⁴ AMERICAN BAR ASS’N, A.B.A. STANDARDS FOR CRIMINAL JUSTICE, STANDARDS RELATING TO PROBATION, Introduction (1970).

¹⁵ Logan, *supra* note 2, at 187.

¹⁶ Petersilia, *Probation*, *supra* note 6, at 149.

¹⁷ *Id.* at 157.

concluding that “nothing works.”¹⁸ While the scholarly community expressed serious concerns about the methodology employed¹⁹ and even Martinson himself tried later to qualify his conclusions,²⁰ Martinson’s “nothing works” conclusion “quickly caught on with the public and politicians”²¹ and became “the rallying cry of a new generation of criminologists.”²² By the end of the 1980s, the abandonment of the rehabilitative ideal in favor of a retributive model of criminal justice was all but complete.²³

With the end of the rehabilitative ideal came an extraordinary and unprecedented movement toward incarceration. Criminologist Michael Tonry describes in stark terms “the modern American preoccupation with absolute severity of punishment and the related widespread view that only imprisonment counts.”²⁴ As a consequence, the United States incarcerates a higher percentage of its citizens for a greater average duration than any other western nation.²⁵ The prison and jail population in the United States increased nearly seven-fold from 1970 to the early twenty-first century,²⁶ with much of that growth coming in the 1990s and beyond. As of 2008, over 2.2 million Americans, one in every 131 people, were incarcerated.²⁷ More than one in ten black males aged 25-29 was in prison or jail.²⁸

One might think that the abandonment of the rehabilitative ideal and the increased reliance on incarceration would have foreshadowed the end of probation as a primary sentencing mode, but such was not to be the case. As the incarceration rates have grown, so too have the rates of defendants being placed on probation. The probation

¹⁸ Robert Martinson, *What Works? Questions and Answers about Prison Reform*, 35 PUB. INT. 22, 48-49 (1974).

¹⁹ See Robert A. Shearer & Patricia Ann King, *Multicultural Competencies in Probation—Issues and Challenges*, FED. PROBATION, June 2004, at 3.

²⁰ See Robert Martinson, *New Findings, New Views: A Note of Caution Regarding Sentencing Reform*, 7 HOFSTRA L. REV. 243, 244 (1979).

²¹ Logan, *supra* note 2, at 190.

²² Shearer & King, *supra* note 19, at 3.

²³ See William D. Burrell, *Trends in Probation and Parole in the States*, in COUNCIL OF STATE GOVERNMENTS, BOOK OF THE STATES 2005 595, 597 (2005).

²⁴ MICHAEL TONRY, SENTENCING MATTERS 128 (1996).

²⁵ *Id.*; THE SENTENCING PROJECT, FACTS ABOUT PRISONS AND PRISONERS (2009), available at http://www.sentencingproject.org/doc/publications/publications/inc_factsaboutprisons_Dec2009.pdf.

²⁶ Logan, *supra* note 2, at 190; THE SENTENCING PROJECT, *supra* note 25.

²⁷ THE SENTENCING PROJECT, *supra* note 25.

²⁸ *Id.*

population in the United States almost tripled between 1980 and 1997, from just over one million to more than three million,²⁹ and that growth has continued unabated. By 2002, the number had climbed to over four million, a 30% increase between 1995 and 2002, and has since continued upward, reaching nearly 4.3 million in 2007.³⁰ Probation cases accounted for over half of the growth in the entire correctional population between 1995 and 2006,³¹ and made up three quarters of the growth in the number of offenders under community supervision in 2007.³² Projections predict continued growth.³³ Within the adult population in the United States, 1.78% are presently on probation.³⁴ The massive expansion of probation appears to be explained in many jurisdictions largely by prison overcrowding and insufficient funds to support further incarceration.³⁵

This extraordinary expansion of the probation system has not been accompanied by any correlating expansion in funding. As the number of probationers continues to rise in staggering proportions, spending on probation has been “stagnant or decreasing.”³⁶ From 1977 to 1990 the number of probationers essentially tripled in size but spending as a percentage of governmental budgets did not change.³⁷ During the same period, spending for prisons and jails doubled.³⁸ “Despite the fact that they handle the vast majority of the offender population, probation and parole receive less than ten

²⁹ COHEN, *supra* note 2, § 1:1 n.2.

³⁰ BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, NCJ 228230, PROBATION AND PAROLE IN THE UNITED STATES, 2008, at 1 (2009) [hereinafter ANNUAL PROBATION SURVEY, 2008], available at <http://bjs.ojp.usdoj.gov/content/pub/pdf/ppus08.pdf>.

³¹ BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, NCJ 220218, PROBATION AND PAROLE IN THE UNITED STATES, 2006, at 1-2 (2007).

³² ANNUAL PROBATION SURVEY, 2008, *supra* note 30, at 3 tbl.1.

³³ Burrell, *supra* note 23, at 595.

³⁴ ANNUAL PROBATION SURVEY, 2008, *supra* note 30, at 1. It is interesting to note for sake of comparison that while 1.78% of the nation’s adult population is on probation today, in 1980 only 1.12% of that same population was under any correctional supervision, including jail, prison, probation and parole. *Id.*

³⁵ See COHEN, *supra* note 2, § 1:25.

³⁶ Joan Petersilia, *A Crime Control Rationale for Reinvesting in Community Corrections*, 75 PRISON J. 479, 484 (1995) [hereinafter Petersilia, *Crime Control*].

³⁷ *Id.* at 483-84.

³⁸ *Id.* at 483.

percent of the correctional funding from state and local governments.”³⁹

Not surprisingly, then, two things have happened over the past few decades: caseloads for probation officers have grown exponentially, and the level of actual support and supervision has declined nearly to the point of non-existence. In the era when probation was viewed as a legitimate rehabilitative enterprise, recommendations for probation officer adult caseloads ranged from the 1967 President’s Crime Commission recommendation of thirty probationers⁴⁰ to what the American Bar Association in 1970 called the “widely recognized standard” of fifty probationers.⁴¹ More recent reports estimate national caseloads averaging as high as 250 probationers per officer.⁴² In data published in 1999, Rhode Island had the highest reported average of any state in the country with an average of over 350 probationers per supervising probation officer.⁴³

As caseloads have skyrocketed, supervision has precipitously declined. For significant numbers of probationers, probation means complete freedom from supervision. On the national level, the percentage of probationers who are even required to report to a probation officer declined from 79% in 1995 to 70% in 2005.⁴⁴ Locally, things appear to be much worse in the urban areas where most offenders live. A 1995 study of the probation system in Los Angeles, reporting caseloads in the hundreds, concluded that at least 60% of all probationers received no services or supervision of any kind.⁴⁵ A similar study in Texas revealed that 95% of the 400,000 adults on probation were required to report only once every three months.⁴⁶ A probation officer testifying in California in 1993, acknowledging that more than half of the probationers on his

³⁹ Burrell, *supra* note 23, at 596; see also Petersilia, *Crime Control*, *supra* note 36, at 484.

⁴⁰ See Petersilia, *Crime Control*, *supra* note 36, at 484.

⁴¹ AMERICAN BAR ASS’N, *supra* note 14, at Standard 6.1 cmt.

⁴² Petersilia, *Probation*, *supra* note 6, at 167.

⁴³ AM. PROB. & PAROLE ASS’N, app. 1 tbl.11 (on file with author) (citing C.G. CAMP & G.M. CAMP, *THE CORRECTIONS YEARBOOK 1999: ADULT CORRECTIONS* (1999)).

⁴⁴ BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, NCJ 215091, *PROBATION AND PAROLE IN THE UNITED STATES, 2005*, at 6 tbl.3 (2006) [hereinafter *ANNUAL PROBATION SURVEY, 2005*], available at <http://bjs.ojp.usdoj.gov/content/pub/pdf/ppus05.pdf>.

⁴⁵ Petersilia, *Crime Control*, *supra* note 36, at 484; Petersilia, *Probation*, *supra* note 6, at 169.

⁴⁶ Petersilia, *Crime Control*, *supra* note 36, at 484.

caseload were completely unsupervised, summarized the situation quite starkly:

On each judicial day hundreds of California judges sentence thousands of offenders to probation, sternly enumerating the many conditions of probation that are enforced by the probation officer. Unfortunately, virtually all of these offenders will never see a probation officer and there will be absolutely no enforcement of the court ordered conditions. Equally unfortunate is that all of the players in this drama—especially the offender—understand that the offenders will go unsupervised.⁴⁷

As a consequence of underfunding and growing caseloads, “probation supervision in many large jurisdictions amounts to simply monitoring for rearrest.”⁴⁸

C. *A Shift in the Underlying Philosophy of Probation*

These trends in probation—exploding caseloads, little to no supervision—have been accompanied by a corresponding change in the prevailing philosophy undergirding and governing probation supervision. Because the history of probation is firmly rooted in the rehabilitative ideal, abandoning that ideal while at the same time increasing reliance on the use of probation required an adjustment in thinking. A “Justice Model” of probation, in which the primary focus of probation is retribution, emerged in the 1980s and remains dominant to this day.⁴⁹ The decision to place an offender on probation has become much more likely to be motivated by a desire to exact retribution for criminal conduct while, at the same time, avoiding the state expense of incarceration.⁵⁰ With the widespread adoption of the “nothing works” mantra, there is no real expectation of rehabilitation. Probation has shifted from being viewed as an alternative to punishment, supplemented with services and support, to being considered a punishment in and of itself, supplemented with obligations and restrictions on freedom.⁵¹ Even the American

⁴⁷ *Id.* at 486 (quoting testimony of Robert Kelgord before the Commission on the Future of the California Courts, 1993).

⁴⁸ *Id.*; see also Robin Campbell & Robert V. Wolf, *Problem-Solving Probation: An Overview of Four Community Based Experiments*, TEX. J. CORRECTIONS, Aug. 2001, at 8, 9 (noting that “at best, a handful of probationers may get the necessary referrals and support to guide them on a path of reform while the vast majority live in the community with virtually no supervision”).

⁴⁹ COHEN, *supra* note 2, § 1:5; CROMWELL, *supra* note 2, at 111.

⁵⁰ COHEN, *supra* note 2, §§ 1:9, 1:25.

⁵¹ *Id.* at § 1:6; Logan, *supra* note 2, at 196.

Bar Association in its 1994 Standards for Criminal Justice abandoned the term “probation” in favor of the term “compliance programs.”⁵² The justice model “repudiates the notion that probation is a sanction designed to rehabilitate offenders in the community, and presents the concept that a sentence of probation represents a proportionate punishment lawfully administered for certain prescribed crimes.”⁵³ Along those lines, the justice model “holds that current practices of counseling, surveillance, and reporting accomplish very little and have minimal impact on recidivism. On the other hand, probation that consists of monitoring court orders for victim restitution or community service and ensures that the imposed deprivation of liberty is carried out, represents a clear and achievable task.”⁵⁴

The adoption of the justice model brought with it a major change in both the staffing and the philosophy of probation departments. Traditionally, probation officers most commonly came from social work backgrounds.⁵⁵ They often referred to themselves as “probation counselors” and to the probationers as “clients.” Under the justice model, the probation officer is much more likely to come from a law enforcement background, to call himself or herself a “probation officer,” and to refer to probationers as “offenders.”⁵⁶ These changes in staffing and in language are reflective of the move away from the rehabilitative model and firmly in the direction of a retributive model.

Viewing probation through a law enforcement perspective rather than a social work perspective has consequences, of course. If probation is about complying with conditions as a form of punishment, then noncompliance must be penalized if the system is to maintain any credibility.⁵⁷ And that penalty is frequently incarceration.⁵⁸ Two broad categories of offenders now flood the prison system: probationers who have failed to comply with some condition of probation—called “technical violators”—and probationers who have been

⁵² AMERICAN BAR ASS’N, A.B.A. STANDARDS FOR CRIMINAL JUSTICE, STANDARDS RELATING TO SENTENCING § 18-3.13 cmt. (3d ed. 1994).

⁵³ CROMWELL, *supra* note 2, at 111.

⁵⁴ *Id.* at 111-12.

⁵⁵ *Id.* at 105-07.

⁵⁶ *See id.* at 105-12.

⁵⁷ *See* TONRY, *supra* note 24, at 101-02.

⁵⁸ *See* Petersilia, *Probation*, *supra* note 6, at 193.

rearrested on a new criminal allegation.⁵⁹ With an almost complete absence of programmatic support or supervision, the fact that each of these categories is substantial ought not be terribly surprising. “Stated simply, offenders who fail while under community supervision constitute the fastest growing component of the prison and jail populations in this country.”⁶⁰ One study reports that probation violators represented 17% of prison admissions nationally in 1980 but by 1999 had doubled to 35%.⁶¹ Another study placed the figure at between 30% and 50% of new admissions.⁶² Some state figures are substantially greater, reaching as high as 80% of new admissions.⁶³ Because of the intractable nature of many of the causes of violations, “these revocation processes result in ‘churning,’ in which individuals repeatedly circulate in and out of custody It has become increasingly clear to correctional administrators and policymakers alike that this is a costly and counterproductive approach.”⁶⁴ It has become equally clear that the “high failure rates of probationers and parolees . . . contribute significantly to prison crowding.”⁶⁵ Something clearly must be done to reverse this path.

III. THE PERVERSION OF THE CRIMINAL JUSTICE PROCESS

A. *The Probation Violation Cycle*

The fact that our prisons are being flooded with probation violators begs the question of how all of those probation violators were sentenced to jail time. The reality is that we have designed a shadow criminal justice system in which probationers can be sent to prison on little evidence and with little procedural protection. Record numbers of offenders are placed on probation each and every year, with probation

⁵⁹ See *id.* at 166; Petersilia, *Crime Control*, *supra* note 36, at 488.

⁶⁰ Faye S. Taxman & James M. Byrne, *Locating Absconders: Results from a Randomized Field Experiment*, FED. PROBATION, Mar. 1994, at 13, 13, see also Petersilia, *Crime Control*, *supra* note 36, at 488.

⁶¹ RYAN S. KING, CHANGING DIRECTION?: STATE SENTENCING REFORMS 2004-2006, at 11 (The Sentencing Project 2007), available at <http://www.sentencingproject.org/doc/publications/sentencingreformforweb.pdf>.

⁶² Petersilia, *Crime Control*, *supra* note 36, at 488.

⁶³ See Petersilia, *Probation*, *supra* note 6, at 166 (noting that Texas reported that 66% of all prison admissions in 1993 were probation or parole violators, while California reported a rate of over 60% and Oregon a rate of over 80%).

⁶⁴ KING, *supra* note 61, at 11.

⁶⁵ Petersilia, *Crime Control*, *supra* note 36, at 488.

serving as the default sentence for any offender who cannot or will not be incarcerated as an immediate consequence of the court's adjudication of the case. In some jurisdictions, almost every misdemeanor is resolved by placing the offender on probation. If the offender sees a probation officer at all—and very many will not—the visit alone will be an end, not a means, of establishing compliance with the terms of probation. Failing to keep that appointment will result in the filing of a technical violation of probation. If there are special conditions attached to the term of probation, the offender will generally be expected to provide some evidence of compliance with those conditions. Because the probation officer has an unmanageable number of probationers to supervise, it is unlikely that any support services beyond referrals to underfunded or unavailable service providers will be offered or received. In the absence of available services, evidence of some effort to obtain services, even if wholly unsuccessful, will often be deemed as compliance. Absent an arrest on a new charge, the probationer will be deemed to have successfully completed the probationary term if he or she can comply with these minimal obligations.

What of the probationer who cannot or does not comply with these obligations? The technical violator—the probationer who fails to appear for a scheduled appointment, fails a drug test, or fails to fulfill a special condition—will in all likelihood be brought before the sentencing court as a probation violator. Although the original criminal charge did not merit a jail sentence and the probationer has not been charged with engaging in new criminal activity, it is more likely than not that the probationer will now be incarcerated, at great expense to the government, and often for an extraordinarily long period of time.⁶⁶ If the goal of the probationary sentence was to deter future criminal behavior, it is hard to justify incarceration in the absence of criminal behavior. The consequence of a probationer's failure to meet what are often unrealistic expectations can frequently be a prison sentence far in excess of what anyone would ever have thought justified by the original criminal charge.⁶⁷ In what might be viewed as a classic

⁶⁶ See TONRY, *supra* note 24, at 105.

⁶⁷ Criminologist Michael Tonry, in his book entitled *Sentencing Matters*, explores the argument that the high rates of technical violations of probation simply “expose the unreality and injustice of conditions—like prohibitions of drinking or expectations that offenders will conform to middle-class behavioral standards they have never observed before—that many offenders will foreseeably breach and that do not involve criminality. Many offenders have difficulty in achieving conventional, law-

example of this scenario, a defendant in Arkansas who had been convicted of theft was eventually sentenced to five years in prison solely for failing to report to his probation officer as required.⁶⁸ The defendant, who had been given permission to leave the state to look for work, explained that he had “moved a lot . . . looking for work, and that he could not always get the report, a stamp, and an envelope together.”⁶⁹ The Court of Appeals of Arkansas upheld the five year sentence.⁷⁰ In just this fashion, we often dedicate scarce prison resources to a failed probationer who committed a minor or non-violent crime rather than to an offender who committed a far more serious offense.

But even more disturbing is the treatment of the probationer who is charged with a new crime. In many jurisdictions this probationer will be incarcerated as a matter of practice or as a matter of law while he or she awaits a probation violation hearing, whether or not the new charge merits incarceration.⁷¹ In all likelihood this probationer will end up incarcerated as a probation violator as a result of the new criminal allegation, and this remains the case even if the new charge is ultimately dismissed or, worse, even if he or she is ultimately acquitted on that charge after a trial.⁷² Most frequently the probation violation allegation will be used as a vehicle to force a resolution of the new criminal charge, leaving that charge completely untested by the normal adjudicative process.

abiding patterns of living and many stumble along the way.” *Id.* He points out that a “traditional social work approach to community corrections would expect and accept the stumbles (so long as they do not involve significant new crimes) and hope that through them, with help, the offender will learn to be law-abiding.” *Id.*

⁶⁸ *Luyet v. State*, CA CR 81-69, 1981 WL 930, at *1 (Ark. Ct. App. Nov. 12, 1981).

⁶⁹ *Id.* at *1.

⁷⁰ *Id.* at *2. Similarly, in *Morgan v. State*, 588 S.W.2d 431 (Ark. 1979), the court upheld a three year prison sentence for a defendant who had pled guilty to forgery and who, while on probation, moved out of state without permission to obtain employment.

⁷¹ Most efforts by probationers to be released while they await a hearing are unsuccessful. Because there is no constitutional presumption of innocence at a probation revocation proceeding, absent a statute allowing judges the discretion to grant bail, probationers will generally be held until their revocation hearing. COHEN, *supra* note 2, §§ 18:5-18:7.

⁷² Most jurisdictions justify this outcome by noting that the standard of proof during a probation violation is by a preponderance of the evidence, whereas during a criminal trial, the standard is beyond a reasonable doubt. The significant gap between these two standards of proof creates a very high likelihood that a probationer will be found guilty during a violation hearing even if they are acquitted during criminal proceedings. *Id.* § 22:15.

B. An End-Run Around the Constitution

The honest truth is that the probation violation mechanism has in many cases completely taken over the mechanical functioning of criminal justice system. With unprecedented numbers of offenders on probation at any time, the likelihood that a defendant charged with a crime is presently on probation is high.⁷³ In that scenario, the system lends itself to an end-run around all of the procedural protections in place to protect the innocent, and the simple exercise of constitutional rights is punished. The primary impact of probation on the criminal justice system is the generation of a shadow criminal justice system in which procedural protections such as the presumption of innocence and the right to a jury trial are disregarded and decisions about incarceration are made essentially by default.

When a probationer is arrested on a new criminal charge, that person is brought before the court to be arraigned on the new charge. It is generally at that very same arraignment that the probationer is generally presented with the allegation that he or she, by committing the new crime, has violated his or her probationary terms. Often, there is a heavy presumption or even a requirement that the probationer will be incarcerated until the probation violation allegation is adjudicated.⁷⁴ In the misdemeanor context, this presumption can frequently have the effect of coercing an immediate resolution of both the alleged probation violation and the new criminal charge.⁷⁵ If a defendant can avoid further detention and obtain release from custody only by admitting a violation and pleading guilty to a new criminal charge, he or she will almost invariably exercise that option regardless of guilt or innocence.⁷⁶

A recent story in the *Providence Journal* chronicled the ugly path that the system can follow when a defendant is placed on probation. A woman engaged in a bitter divorce was repeatedly arrested and charged with misdemeanor offenses

⁷³ In 2007, one in every forty-five adults in the United States was supervised in the community and over 80% of those being supervised were on probation. ANNUAL PROBATION SURVEY, 2008, *supra* note 30, at 1.

⁷⁴ See COHEN, *supra* note 2, §§ 18:5-18:7; ANDREW R. KLEIN, ALTERNATIVE SENTENCING, INTERMEDIATE SANCTIONS AND PROBATION 319 (2d ed. 1997).

⁷⁵ See KLEIN, *supra* note 74, at 329.

⁷⁶ See *id.*

based on allegations made by family members.⁷⁷ On nine separate occasions she was held without bail as an alleged probation violator based on those allegations alone, sometimes for more than a month, until she was ultimately acquitted or the charge was dismissed.⁷⁸ It is unclear whether she was ever offered the opportunity to enter an admission to any of those charges in order to avoid incarceration; if she had been offered that chance, she almost certainly would have taken it. It is the rare defendant indeed who will stay in custody in order to contest a charge when he or she can be released upon an admission of guilt.⁷⁹

With the looming threat of incarceration, the defendant's status as a probationer acts as an almost complete barrier to challenging the veracity of the new criminal allegation or exercising any of the connected constitutional rights because the cost of doing so is more than most defendants can or will bear. While expedient, this process actually serves no constituency very well. Because the veracity and accuracy of the charges is unsubstantiated, the innocent can and do get swept up with the guilty. Because the validity of arrests and charges goes untested, sloppy or unlawful police and prosecutorial work gets rewarded. The defendant, unable to challenge even unjust or untrue charges, accumulates a criminal history from which he or she is unlikely to recover. And because what follows from the new charge is almost invariably yet another term of probation, the defendant walks closer and closer to that line of incarceration. Eventually, and often sooner rather than later, a defendant who has never received any support or social services and whose problems remain untreated winds up incarcerated on charges that nobody truly believes merit incarceration. And the injustice of

⁷⁷ John Hill, *Override Urged in Probation-Violation Veto*, PROVIDENCE J., Dec. 17, 2009, at A13.

⁷⁸ *Id.*

⁷⁹ Another Rhode Island story makes this point in a rather stark fashion, albeit in the context of an alleged bail violation. Accused by an ex-boyfriend of violating a restraining order, a special needs teacher in her fifties was released on bail. Bob Kerr, *She Paid When the Law Came Apart*, PROVIDENCE J., Oct. 12, 2008, at B1. When the ex-boyfriend made another unsupported allegation, she faced the choice of admitting guilt to obtain her release or asserting her innocence enduring two weeks of incarceration to contest the charge. *Id.* On the day of her arraignment she initially asserted her innocence, but then changed her plea to avoid incarceration. *Id.* Unable to live with her false admission, she moved to vacate her plea and, when that motion was granted, she was jailed for two weeks. *Id.* Ultimately, all of the charges against her were dismissed. *Id.*

this system falls disproportionately upon those for whom contact with the criminal justice system is most likely as a matter of sheer probability: the urban poor and people of color.⁸⁰

Contrast that same scenario with a defendant arrested on a new misdemeanor charge committed one day after his or her probation has expired. Because the probationary period has expired, he or she cannot be presented to the court as a probation violator and the defendant is likely to be free to exercise the rights related to challenging the charge without threat of immediate incarceration. And this is most often true even if the alleged crime took place while the person was still on probation.⁸¹ So a defendant arrested and brought to court on a new misdemeanor charge on the last day of his probation can and most often will be incarcerated without bail unless he admits to the new criminal charge, while that same defendant arrested on the same offense but two days later maintains all

⁸⁰ It is a well documented reality that people of color are more likely to be stopped by the police and that their encounters with the police are more likely to result in arrests. Examples abound. The New York City Police Department reported stopping and searching over 500,000 people in 2007; 86% of those stopped and searched were black or Latino. Steven Zeidman, *Time to End Violation Pleas*, N.Y. L.J., Apr. 1, 2008, at 2. In that same year, the Los Angeles Police Department reported that 34.4% of the motor vehicle drivers that it stopped were white, while 18.7% were black, and 37.4% were Hispanic. See Noah Kupferberg, *Transparency: A New Role for Police Consent Decrees*, 42 COLUM. J.L. & SOC. PROBS. 129, 164 app. A (2008). While the numbers of people pulled over appear to have been roughly in proportion to the percentages of each race stopped, a marked difference existed in the number of motorists asked to exit their vehicles and subjected to a search. While just 17.0% of the motorists asked to exit were white, 25.0% were black and 53.2% were Hispanic. *Id.* at 165 app. A. Similarly, of the motorists who were searched once outside of their vehicles, only 11.6% were white, while 31.0% were black and 54.6% were Hispanic. *Id.* Obviously, more stops and more searches will result in more arrests. While any encounter between a police officer and a citizen can escalate into an arrest, people of color are statistically much more likely to be arrested in that kind of encounter. A recent report in Seattle revealed that African-Americans were eight times more likely than whites to be arrested and charged solely with the crime of obstruction, known by local law enforcement officers as “contempt of cop.” Eric Nalder, Lewis Kamb & Daniel Lathrop, *‘Obstructing’ Justice: Blacks Are Arrested on ‘Contempt of Cop’ Charge at Higher Rate*, SEATTLE POST-INTELLIGENCER, Feb. 28, 2008, at A1. In New York City, 87% of the 40,300 people arrested for the lowest-level misdemeanor marijuana possession in 2008 were black or Latino even though research suggests that whites are the heaviest users. Jim Dwyer, *Whites Smoke Pot, But Blacks Are Arrested*, N.Y. TIMES, Dec. 23, 2009, at A24.

⁸¹ Some jurisdictions have enacted statutes that allow them to retain jurisdiction for a “reasonable” period of time following the expiration of the probation period within which hearings can be conducted for violations that occurred while the defendant was on probation. COHEN, *supra* note 2, § 18:19. Even in these jurisdictions, the probation violation hearing is generally avoided if formal revocation proceedings have not commenced prior to the expiration of the probation period. See *United States v. Barton*, 26 F.3d 490, 492 (4th Cir. 1994). In the federal system, for example, unless a warrant or summons has been issued prior to the expiration of probation, the court may not revoke a sentence for probation. 18 U.S.C. § 3565(c) (2006).

of his constitutional rights, including the presumption of innocence and the right to reasonable bail that will generally mean his release from custody. Is the enormous distinction in the treatment of these two defendants justified by any rational public policy? Is it fair? Does it lead to justice?

If the alleged probation violator has the wherewithal and the fortitude to seek a probation violation hearing, that hearing will be one in which virtually all procedural protections for the accused have been removed.⁸² The accused enjoys no right to a trial by jury.⁸³ The rules of evidence are relaxed such that hearsay may be introduced⁸⁴ and illegally obtained evidence may be used.⁸⁵ The right to confront and cross-examine one's accusers is a "conditional right" that a judge can take away.⁸⁶ The burden of proof upon the prosecution, even if the allegation is that the probationer committed a new crime, is significantly reduced.⁸⁷ In some jurisdictions, for example, the government must simply offer evidence such that a judge is "reasonably satisfied" that the probationer has violated a term or condition of probation.⁸⁸ A probationer's ability to obtain discovery in advance of the probation violation hearing is limited,⁸⁹ and because the hearing often takes place before a trial of the new criminal charge is scheduled, probation violation hearings "are frequently held without the benefit of preparation that precedes a criminal trial."⁹⁰

⁸² See, e.g., *Gagnon v. Scarpelli*, 411 U.S. 778, 787-90 (1973); *Morrissey v. Brewer*, 408 U.S. 471, 483-89 (1972); see also *State v. Gautier*, 871 A.2d 347, 359 (R.I. 2005) (noting that probation violation defendants "are afforded considerably less due process protection than that to which they are constitutionally entitled in a full-blown criminal trial").

⁸³ COHEN, *supra* note 2, § 21:49. In fact, revocation hearings may even be presided over by an "independent officer," who is often a probation officer not directly involved in the case. *Morrissey*, 408 U.S. at 486; see also *Gagnon*, 411 U.S. at 781, 786.

⁸⁴ COHEN, *supra* note 2, § 20:11.

⁸⁵ Pa. Bd. of Prob. & Parole v. Scott, 524 U.S. 357, 369 (1998).

⁸⁶ See *Gautier*, 871 A.2d at 359; *Commonwealth v. Durling*, 551 N.E.2d 1193, 1199 (Mass. 1990); see also *United States v. Waters*, 1998 FED App. 0299P (6th Cir.).

⁸⁷ KLEIN, *supra* note 74, at 260-61.

⁸⁸ *Id.* at 260 (internal quotation marks omitted).

⁸⁹ COHEN, *supra* note 2, §§ 21:29-30. Courts have held that, unlike in a criminal proceeding where a defendant is entitled to disclosure of evidence if it is material to his or her case, in a probation violation hearing due process may not be denied if the government fails to disclose evidence, even potentially exculpatory evidence, so long as the government does not plan to use that evidence during a violation proceeding. See *United States v. Neal*, 512 F.3d 427, 436 (7th Cir. 2008); *United States v. Derewal*, 66 F.3d 52, 55 (3d Cir. 1995).

⁹⁰ *Commonwealth v. Cosgrove*, 629 A.2d 1007, 1011 (Pa. Super. Ct. 1993).

In practice, unless the prosecution fails to present any evidence at all, the outcome of a probation violation hearing is often all but a foregone conclusion. When a probationer is found after a hearing to have violated the terms of his or her probation by committing a new crime, that probationer is often sentenced in a fashion that, in reality, is intended to punish the probationer for having committed the new crime. While the sentence is legally justified not as a sentence for the new offense, but rather as a sentence for violating the terms of probation,⁹¹ any honest assessment of the situation acknowledges the truth as perceived by all of the relevant players: the sentence is punishment for the new offense. Often the severity of the probation violation sentence is sufficient to allow the government either to offer a disposition on the new charge with a sentence that functionally merges with the probation violation sentence, or to forgo the prosecution of the new charge altogether. The outcome is that the prosecution gets the sentence it was seeking on the new charge without the burden of ever having to prove it. There is no need, under this system, to have a criminal trial, and our entire system of procedural protections for the accused is left on the sidelines. Something must be done to correct this abuse and restore the legitimacy of our criminal justice system.

IV. ALTERNATIVES TO USING PROBATION AS A DEFAULT SENTENCE

National reports indicate that as many as 80% of adult misdemeanor convictions result in sentences of probation.⁹² The sheer volume of misdemeanor probationers completely overwhelms the system, preventing probation from achieving any measure of effectiveness. It does not have to be this way. Virtually all jurisdictions employ alternative sentencing mechanisms besides probation to resolve criminal cases. If

⁹¹ See *Lucido v. Superior Court*, 795 P.2d 1223, 1230 (Cal. 1990) (“The fundamental role and responsibility of the hearing judge in a revocation proceeding is not to determine whether the probationer is guilty or innocent of a crime, but whether a violation of the terms of probation has occurred”); *Gautier*, 871 A.2d at 361 (“[A] probation-revocation hearing is considered a continuation of the original prosecution for which probation was imposed—in which the sole purpose is to determine whether a criminal defendant has breached a condition of his existing probation, not to convict that individual of a new criminal offense.”); *Cosgrove*, 629 A.2d at 1011 (“It is neither [a probation hearing’s] purpose nor function to serve as a final arbiter of an individual’s guilt or innocence of criminal charges.”).

⁹² Petersilia, *Probation*, *supra* note 6, at 173.

probation is no longer viewed as serving a rehabilitative function, then presumably probation is being used for its retributive or deterrent value. Non-probationary sentences, such as the imposition of time served, of a fine, of community service, or even of a finding of guilt without further punishment, can certainly carry as much retributive value as a probationary period that involves little supervision or, more commonly, no supervision at all. If the retributive value comes from conditions that might be attached to probation, those conditions can be enforced without reliance on probation. Similarly, the deterrent value of a probationary sentence, if there is any in fact, can frequently be equaled by the imposition of a non-probationary sentence.

A. *Debunking the Current Rationales for Probation*

As the system presently exists, the stated rationales supporting the extensive reliance on probation as a sentencing mechanism do not withstand scrutiny. The primary rationale—that probation is cheaper than incarceration and that we simply do not have room in our jails and prisons for all of these defendants—relies on the premise that most or all of those who are placed on probation should be incarcerated. When as many as 80% of all misdemeanor convictions result in a period of probation, it is clear that these defendants are not being placed on probation as an alternative to incarceration.⁹³ What the casual use of probation actually accomplishes for these defendants is the prospect of incarceration for a probation violation that would not otherwise exist if the person had not been placed on probation in the first place. This use of probation does not drive incarceration costs down, but rather quite the opposite.

Another rationale for the reliance on probation as a sentencing mechanism is that probation is a form of retributive sentence. This might make sense in a context in which compliance with probation was onerous. If the vast majority of probationers report rarely or never, and if the level of supervision is diluted to the point of virtual non-existence, it is very hard to comprehend how probation exacts a form of retribution. The honest reality is that for most probationers, probation serves as little more than a noose around their neck,

⁹³ *Id.*

waiting to be tightened when or if they have an encounter with the law. Any system that relies on a future encounter with the law as a triggering mechanism will have a grossly disproportionate impact on the urban poor and people of color.⁹⁴ As noted above, the retributive value of any conditions that might be attached to probation can be achieved by imposing those same conditions without imposing probation.⁹⁵

Yet another rationale for the reliance on probation as a sentencing mechanism is the notion that the mere fact that the offender is on probation will serve as a deterrent to future criminal conduct. But there are several flaws with this reasoning. There is very little empirical data supporting the general notion of deterrence theory with respect to probation.⁹⁶ Experts agree that a low probability threat of a severe sanction is not effective.⁹⁷ For the vast majority of probationers who are obtaining little to no supervision, a violation of probation will occur only if there is an arrest for a new offense. Apprehension for criminal behavior is often a relatively low probability event.⁹⁸ To the extent that a crime involves any premeditation rather than a response to impulse, the offender's estimation of the probability of apprehension will certainly be low in an offender's mind. Presumably the potential sentence for that new crime already serves as a deterrent, so the relevant deterrent value is the differential in deterrence that can be derived solely from one's status as a probationer. With a complete absence of data on this question, it seems relatively

⁹⁴ See *supra* note 80.

⁹⁵ See *supra* Part IV.

⁹⁶ See Petersilia, *Probation*, *supra* note 6, at 154-55; Faye S. Taxman, *Supervision—Exploring the Dimensions of Effectiveness*, FED. PROBATION, Sept. 2002, at 14.

⁹⁷ See COHEN, *supra* note 2, § 1:7 (citing research suggesting that “certainty of punishment is a greater deterrent than severity of punishment”); Michael Tonry, *The Functions of Sentencing and Sentencing Reform*, 58 STAN. L. REV. 37, 52 (2005) (“Current knowledge concerning deterrence is little different than eighteenth-century theorists supposed it to be: certainty and promptness of punishment are much more powerful deterrents than severity.”); Angela Hawken & Mark Kleiman, *H.O.P.E. for Reform: What a Novel Probation Program in Hawaii Might Teach Other States*, AM. PROSPECT, Apr. 10, 2007, http://www.prospect.org/cs/articles?article=hope_for_reform (noting that crime “attracts reckless and impulsive people, for whom deferred and low-probability threats of severe punishment are less effective than immediate and high-probability threats of mild punishment”).

⁹⁸ See Tonry, *supra* note 97, at 53; see also Richard S. Frase, *Punishment Purposes*, 58 STAN. L. REV. 67, 79 (2005) (“[T]he detection rates for most crimes are very low, and the probability of an offender receiving a custody sentence is often less than one out of every one hundred crimes committed.”).

safe to assume that this differential is minimal if not non-existent.

The remaining rationale for the heavy reliance on probation as a sentencing mechanism, if one is honest about how it works, is that it makes the processing of a future criminal charge faster and easier for the prosecution. But this rationale, despite its efficiency, is the one that is so deeply troubling. It makes for very poor public policy choices in a variety of ways—not just who we incarcerate and for how long, but also how quickly we allow offenders to accumulate criminal records that render them unemployable and ineligible for most rental housing. And this process serves to seriously undermine the public perception of the fairness of the system.⁹⁹

If the legitimate justifications for such extraordinarily heavy reliance on probation do not hold up, the obvious solution is to stop using probation as the default non-jail sentence and start relying more heavily on other non-jail dispositions, particularly for misdemeanor offenses. This simple step can help restore the viability and credibility of the probationary sanction by precipitously reducing caseloads. With smaller caseloads, real support and supervision is an attainable goal and there is substantial research suggesting that it can make a real difference.¹⁰⁰ Probation should be imposed sparingly and deliberately in the way in which it was historically intended: as a means of providing support and supervision to those select offenders for whom such support and supervision seems likely to make a difference. There is little value in using probation as a means of monitoring an offender's performance of an identifiable condition of probation; that function can be served either directly by the court or by referral to an outside agency.¹⁰¹ Those offenders who are placed

⁹⁹ A prime example of the public perception of the probation violation system can be found in an article published in the *Providence Phoenix* in 1997, the title of which tells the reader all he or she needs to know. Jody Ericson, *Take a Ride on Rhode Island's Revocation Railroad: Make One False Move While on Probation and Go Directly to Jail*, PROVIDENCE PHOENIX, Oct. 3, 1997, at 9. A similar message can be found a decade later in the magazine *Rhode Island Monthly*. *Guilty, Even While Innocent*, R.I. MONTHLY, Dec. 2008.

¹⁰⁰ See *infra* notes 138-144 and accompanying text.

¹⁰¹ In Rhode Island, a private not-for-profit entity called Justice Assistance has a contract with the courts to monitor compliance with conditions such as community service, domestic violence counseling, substance abuse counseling, mental health counseling, and restitution in cases in which probation is not ordered. See Justice Assistance, www.justiceassistance.org (last visited March 6, 2010). The agency reports back to the court to indicate compliance or non-compliance. *Id.*

on probation must receive much more than just monitoring, but also intervention, support, and supervision.

B. Alternatives to Probation

A wide variety of non-probationary sentences is available. One common non-probationary sentence is the “unconditional discharge” found in many state statutes. In New York, for example, a court may impose a sentence of unconditional discharge “if the court is of the opinion that no proper purpose would be served by imposing any condition upon the defendant’s release.”¹⁰² The statutory provision governing an unconditional discharge in Connecticut uses precisely the same language.¹⁰³ In New Hampshire, an unconditional discharge may be imposed if the court is of the opinion that neither supervision nor any other condition would serve a proper purpose.¹⁰⁴ The statutes in each of these states provide that a sentence of unconditional discharge “is for all purposes a final judgment of conviction.”¹⁰⁵

Pennsylvania uses different language to accomplish essentially the same function, explicitly allowing a court to impose a sentence of “guilt without further penalty.”¹⁰⁶ In other jurisdictions, a plea of guilty followed by a sentence of “time served” has the same effect, creating a criminal conviction and discharging the offender with no further obligations to the court.¹⁰⁷

The statutory sentencing schemes in some states seem designed to discourage or prevent the overuse of probation by statute. In New Hampshire, for example, probation is not a permissible sentence for a Class B misdemeanor and may be imposed only if the offense is a felony or a Class A misdemeanor.¹⁰⁸ Pennsylvania’s Sentencing Guidelines suggest

¹⁰² N.Y. PENAL LAW § 65.20 (McKinney 2009).

¹⁰³ See CONN. GEN. STAT. §53a-34 (2007).

¹⁰⁴ See N.H. REV. STAT. ANN. § 651:2 (VIII) (2007).

¹⁰⁵ CONN. GEN. STAT. §53a-34 (b); N.H. REV. STAT. ANN. § 651:2 (VIII); N.Y. PENAL LAW § 65.20.

¹⁰⁶ PA. CONS. STAT. § 9753 (2007).

¹⁰⁷ See, e.g., COLO. REV. STAT. § 16-7-206 (2009) (providing that the court’s acceptance of a guilty plea “acts as a conviction for the offense); MINN. STAT. § 609.02, subd. 5 (2007) (defining a “conviction” as a plea of guilty or a verdict of guilty that is “accepted and recorded by the court”); N.Y. CRIM. PROC. § 1.20(13) (McKinney 2009) (defining a “conviction” as “the entry of a plea of guilty to, or a verdict of guilty upon, an accusatory instrument”).

¹⁰⁸ N.H. REV. STAT. ANN. § 651:2 (I), (III).

“the use of the least restrictive, non-confinement sentencing alternatives” appropriate to the case, including the “determination of guilt without further penalty.”¹⁰⁹ Maine has gone much further, prohibiting the use of probation as a sentence in the majority of misdemeanor cases and making a sentence of unconditional discharge the default sentence even in those situations where probation is permissible.¹¹⁰ The Maine statute provides that a court may impose probation as a sentence only if it affirmatively finds that “the person is in need of the supervision, guidance, assistance or direction that probation can provide.”¹¹¹ In the alternative, an offender “for whom the court determines that no other authorized sentencing alternative is appropriate punishment *must* be sentenced by the court to an unconditional discharge.”¹¹² The adoption of these sentencing policies in Maine made an enormous difference in a very short period of time, with the number of probationers under supervision declining by over one-third between 2004 and 2007.¹¹³ During the same time frame the percentage of prison inmates incarcerated in Maine on a probation violation declined from 30% of the prison population to 25%.¹¹⁴ By 2005, Maine was among the top ten states in the country with the smallest percentage of its adult population under probation supervision.¹¹⁵ In New York, substantial use of the sentence of “time served” has helped keep probation numbers quite low.¹¹⁶ Statewide in 2007, more than 12% of misdemeanor convictions in New York were

¹⁰⁹ 10A PA. PRACTICE SERIES § 27:14, *Driving Under the Influence* (2009).

¹¹⁰ See ME. REV. STAT. ANN. tit. 17-A, § 1201 (2009).

¹¹¹ *Id.* § 1201(2).

¹¹² *Id.* § 1346 (emphasis added). The commentary to § 1201 notes that “probation should be used if it appears that the convicted person would be helped thereby” but that, “[a]bsent such a need, an unconditional discharge is warranted.” *Id.* § 1201 cmt.

¹¹³ MARK RUBIN, TARGETED INTERVENTIONS COULD EASE MAINE’S PRISON AND JAIL POPULATIONS (2008), available at http://muskie.usm.maine.edu/justiceresearch/Publications/Adult/Targeted_Interventions_Could_Ease_ME_Prison_Jail_Population.pdf.

¹¹⁴ *Id.*

¹¹⁵ ANNUAL PROBATION SURVEY, 2005, *supra* note 44, at 3 tbl.1.

¹¹⁶ Like Maine, in 2005 New York was among the top ten states in the country with the smallest percentage of its adult population under probation supervision. *Id.* As will be developed elsewhere, this status can also be attributed to the widespread use of conditional discharge sentences and the imposition of fines. See *infra* notes 123-128 and accompanying text.

resolved with a sentence of time served,¹¹⁷ while in New York City the percentage exceeded 17%.¹¹⁸

The retributive and deterrent value of an unconditional discharge, a sentence of guilt without further penalty, or a sentence of time served is clear. In all of the statutory schemes cited above, the imposition of a sentence creates a criminal conviction. The mere fact of the criminal conviction carries all of the same retributive characteristics of a period of probation that entails no supervision. The criminal conviction is a matter of public record and available for all of the world to see. The stigma connected with being a convicted criminal is equally poignant without the accompanying period of probation, as are the adverse consequences for future employment and housing. And the fact of the conviction remains accessible and available for use against the defendant in any future court proceeding or sentence. The Supreme Court of Pennsylvania acknowledged this general logic some time ago:

In some instances, the court may decide that the needs of justice are fulfilled by a determination of guilt alone, without necessity for further penalty. The shame and trauma of public conviction may be punishment enough and there may be no need of any plan for 'reformation' or control. In such cases, the courts should be free to make such a judgment without requiring useless probation.¹¹⁹

Whatever deterrent value may be served by an offender's awareness that the commission and detection of a new crime while on probation may carry an enhanced penalty—and there is no available evidence to suggest that such deterrent value even exists—can be replicated by a more intelligent graduated sentencing scheme for repeat offenders.

The unconditional discharge is, of course, far from the only way to achieve the desired result of reducing excess reliance on probation while at the same time imposing a sentence that has retributive and deterrent value. Most states list a variety of alternative non-jail sentences in their array of sentencing possibilities, including community service, restitution, various counseling or educational regimens, and

¹¹⁷ N.Y. STATE DIV. OF CRIMINAL JUSTICE SERVICES, DISPOSITION OF ADULT ARRESTS, NEW YORK STATE 5 (2009), available at <http://www.criminaljustice.state.ny.us/crimnet/ojsa/dispos/nys.pdf> [hereinafter DISPOSITION OF ADULT ARRESTS, NEW YORK STATE].

¹¹⁸ *Id.*

¹¹⁹ Commonwealth v. Rubright, 414 A.2d 106, 109 (Pa. 1980) (quoting S. TOLL, PENNSYLVANIA CRIMES CODE ANNOTATED § 1323 (Supp. 1978)).

finer. These conditions can and do have retributive value. Indeed, for many offenders a community service obligation is much more onerous than a period of probation, particularly if that probation is essentially unsupervised.¹²⁰ Research studies have concluded that, as measured by recidivism rates, a community service sentence has no less deterrent value than a sentence of probation.¹²¹ Each of these sorts of conditions can be monitored either directly by the court through a future court appearance or through some outside agency without any need for probationary supervision. Indeed, the use of a probationary sentence to accomplish nothing more than monitoring of compliance with a specific condition is one of the primary reasons that probation has been so grossly overused.

The easiest mechanism for overseeing the imposition of a specific alternative sanction is the use of the “conditional discharge.” In New York, for example,

[A] court may impose a sentence of conditional discharge for an offense if the court, having regard to the nature and circumstances of the offense and to the history, character and condition of the defendant, is of the opinion that neither the public interest nor the ends of justice would be served by a sentence of imprisonment and that probation supervision is not appropriate.¹²²

Other states have quite similar provisions. In New Hampshire, for example, a defendant “may be sentenced to a period of conditional discharge if such person is not imprisoned and the court is of the opinion that probationary supervision is unnecessary, but that the defendant’s conduct should be according to conditions determined by the court.”¹²³

Reliance on non-probationary alternative sentences has allowed some jurisdictions to keep their probation rates relatively under control. In 2007, almost one-third of all misdemeanor convictions in the state of New York resulted in a sentence of conditional discharge.¹²⁴ When added to the misdemeanor cases resolved by sentences of time served and those resolved with the imposition of a fine, the total percentage of misdemeanor convictions resolved without resort

¹²⁰ Michael Tonry, describing alternative sentencing in Europe, reports that “In law and in practice, CSOs (community service orders) are regarded in England as more intrusive and punitive than probation.” TONRY, *supra* note 24, at 122.

¹²¹ *Id.* at 122-23.

¹²² N.Y. PENAL LAW § 65.05 (1)(a) (2009).

¹²³ N.H. REV. STAT. ANN. § 651:2 (VI)(a) (2007).

¹²⁴ See DISPOSITION OF ADULT ARRESTS, NEW YORK STATE, *supra* note 117, at 5.

to probation or incarceration was just under 75%.¹²⁵ Fewer than 5% of misdemeanor convictions resulted in probationary sentences.¹²⁶ In New York City the numbers were even more pronounced, with over 40% of misdemeanor convictions being resolved with a conditional discharge and not even 1% sentenced to probation.¹²⁷ Not surprisingly, then, in 2005 New York was listed among the top ten states in the country with the lowest percentage of its adult population under probationary supervision.¹²⁸ And the vast majority of those adults on probation appear to be on probation for felony offenses, presumably a much wiser use of the limited supervisory resources available to the probation department. Similarly, as noted earlier, Maine has achieved substantial reductions in number of probationers by prohibiting the use of probation as a misdemeanor sentence except on a select category of misdemeanors.¹²⁹

Data in North Carolina indicate that of all cases resolved with a sentence defined as “community punishment” only one-third were sentenced to a period of supervised probation.¹³⁰ Despite that fact, North Carolina’s percentage of adults on probation is nearly as high as the national average.¹³¹ This anomalous result may be explained by what appears to be a quite unfortunate and ill-advised reliance on unsupervised probation, which is imposed in 44% of community punishment cases.¹³² That undue reliance may in turn be explained by the Criminal Code Commission’s rejection of a recommendation to include unconditional discharge as a sentencing option.¹³³ If the high volume of unsupervised probation cases were excluded, the percentage of adults being supervised by probation officers would presumably be significantly reduced.

¹²⁵ *Id.*

¹²⁶ *Id.*

¹²⁷ See DISPOSITION OF ADULT ARRESTS, NEW YORK CITY, *supra* note 118, at 5.

¹²⁸ ANNUAL PROBATION SURVEY, 2005, *supra* note 44, at 3 tbl.1.

¹²⁹ See *supra* notes 110-115 and accompanying text.

¹³⁰ N.C. SENTENCING AND POLICY ADVISORY COMM’N, STRUCTURED SENTENCING STATISTICAL REPORT FOR FELONIES AND MISDEMEANORS 50-51 (Feb. 2008), available at <http://www.nccourts.org/Courts/CRS/Councils/spac/Documents/06-07statisticalreport.pdf>.

¹³¹ ANNUAL PROBATION SURVEY, 2005, *supra* note 44, at 3 tbl.1.

¹³² N.C. SENTENCING AND POLICY ADVISORY COMM’N, *supra* note 130, at 50-51.

¹³³ See N.C. GEN. STAT. § 15A-1301 cmt. (2009).

C. *The Potential Benefits of Reform*

A substantial reduction in probation caseloads, particularly on the misdemeanor level, can have significant crime control ramifications with what would seem to be very little to no risk of adverse consequences. National statistics reveal that 75% of misdemeanor probationers complete their period of probation without violation.¹³⁴ Since the majority of these probationers receive little to no support or supervision, one logical conclusion is that these offenders were not in need of any supervision.¹³⁵ If that is the case, any potential benefits of the probationary sentence would seem to be far outstripped by the costs.¹³⁶ The costs of placing enormous numbers of misdemeanor defendants on probation are very real. There are administrative and transactional costs connected to each probationer, even if he or she is totally unsupervised.¹³⁷ For those probationers who do not succeed, there are costs connected to the entire violation process as well as to the potential escalation of a non-jail case into incarceration. With each failure the reputation of probation as a potentially effective crime control mechanism suffers. But perhaps most importantly, the opportunity cost—in both human and financial terms—connected with failing to provide actual support and supervision in a fashion that has some possibility of efficacy is immeasurable.

Despite the popularity of the “nothing works” philosophy that first took hold in the 1970s, in fact there is a great deal of evidence that the provision of support services and supervision can work quite well in reducing recidivism and helping to control crime. The study that created the “nothing works” furor came under persistent and compelling attack from the moment of its publication. As a National Academy of Sciences Panel concluded in reevaluating the original “nothing works” study just three years after its publication, “when it is

¹³⁴ See COHEN, *supra* note 2, § 1:23 n.3; Petersilia, *Probation*, *supra* note 6, at 180-81.

¹³⁵ Another logical conclusion may be that some of these probationers violated their probation but the violations went undetected. The higher the number of probationers in this category, the less value probation would seem to have as any sort of deterrent to future criminality.

¹³⁶ See Petersilia, *Probation*, *supra* note 6, at 181 (questioning “the wisdom of placing such low-risk persons on probation in the first place” because the costs appear to outstrip the benefits).

¹³⁷ *Id.*

asserted that ‘nothing works,’ the panel is uncertain as to just what has even been given a fair trial.”¹³⁸ The programs that made up the basis of the “nothing works” study were “often not only underfunded and understaffed, but typically staffed by poorly trained and often unmotivated people.”¹³⁹

More recent research strongly supports the proposition that support services and supervision can have a meaningful impact on recidivism. In a leading study published in 1987, Professors Paul Gendreau and Robert Ross surveyed over 200 studies on rehabilitative programs, concluding that “successful rehabilitation of offenders had been accomplished, and continued to be accomplished quite well.”¹⁴⁰ They found that “reductions in recidivism, sometimes as substantial as 80 percent, had been achieved in a considerable number of well-controlled studies.”¹⁴¹ Research continuing on through the 1990s, now known as the “what works” literature, consistently found similar results.¹⁴² In the case of drug addicted offenders, there is “rather solid empirical evidence that ordering offenders into treatment, and getting them to participate, reduces recidivism.”¹⁴³ But these reductions in recidivism were seen only in “programs in which offenders both received surveillance (e.g., drug tests) and participated in relevant treatment.”¹⁴⁴

The plain reality is that probation can have a rehabilitative impact only if we return to the rational and judicious use of probation as a criminal sanction, allowing probation officers to engage constructively with probationers. That requires a manageable case load that can involve actual interaction and supervision, complete with referrals to viable treatment programs and adequate follow up to assure compliance. The lost opportunity to have a meaningful impact on an offender’s prospects for rehabilitation cannot be justified.

¹³⁸ See Miller, *supra* note 1.

¹³⁹ *Id.* (quoting criminologist Elliott Currie).

¹⁴⁰ Paul Gendreau & Robert R. Ross, *Revivification of Rehabilitation: Evidence from the 1980s*, 4 JUST. Q. 349, 350-51 (1987).

¹⁴¹ *Id.*

¹⁴² See generally WHAT WORKS?: REDUCING REOFFENDING (James McGuire ed., 1995).

¹⁴³ Petersilia, *Crime Control*, *supra* note 36, at 489.

¹⁴⁴ *Id.*

V. THE RETURN TO A SYSTEM THAT APPROXIMATES JUSTICE

If the American criminal justice system is to be true to its name and its purported mission, it must stop using the probation violation system as an end-run around due process to resolve new charges for those who are charged with committing a new offense while on probation. New criminal allegations should be prosecuted using the procedural mechanisms that have been developed throughout our history for the prosecution of criminal charges, whether or not the accused happens to be on probation at the time of the alleged offense or prosecution. While it may be appropriate to hold a probationer to a higher standard of behavior, it is not appropriate to let a probationer be prosecuted for a new criminal offense under a process that has been stripped of virtually all of its procedural protections. Creating protections against this sort of abuse of the probation violation system will reduce the temptation on the part of some sentencing judges to use probation as nothing more than a noose around an offender's neck. Correcting this misguided use of probation will create both the appearance and, more importantly, the reality of observing constitutional principles and assuring fundamental fairness in this very broken part of the criminal justice system.

It is plain to any observer, despite judicial protestations to the contrary, that judges frequently impose probation violation sentences based upon a new criminal allegation in a fashion that is designed to punish the probationer for the new criminal allegation. The consequences in terms of fairness, both in actuality and in the public perception, are devastating. In Rhode Island, media coverage of the issue has generated headlines including "Found Innocent, But Still Jailed,"¹⁴⁵ "Guilty, Even While Innocent,"¹⁴⁶ and "Take a Ride on Rhode Island's Revocation Railroad."¹⁴⁷ Each of these articles lays out in compelling terms multiple scenarios in which probation violation hearings were held in advance of, and used as substitutes for, criminal trials based upon new criminal allegations. Even when a probationer has been acquitted after a trial of the new criminal charge, a lengthy sentence based upon that same conduct continues unabated. Often, the

¹⁴⁵ John Hill, *Found Innocent, But Still Jailed*, PROVIDENCE J., Aug. 9. 2009, at A1.

¹⁴⁶ *Guilty, Even While Innocent*, *supra* note 99.

¹⁴⁷ Ericson, *supra* note 99.

prosecution of a new criminal charge is abandoned or short-circuited after a violation hearing because the probationer has already received the desired sentence on the probation violation and the adversarial testing of the new criminal allegation never takes place.

Tellingly, there is but one context in which the courts have routinely recognized the inadequacy of using the probation violation hearing as a substitute for a criminal trial: when the accused wins. The courts seem to have little trouble upholding lengthy sentences following from probation violation hearings conducted with minimal procedural protections for the innocent. But when a hearing court has found that the government's evidence is insufficient to meet even the reduced burden of proof used at a violation hearing, the majority of jurisdictions have rejected the application of collateral estoppel to prevent the government from nonetheless proceeding with a trial based on the same allegations.¹⁴⁸ When faced with a not guilty finding at a violation hearing, those courts have maintained that the criminal trial process is "the intended forum for ultimate determinations as to guilt or innocence of newly alleged crimes"¹⁴⁹ and that applying collateral estoppel to prevent the criminal prosecution of the new charge "would undesirably alter the criminal trial process by permitting informal revocation determinations to displace the intended factfinding function of the trial."¹⁵⁰ The Superior Court of Pennsylvania, in rejecting the application of collateral estoppel to a not guilty finding at a probation violation hearing, explained that:

It is neither the[] purpose nor function [of a violation hearing] to serve as a final arbiter of an individual's guilt or innocence of criminal charges. It is only through a criminal trial at which the defendant is presumed innocent and the [government] bears the burden of proof of guilt beyond a reasonable doubt that contested

¹⁴⁸ See, e.g., *Lucido v. Superior Court*, 795 P.2d 1223, 1232-33 (Cal. 1990); *State v. McDowell*, 699 A.2d 987, 990 (Conn. 1997); *Commonwealth v. Cosgrove*, 629 A.2d 1007, 1011 (Pa. Super. Ct. 1993); *State v. Gautier*, 871 A.2d 347, 360-61 (R.I. 2005); *State v. Brunet*, 806 A.2d 1007, 1008 (Vt. 2002). See generally George L. Blum, Annotation, *Determination that State Failed to Prove Charges Relied Upon for Revocation of Probation as Barring Subsequent Criminal Action Based on Same Underlying Charges*, 2 A.L.R. 5th 262 (1992 & Supp.) (collecting and discussing cases deciding whether the government's failure to prove a probation violation at a revocation hearing precludes a subsequent criminal prosecution based upon the same underlying conduct).

¹⁴⁹ *Lucido*, 795 P.2d at 1230-31.

¹⁵⁰ *Id.* at 1229.

issues of criminal culpability are determined with finality. To cede this responsibility to a setting that does not adhere to the procedural safeguards necessary for a fair adjudication of guilt, such as a probation revocation hearing, would result in a perversion of the criminal justice system.¹⁵¹

More than one judge has described this process as a “Heads I win, tails I flip again” proposition,¹⁵² allowing the government to present minimal evidence at a violation hearing with an option to try again at a trial if unsuccessful. The accused, on the other hand, must litigate fully at the probation violation hearing because he or she faces dire consequences if found to be a violator.

This scenario can easily be avoided by sequencing the events differently. If a new criminal charge is adjudicated in advance of the probation violation hearing, the substitution of the violation hearing for the trial will never take place. If the probationer admits to or is convicted of the new offense, the probation violation has been established without sacrificing the procedural screening mechanisms upon which we rely. And if the probationer is acquitted at a trial or the charge is dismissed, under present law the prosecution can generally still proceed with a probation violation allegation.¹⁵³ The fact that prosecutors in so many jurisdictions resist all attempts to sequence events in this fashion, despite pleas from the American Bar Association¹⁵⁴ and sometimes from their own courts¹⁵⁵ to do so, reveals a great deal about the motivations

¹⁵¹ *Cosgrove*, 629 A.2d at 1011.

¹⁵² *Lucido*, 795 P.2d at 1243 (Broussard, J., dissenting) (internal quotation marks omitted); *McDowell*, 699 A.2d at 992 (Berdon, J., dissenting) (internal quotation marks omitted) (quoting *Lucido*, 795 P.2d at 1243 (Broussard, J., dissenting)); *Brunet*, 806 A.2d at 1017 (Johnson, J., dissenting) (internal quotation marks omitted) (quoting *Lucido*, 795 P.2d at 1243 (Broussard, J., dissenting)).

¹⁵³ See COHEN, *supra* note 2, § 22:15. Simple fairness, in addition to respect for the values underlying the criminal justice system, would suggest that this practice be abandoned.

¹⁵⁴ AMERICAN BAR ASSOC., ABA STANDARDS FOR CRIMINAL JUSTICE, SENTENCING § 18-7.4 (h) (3d ed. 1994) [hereinafter ABA STANDARDS FOR CRIMINAL JUSTICE, SENTENCING] (“When an alleged violation is based solely on the alleged commission of another offense, the rules should provide that the final hearing on the alleged violation ordinarily should be held after disposition of the new criminal charge.”); see also AMERICAN BAR ASSOC., STANDARDS RELATING TO PROBATION § 5.3 (1970) [hereinafter ABA STANDARDS RELATING TO PROBATION] (“A revocation proceeding based solely upon commission of another crime ordinarily should not be initiated prior to the disposition of that charge.”).

¹⁵⁵ See, e.g., *People v. Coleman*, 533 P.2d 1024, 1046 (Cal. 1975) (“[W]e wish to note that the most desirable method of handling the problems of concurrent criminal and probation revocation proceedings may well be for revocation proceedings not even to be initiated until after disposition of the related criminal proceedings.”); *State v.*

behind their use of probation and the probation violation system.

Several procedural requirements could be implemented that would make the probation violation process much fairer. Sequencing events so that a trial on a new criminal allegation precedes a probation violation hearing based upon that same conduct, the most obvious of these reforms, has been promoted by the American Bar Association for decades. Section 18-7.4 (h) of the American Bar Association Standards for Criminal Justice provides that “[w]hen an alleged violation is based solely on the alleged commission of another offense, the rules should provide that the final hearing on the alleged violation ordinarily should be held after disposition of the new criminal charge.”¹⁵⁶ Commentary to an earlier version of the standards, explaining this sequencing recommendation, explicitly recognized the danger at issue, noting that the relaxation of the rules of evidence, the absence of a jury, and the lowering of the burden of proof “can lead to an abuse of the proceeding by basing revocation upon a new criminal offense when the offense could not be proved in an ordinary criminal trial.”¹⁵⁷ The commentary further noted that “it would be unseemly for the probation court to conclude, counter to the result of a criminal trial, that an offense has occurred and that it could provide a basis for a revocation.”¹⁵⁸

The First Circuit has likewise acknowledged the serious potential for abuse in holding a violation hearing based upon a new criminal allegation before the disposition or trial of the new charge. In *Flint v. Mullen*,¹⁵⁹ a case in which a probationer was sentenced to twelve years on a violation based solely on a criminal charge upon which he was ultimately acquitted, the court indicated its view that “it would be preferable for the state to have held the violation hearing after the . . . trial,”¹⁶⁰

Chase, 588 A.2d 120, 123 (R.I. 1991) (“The disposition of the instant case should encourage the state to initiate future probation-revocation proceedings with more concern for judicial economy.”); *State v. Begins*, 514 A.2d 719, 723 (Vt. 1986) (“We caution against a policy of scheduling probation revocation hearings prior to trial, without exercising discretion in each case. . . . [T]he better method of dealing with problems of concurrent criminal and probation revocation jurisdiction is to postpone the probation proceedings until after disposition of related criminal proceedings . . .”).

¹⁵⁶ ABA STANDARDS FOR CRIMINAL JUSTICE, SENTENCING, *supra* note 154, § 18-7.4 (h).

¹⁵⁷ ABA STANDARDS RELATING TO PROBATION, *supra* note 154, § 5.3 cmt.

¹⁵⁸ *Id.*

¹⁵⁹ 499 F.2d 100 (1st Cir. 1974).

¹⁶⁰ *Id.* at 105.

adding that it could “see little public interest served by this kind of timing.”¹⁶¹ As the court explained:

Were the order reversed, the alleged violator could be held on high bail or without bail if he were a poor bail risk. If there were a criminal conviction, the subsequent violation decision would be simple; if there were an acquittal, the court conducting the violation hearing could proceed with full knowledge of that result, remaining free to weigh evidence by a lower standard, but having in mind the acquittal. The result is apt to be, if not also appear, more just.¹⁶²

The Supreme Courts of Vermont and California have both opined that the “better” or “most desirable” method of handling concurrent criminal and probation violation proceedings is for the trial to proceed first.¹⁶³

One is hard pressed to find legitimate justifications for holding a violation hearing based upon a new criminal allegation in advance of a criminal trial. The only justification that appears in any of the case law concerns the issue of detention in advance of the hearing, particularly in light of the constitutional requirement that a violation hearing take place “within a reasonable time after the [probationer] is taken into custody.”¹⁶⁴ Several responses to this potential objection make its resolution rather easy. In many settings, the primary justification for detention lies not in the person’s status as a probationer, but rather in his or her status as a person with a criminal history accused of a new crime. That detention can be accomplished by the setting of appropriate bail (or holding the accused without bail when permitted) on the new criminal offense. In such a scenario, there would be no need for the prosecution to file the probation violation allegation until after the new criminal charge is resolved. Another response could be to detain the probationer on the alleged violation and put the decision about sequencing in the hands of the probationer, allowing the probationer to waive the right to a prompt violation hearing in order to delay it until after the resolution of the new criminal charge.

When one pushes past objections about detention while awaiting a violation hearing, it becomes apparent that a primary reason for sequencing the events as many states do is

¹⁶¹ *Id.*

¹⁶² *Id.*

¹⁶³ *State v. Begins*, 514 A.2d 719, 723 (Vt. 1986); *People v. Coleman*, 533 P.2d 1024, 1046 (Cal. 1975).

¹⁶⁴ *Morrissey v. Brewer*, 408 U.S. 471, 488 (1972).

to accomplish just what must be prohibited: the creation of a system in which the right to a trial by jury, the right to fully confront witnesses, and the right to put the government to its burden of proof beyond a reasonable doubt recede into the background and prosecution by violation hearing becomes the norm. Perhaps the most obvious and extreme abuses come in the cases that buck the trend of this shadow system, those in which the probationer prevails at the probation violation hearing and is nonetheless prosecuted for the underlying new crime, and those in which the probationer, having been found in violation and sentenced severely, prevails at the criminal trial. In either scenario, the perception, if not the reality, is that an end-run has been made around the Constitution. Even if one were to tolerate a system in which a probation violation hearing comes first, these particular abuses could be stopped.

If the government chooses to present a probationer as a violator and move forward with a violation hearing, it does not seem unreasonable to force the government to live with the consequences of its decision. If the government, even with the benefit of relaxed rules of evidence, cannot meet a reduced burden of proof at a violation hearing, it is unclear why the government should then be allowed another chance to try to prove the same allegations. But in the majority of jurisdictions in this country, the government enjoys just that privilege. Ironically, as noted earlier, courts ruling in this fashion have generally relied on the argument that the procedures employed at a violation hearing are insufficiently reliable to justify using them to resolve a criminal charge. These courts seem not to recognize the irony that this is precisely how probation violation hearings are used on a daily basis in thousands of cases. Common principles of collateral estoppel should be employed, as they are in some jurisdictions,¹⁶⁵ to prevent the government from relitigating an issue that it lost.

¹⁶⁵ See *People v. Bone*, 412 N.E.2d 444, 447 (Ill. 1980) (noting that collateral estoppel will apply when “an issue of ultimate fact was decided in the prior revocation proceeding which was determinative of the issues in the criminal prosecution for the offenses”); *People v. Anzures*, 670 P.2d 1258, 1260 (Colo. Ct. App. 1983) (noting that collateral estoppel will apply although the revocation hearing and criminal charge are “technically based on the commission of wholly separate offenses, but where the same facts are determinative of guilt for each”); *State v. Bradley*, 626 P.2d 403, 406 (Or. Ct. App. 1988) (noting that an “express finding on a matter of fact material to a probation revocation proceeding will collaterally estop the state from” relitigating the same issue where the issue was “fully litigated at the probation revocation proceeding” (emphasis omitted)).

When a probationer is found to have been in violation of probation based upon a new criminal charge and is ultimately acquitted of that new charge, again the probationer seems to be a victim of a gaming of the system. This scenario would be avoided by sequencing the events properly, but if the hearing must proceed first, it does not seem unreasonable to let the issue be revisited in the light of an acquittal after a full trial replete with constitutional protections. Another way of reducing the likelihood of this scenario, and of enhancing the reliability and fairness of a probation violation hearing, would be to elevate the government's burden of proof at a probation violation hearing. The greater the disparity between the government's burden at a violation hearing and the government's burden at a trial, the greater the likelihood of unjust or disparate outcomes. When the government's burden at a violation hearing is as low as the "reasonable satisfaction" of the judge, it is far from surprising when a charge that cannot be proved beyond a reasonable doubt results in a finding of violation. Do we really intend to have a system in which probationers can be convicted of new crimes based on a lesser standard of proof achieved through the introduction of evidence that would normally be inadmissible? The distance between the language found in court decisions explaining the purported purpose of probation violation hearings and the reality as experienced by participants in the criminal justice system is staggering.

VI. CONCLUSION

Something has gone terribly wrong in the American criminal justice system. In the process of moving from a system focused on the rehabilitative potential of the defendant to a system myopically focused on retribution, we have trampled not only upon the tool with the greatest rehabilitative potential, but also upon the due process protections that we supposedly hold most dear. By using probation as a default sentence for all of those whom we choose not to incarcerate, we have created burgeoning caseloads that prevent probation from serving any useful rehabilitative function. Many probationers go without any supervision whatsoever, and those who are in need of social services and support rarely get it. Not surprisingly, then, high percentages of probationers do not succeed on probation. Many of those wind up incarcerated, even though the system's conclusion was that the underlying

crime did not justify incarceration. The “crime” that we punish with incarceration is the inability to live up to the terms and conditions of probation, even if it was entirely unrealistic to expect the probationer to live up to those terms and conditions and entirely predictable that the probationer would fail. This is a peculiar way indeed to make determinations about whom to incarcerate. A far more logical system would use probation only when it can serve a real function. In that fashion, probation officers could actually do their jobs, future criminality could be dealt with on its own terms, and a simple failure to abide by imposed norms of behavior and conformity would not become a cause for incarceration.

One reason it may be hard to convince some constituencies to abandon the abuse of probation is that they are wedded to what follows from that abuse: a shadow criminal justice system in which huge numbers of cases are processed not through the due process protections that come with the prosecution of a criminal charge, but through a violation hearing process that is devoid of virtually all of these protections. This process is certainly efficient, but does not reflect the values of justice that our system is supposed to represent. It is simply inappropriate to hold a violation hearing at which a criminal charge is adjudicated not through a criminal trial replete with protections for the innocent, but rather through a truncated procedure designed for a very different purpose.

If we persist in proceeding with a probation violation hearing in advance of a criminal trial on a new charge, we can at least aspire to level the playing field just a little bit. Those jurisdictions with very low burdens of proof can require at least proof by a fair preponderance of the evidence. And if the accused manages to prevail at a probation violation hearing based solely on a new criminal charge, traditional principles of collateral estoppel should prevent the government from trying a second time to prosecute the accused for the same behavior.

We are all losers when we engage in a process for which the thinly veiled legal justification is readily transparent to all as a fraud. All criminal charges should be adjudicated on the merits. When a probation violation is predicated on a new criminal charge, the adjudication of that new charge should normally resolve the issue of whether or not the terms and conditions of probation have been violated. If the criminal charge cannot be successfully prosecuted, it should follow that the probation violation should be dismissed. And if we insist on

a process that adjudicates the probation violation first, we can at least abide by procedural rules that more closely approximate fairness. We owe it to ourselves to restore the public's faith in the integrity of the prosecutorial function and to put the concept of justice back into the criminal justice system.

Dignity, Legal Pluralism, and Same-Sex Marriage

Jeffrey A. Redding[†]

For the first time in living memory, we can realistically hope to see lesbian and gay couples happily joined on an equal footing with our non-gay brothers and sisters—if those who favor equality can put aside their divisions and unite to secure ultimate victory. For this reason, I have urged that we end, or at least suspend, the intra-community debate over whether to seek marriage. The ship has sailed.¹

—Evan Wolfson (1993)

Only marriage between a man and a woman is valid or recognized in California.²

—Proposition 8 (2008)

[†] Assistant Professor, Saint Louis University School of Law. Portions of this Article were presented previously at the American Society of Comparative Law's 2008 Annual Meeting, two symposia on California's Proposition 8 at Chapman University School of Law during the 2008-09 academic year, and a faculty workshop at Saint Louis University's School of Law. I thank participants at each forum for their questions and feedback, and also Mary Anne Case, Glenn Cohen, Katherine Darmer, Adrienne Davis, Moon Duchin, Chad Flanders, Holning Lau, Robert Leckey, Sebastian Lourido, Eric Miller, Doug Nejaime, Karen Petroski, Marc Poirier, Darren Rosenblum, Laura Rosenbury, Kerry Ryan, Pete Salsich, Molly Walker Wilson, and Robin Fretwell Wilson for especially insightful individual conversations and suggestions. Dallin Merrill, Kate Mortensen, and Kevin Salzman all provided excellent research assistance for this Article, as did the Saint Louis University Law Library staff (and especially Peggy McDermott). Both Yale's Fund for Lesbian and Gay Studies (FLAGS) and Saint Louis University School of Law provided generous support for research leading to this Article. Of course, all errors of fact and judgment remain mine alone. This Article is dedicated to Rehaan Engineer, for never letting his dignity get in the way of his love.

¹ Evan Wolfson, *Crossing the Threshold: Equal Marriage Rights for Lesbians and Gay Men and the Intra-Community Critique*, 21 N.Y.U. REV. L. & SOC. CHANGE 567, 611 (1994). Evan Wolfson is Executive Director of the organization Freedom to Marry. See <http://www.freedomtomarry.org/> (last visited Mar. 7, 2010). In 1993, he was writing in the wake of the Hawaii Supreme Court decision in *Baehr v. Lewin*, 852 P.2d 44 (Haw. 1993), which found that Hawaii's prohibitions on same-sex marriage potentially violated the Hawaiian state Constitution's guarantees of equality. *Id.* at 67.

² Cal. Prop. 8 (2008) (codified as CAL. CONST. art. I § 1.5). Proposition 8 is also known as the California Marriage Protection Act, and it was approved by voter-ballot initiative and enacted into law by Californians on November 4, 2008. For more information on Proposition 8, see <http://www.voterguide.sos.ca.gov/title-sum/prop8-title-sum.htm>.

There is a position—not at all unfamiliar in contemporary discussion—which says that to be a citizen is essentially and simply to be under the rule of the uniform law of a sovereign state. . . . [T]his is a very unsatisfactory account of political reality in modern societies.³

—Archbishop of Canterbury Rowan Williams (2008)

American family law is in tumult, and that is a good thing. The debate over same-sex marriage has opened the floodgates of contestation, debate, and imagination over the regulation of interpersonal relationships in the United States.⁴ The faltering of one major American taboo—that of same-sex intimacy—has encouraged citizens, activists, and lawyers to question other social and legal taboos and, also, to attempt to construct new ones. For example, active debates concerning whether the state might permit and regulate (or at least decriminalize) polygamy are now occurring,⁵ as are discussions concerning the wisdom of the state sponsoring marriage in the first place.⁶ Startling proposals to constitutionalize family law

³ See Rowan Williams, Archbishop of Canterbury, *Civil and Religious Law in England: A Religious Perspective*, Lecture at the Royal Courts of Justice (Feb. 7, 2008), available at www.archbishopofcanterbury.org/1575.

⁴ See Kerry Abrams & Peter Brooks, *Marriage as a Message: Same-Sex Couples and the Rhetoric of Accidental Procreation*, 21 *YALE J. L. & HUMAN.* 1, 2 (2009) (arguing that “[t]he more American courts, and the American people, weigh in on same-sex marriage, the more problematic the very concept of ‘marriage’ becomes”).

⁵ See, e.g., Michèle Alexandre, *Lessons from Islamic Polygamy: A Case for Expanding the American Concept of Surviving Spouse So As to Include De Facto Polygamous Spouses*, 64 *WASH. & LEE L. REV.* 1461, 1461 (2007) (discussing the desirability of creating “legal remedies for vulnerable individuals living and operating in de facto polygamous unions”); Courtney Megan Cahill, *Same-Sex Marriage, Slippery Slope Rhetoric, and the Politics of Disgust: A Critical Perspective on Contemporary Family Discourse and the Incest Taboo*, 99 *NW. U. L. REV.* 1543, 1548 (2005) (questioning “the privileged position that the incest taboo has maintained in the law governing sexuality and the family . . . and [] propos[ing] that the law reappraise the extent to which disgust, rather than reasoned argument, sustains laws directed at sexual and familial choice”); Adrienne Davis, *The Game of Love: Polygamy, Default Rules, and Bargaining for Equality* (Wash. Univ. Sch. of Law, Working Paper No. 09-09-01, 2009), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1480906; Emily J. Duncan, *The Positive Effects of Legalizing Polygamy: “Love is a Many-Splendored Thing”*, 15 *DUKE J. GENDER L. & POL’Y* 315 (2008); Shayna M. Sigman, *Everything Lawyers Know About Polygamy is Wrong*, 16 *CORNELL J.L. & PUB. POL’Y* 101 (2006) (arguing against the criminalization of polygamy, but not necessarily for formal recognition by the state of polygamous relationships).

⁶ See, e.g., NANCY D. POLIKOFF, *BEYOND (STRAIGHT AND GAY) MARRIAGE: VALUING ALL FAMILIES UNDER THE LAW* (2008); John G. Culhane, *Marriage Equality? First, Justify Marriage (If You Can)*, 1 *DREXEL L. REV.* 485, 511 (2009), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1428971 (asking “Which of marriages [sic] many benefits are justified, and to what extent? . . . How might we tie benefits and burdens to facts rather than to status?”); see also Melissa Murray, *Equal Rites and Equal Rights*, 96 *CAL. L. REV.* 1395 (2008) (discussing the California

have been another consequence of the same-sex marriage debate.⁷

American family law now has an energetic politics, which can only be a welcome development after years of widespread complacency towards an entrenched and encrusted mediocrity called “marriage.” It is no longer possible (nor has it ever been desirable) to wish an end to these family law debates, whether these debates occur within the gay and lesbian community, or without, and whether these debates concern same-sex marriage or its slippery-slope progeny.

Moreover, like the United States itself, American family law does not exist in a nationalistic bubble in a globalized world. Indeed, the American discussion of same-sex marriage has always been an especially rich one, and has also maintained vitality in the face of great odds, because of this discussion’s transnational character. Defying the commonplace image of family and family law being exclusively domestic concerns, the gay, lesbian, and bisexual movement⁸ for same-

Supreme Court’s opening the door to the possibility that the State of California may create a new type of officially-recognized relationship, equally available to all people, which is not called “marriage”).

⁷ I am referring here specifically to amendments to the U.S. Constitution proposed in both the House of Representatives and Senate in 2005-06 which would have constitutionalized an opposite-sex definition of marriage for the United States. *See, e.g.*, Marriage Protection Amendment, H.R.J. Res. 88, 109th Cong. (2006), Marriage Protection Amendment, S.J. Res. 1, 109th Cong. (2005) (both proposing constitutional language that “[m]arriage in the United States shall consist only of the union of a man and a woman.”).

⁸ This movement is also known as the “LGBT” (lesbian, gay, bisexual, and transgendered) rights movement. This Article uses the expression “gay, lesbian, and bisexual” (or “gay and lesbian” as an unfortunately useful shorthand) instead of the more-inclusive “LGBT” terminology, since many of the issues concerning same-sex marriage are only occasionally issues for transgendered people. In this respect, “[s]ame-sex marriages already exist in the transgender community.” Phyllis Randolph Frye & Alyson Dodi Meiselman, *Same-Sex Marriages Have Existed Legally in the United States for a Long Time Now*, 64 ALB. L. REV. 1031, 1036 (2000). Same-sex marriages can arise in the transgendered community as a result of post-marital changes in the natal sex of one partner in an opposite-sex marital relationship. *See generally* Jennifer L. Levi, *Marriage and Civil Unions*, in REPRESENTING NONTRADITIONAL FAMILIES (2006) (noting that there is a strong presumption of continuing legality of marriages in such situations since “[a]ll states abide by a strong public policy in favor of validating marriages, and an otherwise lawful marriage may only be terminated by death or divorce”). Same-sex marriages can also arise in the transgendered community when states refuse to legally recognize post-natal sex changes. As a consequence, “same-sex-appearing marriages,” Frye & Meiselman, *supra*, at 1033, can result when one person in an opposite-sex relationship transitions between sexes yet is still allowed to marry a partner of the “same” sex because the state refuses to legally recognize the sex change. This ironic result of the refusal of a state to permit/recognize post-natal changes in the legal sex of individuals can be found in Tennessee, Texas, Kansas, Florida, and Ohio. *See* Julie A. Greenberg, *When Is a Same-Sex Marriage Legal? Full Faith and Credit and Sex Determination*, 38

sex marriage has been an especially transnational one. For example, news of same-sex marriage developments in the Netherlands, South Africa, Spain, and elsewhere redounded quickly to the United States,⁹ and comparable developments in Massachusetts, California, and Connecticut now reverberate around the world.¹⁰ Yet, despite the transnational character of the same-sex marriage debates, rigorous transnational analyses with respect to some of the key concepts at play in today's (and tomorrow's) debates are few in number.¹¹

With this situation in mind, this Article's goals are two-fold and related, namely 1) to contribute to the radical re-thinking of family law that is on-going in the contemporary United States by 2) analyzing recent U.S. developments with respect to same-sex marriage from a transnational perspective. In doing so, this Article argues against the odd and overstated quality of recent American state court discussions concerning the *necessary* relationship between dignity and family law pluralism. These discussions, and the conclusions that they have given rise to, have resulted not only in the erasure¹² of

CREIGHTON L. REV. 289, 296-98 (2005). Finally, and more theoretically, transgendered individuals may object to the entire methodology concerning the duality or even knowability of "sex" that is often deployed when gay and lesbian activists advocate for "same-sex" marriage. See Mary Coombs, *Sexual Dis-Orientation: Transgendered People and Same-Sex Marriage*, 8 UCLA WOMEN'S L.J. 219, 220 (1997) (noting that "[b]oth the opponents and the proponents of same-sex marriage have generally assumed that 'same-sex marriage' is equivalent to 'gay or lesbian marriage'").

⁹ See, e.g., N.J. CIVIL UNION REVIEW COMMISSION, *THE LEGAL, MEDICAL, ECONOMIC & SOCIAL CONSEQUENCES OF NEW JERSEY'S CIVIL UNION LAW 37* (2008) (discussing international developments in same-sex relationship-recognition); Anthony Deutsch, *Dutch Lawmakers Make Marriage Legal for Gays*, CHI. SUN-TIMES, Sept. 12, 2000, at 3 (describing the enactment of a bill "converting the country's 'registered same-sex partnerships' into full-fledged marriages" and giving gay couples "divorce guidelines" as well as "wider adoption rights"); Chris Rovzar, *Spanish Lessons*, THE ADVOCATE, Jan. 16, 2007, available at http://www.advocate.com/issue_story_ektid41071.asp; Michelangelo Signorile, *Gay Marriage in U.S. Gets Boost*, NEWSDAY, Jun. 20, 2003 (describing recent same-sex marriage developments in Canada and their potential impact on the U.S.).

¹⁰ See, e.g., Sarah Beresford & Caroline Falkus, *Abolishing Marriage: Can Civil Partnership Cover It?*, 30 LIVERPOOL L. REV. 1, 3-5 (2009) (discussing U.S. same-sex marriage developments in the context of debates in the United Kingdom over the recognition of civil partnerships, as opposed to marriages, for same-sex couples); Tarunabh Khaitan, *Beyond Reasonableness: A Rigorous Standard of Review for Article 15 Infringement*, 50 J. INDIAN L. INST. 177, 180 n.13 (2008) (mentioning California same-sex marriage litigation while arguing for a heightened standard of review in the enforcement of Indian constitutional equality norms).

¹¹ POLIKOFF, *supra* note 6, at 110-22, provides a welcome exception to this general rule.

¹² For example, in both Connecticut and New Hampshire, after the legalization of same-sex marriages in those two states, civil unions were automatically converted into "marriages." See GAY & LESBIAN ADVOCATES & DEFENDERS, QUESTIONS

profound and enviable gay and lesbian legal achievements—“domestic partnerships,” “civil unions,” and the like—but also to a severe backlash in the form of Proposition 8 and similar state ballot initiatives.

The conventional (liberal) view is that Proposition 8 and similar laws which create “separate but equal” relationship-recognition regimes for homosexuals¹³ (as opposed to traditional heterosexual marriage) pose insurmountable affronts to gay and lesbian dignity. Using a transnational perspective and analysis, however, this Article proposes an alternative, more optimistic take on the relationship between dignity, same-sex marriage, and legal pluralism. Indeed, while the political campaign around Proposition 8 was heated and at times vitriolic, the ballot initiative ultimately returned California to a situation of family law pluralism, i.e. a situation where same-sex and opposite-sex couples are each governed by different (family) laws. In this instance, these different family laws grant essentially *the same* rights and responsibilities to each sort of couple. As this Article argues, however, there are other possible results from maintaining a separate system of family law for gays and lesbians, namely the possibility of gay and lesbian people exercising *agency* with respect to the family laws which directly affect them. With this agency, gays and lesbians would have the possibility of experiencing something more than “separate but equal” family laws being applied to them. Indeed, gays and lesbians would have the opportunity to *author*—or, in other words, to *exercise agency* with respect to—their own “separate *and better*” alternatives to (heterosexually-

AND ANSWERS ABOUT CONNECTICUT’S TRANSITION FROM CIVIL UNIONS TO MARRIAGE 4 (2009), <http://www.glad.org/uploads/docs/publications/ct-cu-to-marriage.pdf> (last visited Feb. 25, 2010); Andrew J. Manuse, *New Hampshire Legalizes Gay Marriage*, REUTERS, Jun. 3, 2009, available at <http://www.reuters.com/article/idUSTRE5526NV20090603>. In Vermont, after the legalization of same-sex marriage there in September, 2009, no new civil unions could be entered into, but existing civil unions were not automatically converted into marriages. See Vermont.com, Guide to Vermont Civil Marriage, <http://www.vermont.com/civilmarriagefaq.cfm> (last visited Feb. 25, 2010).

These examples suggest that William Eskridge’s optimistic prediction of the “sedimentary” effects of same-sex marriage laws has been somewhat disproved. See WILLIAM N. ESKRIDGE, JR., EQUALITY PRACTICE: CIVIL UNIONS AND THE FUTURE OF GAY RIGHTS 121 (2002) (“Each step toward same-sex marriage is typically (but not always) *sedimentary*: rather than displacing earlier reforms, the new reform simply adds another legal rule or institution on top of an earlier one.”); see also *id.* at 210, 218-19.

¹³ When this Article uses the term “homosexual,” it does so in the manner that one finds the term “heterosexual” being used, i.e. in a purely descriptive, generic, and non-judgmental manner.

authored) “majoritarian marriage.”¹⁴ These positive aspects to Proposition 8, and family law pluralism more generally, should not be overlooked, and this Article explores how they can be capitalized upon in a principled, dignity-oriented manner.

In refusing to be defeated by either the hate or the hopelessness that has infused the debate over Proposition 8 (and similar measures), this Article attempts to help the American gay and lesbian civil rights movement find a dignified way out of its current quagmire with (ostensibly) anti-gay forces,¹⁵ and the costly and counter-productive war over same-sex marriage. The traditional civil rights paradigms and strategies disparaging “separate but equal”¹⁶ laws that this

¹⁴ For a more detailed discussion of how I understand and use the term “agency” in this Article, see *infra* Part III.

¹⁵ Many people who are working to preserve “marriage” for heterosexuals only would contest the assertion that they are “anti-gay,” arguing rather that they are simply “pro-traditional marriage.” I am not convinced by a great number of these people, and I believe that a certain virulent homophobia underlies much of their opposition to same-sex marriage. In this, I am in partial accord with Martha Nussbaum’s diagnostic (and critical) observations concerning “traditional marriage” arguments that implicitly or explicitly assume that “to associate traditional marriage with the sex acts of same-sex couples is to defile or contaminate [traditional marriage].” Martha Nussbaum, *A Right to Marry? Same-sex Marriage and Constitutional Law*, DISSENT (Summer 2009), available at <http://www.dissentmagazine.org/article/?article=1935>. That being said, I also believe that many same-sex marriage advocates are less than homophobic, especially when they disparage activity that is typically ascribed to gay men (e.g. sexual promiscuity). For example, describing the factors that he thinks contributed historically to the gay and lesbian push for same-sex marriage rights, well-known professor and same-sex marriage advocate William Eskridge has written:

Whatever gravity gay life may have lacked in the disco seventies it acquired in the [AIDS] health crisis of the eighties. What it lost in youth and innocence it gained in dignity. Gay cruising and experimentation . . . gave way somewhat in the 1980s to a more lesbian-like interest in commitment. Since 1981 and probably earlier, gays were civilizing themselves. Part of our self-civilization has been an insistence on the right to marry.

WILLIAM N. ESKRIDGE, JR. THE CASE FOR SAME-SEX MARRIAGE 58 (1996); see also David B. Cruz, “*Just Don’t Call it Marriage: The First Amendment and Marriage as an Expressive Resource*,” 74 S. CAL. L. REV. 925, 943-44 (2001) (arguing that “[s]ame-sex couples, precluded by the mixed-sex requirement from using civil marriage to express the integrity of their sexuality, are . . . subjected to the ‘sex as lifestyle’ presumption”).

¹⁶ For previous examples of work that invokes arguments about the unconstitutionality of “separate but equal” family law institutions, using case law from previous civil rights struggles involving race and sex, see David Buckel, *Government Affixes a Label of Inferiority on Same-Sex Couples When It Imposes Civil Unions and Denies Access to Marriage*, 16 STAN. L. & POL’Y REV. 73, 74 (2005); Barbara Cox, *But Why Not Marriage: An Essay on Vermont’s Civil Unions Law, Same-Sex Marriage, and Separate but (Un)Equal*, 25 VT. L. REV. 113 (2000); Michael Mello, *For Today, I’m Gay: The Unfinished Battle for Same-Sex Marriage in Vermont*, 25 VT. L. REV. 149, 156 (2000); Mark Strasser, *Mission Impossible: On Baker, Equal Benefits, and the Imposition of Stigma*, 9 WM. & MARY BILL OF RTS. J. 1 (2000). All of these articles claim a parallel between the modern-day system of reserving “marriage” for heterosexuals, while granting “civil unions” or “domestic partnerships” to homosexuals, and the

movement heavily leans upon are not gaining widespread traction with respect to same-sex relationship recognition, even taking into account recent same-sex marriage developments in the District of Columbia, Iowa, and some Northeastern states.¹⁷ Moreover, even if these paradigms were to gain more widespread currency, there are real harms to gay and lesbian agency—and, as a result, dignity—that accompany gay and lesbian absorption into majoritarian family law, and these harms should not be overlooked.

Part I begins this Article with an exploration and excavation of two recent and important state supreme court judgments, from California and Connecticut, which exemplify the current state of mainstream liberal legal thinking with respect to the legalization of same-sex marriage. This Part focuses on how the crucial concept of “dignity” is deployed in these two legal decisions in support of the argument that gays and lesbians are denied dignity, and made second-class citizens, when the state recognizes dyadic, intimate, same-sex relationships differently than it does comparable opposite-sex relationships.¹⁸ According to both states’ supreme courts, any relationship-recognition system that grants heterosexuals the possibility of “marriage,” while only holding out “domestic partnerships” or “civil unions” to homosexuals, smacks of the now-repudiated idea that institutions can be “separate but equal.”

Part II contests the California and Connecticut Supreme Courts’ understanding of how dignity and legal uniformity must *necessarily* be connected. It does so by broadening the discussion of dignity and family law to look at both outside of the United States. While liberal advocates in the United States have argued that transnational and comparative experience is relevant and important with respect to some of the leading

nineteenth-century system of maintaining “separate but equal” railway cars for persons of different races, the constitutionality of which was upheld in the now widely-disparaged U.S. Supreme Court decision in *Plessy v. Ferguson*, 163 U.S. 537 (1896).

¹⁷ Marriage between same-sex partners is now legally available in five states: Connecticut, Iowa, Massachusetts, New Hampshire and Vermont. Lambda Legal, Status of Same-Sex Relationships Nationwide, <http://www.lambdalegal.org/publications/articles/nationwide-status-same-sex-relationships.html> (last visited Feb. 22, 2010). Same-sex marriage has also very recently become available in the District of Columbia.

¹⁸ This Article uses the conventional expression “opposite-sex relationships” to describe male-female pairings, but does not intend to endorse the view that there are only two sexes or that, even if there are, that they have to be configured as dyadic and “opposite.”

legal issues of the day,¹⁹ the ostensibly liberal California and Connecticut Supreme Court decisions are astonishing in their overwhelmingly domestic focus. Part II compensates for this lack by showing what a more rigorous transnational investigation centered on dignity and family law pluralism would unearth. The foreign jurisdictions examined in this Part include Canada, the United Kingdom, and India. As this Part discusses, in these national contexts, dignity and its conceptual cognates (e.g. respect, tolerance, minority rights) have been invoked not to amalgamate minorities into a unitary, common family law system but, instead, to provide minorities with legal space in which to implement non-majoritarian visions of family, community, and the good life.

Part III brings the discussion back home, showing how a domestic consideration of transnational notions of dignity and family law pluralism could play out in the United States. Provocatively, this Part argues that the dignity of gay and lesbian people could be *enhanced* by a separate system of relationship-recognition and family law for same-sex unions. Such a separate system would create legislative space and freedom for the exercise of gay and lesbian *agency*, and the elaboration of “separate *and better*” alternatives to the straitjacket of majoritarian marriage. However, as this Part also discusses, in order for this potential to be realized, there will have to be transformations in the imagination and aims of the current gay and lesbian rights movement, as well as in the larger social and legal context in which this movement is situated.

This Article thus ends by confronting squarely but confidently the reality of a twenty-first century United States—one where same-sex marriage has little traction or instantiation, and one where conservatives’ success at colonizing family law more generally cannot be elided. Indeed, instead of perpetually lamenting this reality, this Part concludes by working to creatively generate new legal thinking²⁰ which de-links the dignity of gay and lesbian people

¹⁹ See, e.g., Vicki C. Jackson, *Comparative Constitutional Federalism and Transnational Judicial Discourse*, 2 INT’L J. CONST. L. 91, 92-93 (2004); WILLIAM N. ESKRIDGE, JR., *EQUALITY PRACTICE: CIVIL UNIONS AND THE FUTURE OF GAY RIGHTS* 83 (2002).

²⁰ Of course, knowledge is (often) cumulative, and my arguments here clearly build off of a great deal of previous important work in queer theory, political theory, and even linguistics. For previous examples of work that has made similar—yet also quite different—points with respect to some of the arguments presented in this Article,

with majoritarian marriage and, instead, locates this dignity in the agency of gay and lesbian people with respect to their own lives, their own families, and their own laws. Indeed, one way of both inhabiting and expressing this agency, and dignity, would be to assert political and legislative control over a separate body of family law for gay and lesbian people and families. Such a move would not be motivated by compromise or capitulation,²¹ or utopian thinking, but by a deeply principled quest for dignity in a contemporary United States that has demonstrated its eager readiness to permit gays and lesbians to occupy a different legal arena than heterosexuals. With this

see generally Shahar Lifshitz, *Married Against Their Will?: Toward a Pluralist Regulation of Spousal Relationships*, 66 WASH. & LEE L. REV. 1565, 1573 (2009) (distinguishing opposite-sex and same-sex couples and “suggest[ing] a unique legal regime for the latter,” but primarily as an unfortunately necessary result of the fact that same-sex couples face “legal restrictions from getting married”); POLIKOFF, *supra* note 6; Douglas W. Allen, *An Economic Assessment of Same-Sex Marriage Laws*, 29 HARV. J.L. & PUB. POL’Y 949, 980 (2006) (advocating but only briefly developing “a separate legal structure called ‘homosexual marriage,’” and doing so from a hetero-centric perspective which valorizes “traditional marriage”); Marie A. Failinger, *A Peace Proposal for the Same-Sex Marriage Wars: Restoring the Household to Its Proper Place*, 10 WM. & MARY J. WOMEN & L. 195, 198 (2004) (characterizing same-sex marriage advocacy as “ultimately mimic[ing] rather than resolv[ing] the problems with using the ‘choice’-based nuclear family as the favored legal model for ordering intimate relationships”); ESKRIDGE, *supra* note 12 (evinced interest in pluralistic “tailor-made regulatory regimes” for families but repeatedly characterizing any non-marital regime for same-sex partners as of a “compromise” nature); Barbara Stark, *Marriage Proposals: From One-Size-Fits-All to Postmodern Marriage Law*, 89 CAL. L. REV. 1479, 1490-91 (2001) (diagnosing and expressing skepticism towards “metanarratives” about marriage); Paula L. Ettelbrick, *Avoiding a Collision Course in Lesbian and Gay Family Advocacy*, 17 N.Y.L. SCH. J. HUM. RTS. 753, 758 (2000) (proposing a “continuum of family recognition options,” all of which would be open to both homosexuals and heterosexuals on the basis of formal equality); MICHAEL WARNER, *THE TROUBLE WITH NORMAL: SEX, POLITICS, AND THE ETHICS OF QUEER LIFE* (1999) (providing perhaps the most rousing and wide-ranging queer critique of same-sex marriage advocacy that has been made in the past many years).

²¹ While I am sympathetic to the proposals recently put forward by David Blankenhorn and Jonathan Rauch with respect to the legislation of a federal civil union regime—as distinguished from marriage—I would resist their characterization of this as a “compromise.” See David Blankenhorn & Jonathan Rauch, *A Reconciliation on Gay Marriage*, N.Y. TIMES, Feb. 22, 2009; see also William N. Eskridge, Jr., *How Government Unintentionally Influences Culture (The Case of Same-Sex Marriage)*, 102 NW. U. L. REV. 495, 496 (2008) (identifying domestic partnerships as a “compromise” between same-sex marriage advocates and opponents); Nussbaum, *supra* note 15 (identifying civil unions as a “compromise offer”). For more discussion on how principle can provide the foundation for belief in legal pluralism, see Martha Minow, *Is Pluralism an Ideal or a Compromise?: An Essay for Carol Weisbrod*, 40 CONN. L. REV. 1287 (2007). For another articulation of the relationship between legal pluralism and higher ideals, see Katharine Bartlett’s argument that “in reducing the power of individuals to make their own family decisions, family-standardizing reform reduces the capacity of individuals to develop as *moral beings*.” Katharine T. Bartlett, *Saving the Family from the Reformers*, 31 U.C. DAVIS L. REV. 809, 817 (1998) (emphasis added).

in mind, this Article aims to imagine how gay and lesbian dignity might be enhanced rather than diminished by looking broadly, traveling widely, and viewing the world with curiosity and xenophilia, rather than dread and homophobia.

I. CALIFORNIA AND CONNECTICUT

Few expressions call forth the nod of assent and put an end to analysis as readily as “the dignity of man.”²²

—Bertram Morris (1946)

This Part explains and explores two recent and important state supreme court judgments, from California and Connecticut, which exemplify the current state of mainstream liberal legal thinking with respect to the legalization of same-sex marriage. This Part concentrates on these two state high court judgments because they are the most recent state supreme court judgments that explicitly invoke the concept of dignity in their resolution of the question presented in each case. By way of comparison, the recent Iowa Supreme Court judgment legalizing same-sex marriage in that state did not use the word “dignity” even once in its judgment.²³ Prior to the California and Connecticut high court decisions, the Supreme Judicial Court of Massachusetts had issued an advisory opinion in 2004 to that state’s Senate on a question very similar to the one that both the California and Connecticut courts addressed in their opinions, namely the constitutionality of a state government naming officially-recognized, otherwise-equivalent *same-sex* relationships something different than “marriage.”²⁴ However, I do not discuss this opinion in detail in this Part because so much of the analysis in that opinion is relied upon and utilized by the California and Connecticut Supreme Courts.²⁵ In the dual interests of brevity and currency, this Part

²² Bertram Morris, *The Dignity of Man*, 57 ETHICS 57, 57 (1946).

²³ See *Varnum v. Brien*, 763 N.W.2d 862 (Iowa 2009).

²⁴ See *Opinions of the Justices to the Senate*, 802 N.E.2d 565, 569 (Mass. 2004). Earlier, of course, the Massachusetts Supreme Judicial Court had issued its path-breaking opinion, *Goodridge v. Dept. of Pub. Health*, 798 N.E.2d 941 (Mass. 2003), legalizing same-sex marriage in the first place. The concept of dignity played a role in this opinion as well, with the court declaring that “[t]he Massachusetts Constitution affirms the dignity and equality of all individuals. It forbids the creation of second-class citizens.” *Id.* at 948.

²⁵ See, e.g., *In re Marriage Cases*, 183 P.3d 384, 398 n.3 (Cal. 2008); *Kerrigan v. Comm’r of Pub. Health*, 957 A.2d 407, 417 (Conn. 2008).

focuses on these two most recent state supreme court opinions instead.

In the spring of 2008, the California Supreme Court handed down its groundbreaking decision concerning same-sex marriage, *In re Marriage Cases*.²⁶ In this case, the court was asked to decide whether California's relationship-recognition system was consistent with the California state constitution's protections of the right to marry and the right to equality.²⁷ Under this relationship-recognition system, "marriage" was reserved for opposite-sex couples, while same-sex couples had access only to a parallel "domestic partnership" regime.²⁸ Like California, some other states had also created two parallel systems of family law within their borders,²⁹ but California's regime of separate laws for different sexual orientations was unusual in that it accorded domestic partners "virtually all of the same substantive legal benefits and privileges, and . . . legal obligations and duties . . . that California law affords to and imposes upon a married couple."³⁰ Accordingly, what the California Supreme Court had to decide in this case was whether California's "separate but equal"³¹ family law system was constitutional under the California Constitution. Ultimately, the court held that this system was not constitutional, and that same-sex couples had to be given "marriage" licenses just like opposite-sex couples.³²

²⁶ *In re Marriage Cases*, 183 P.3d 384, *superseded by* CAL. CONST. art. I, § 7.5.

²⁷ *See id.* at 400.

²⁸ *See id.* at 409, 413.

²⁹ For example, Hawaii has enacted a law concerning "reciprocal beneficiaries" and Wisconsin has adopted a form of "domestic partnership," but neither scheme provides the same rights and obligations as "marriage," or California's expansive, marriage-like "domestic partnership" regime. *See* Lambda Legal, Status of Same-Sex Relationships Nationwide, <http://www.lambdalegal.org/publications/articles/nationwide-status-same-sex-relationships.html> (last visited Jan. 15, 2010).

³⁰ *In re Marriage Cases*, 183 P.3d at 398. According to the court, nine differences remain between domestic partnerships and marriages in California. *Id.* at 416-17 n.24. Some of these differences are arguably to the benefit of people entering into a domestic partnership, while others arguably impose burdens that people entering marriage do not face. An example of an advantage would be that domestic partnerships are easier to dissolve than marriages in California. An example of a burden placed solely on people wishing to enter a domestic partnership is the requirement that such people have a common residence. There is no such common-residence requirement for people marrying. *See id.*

³¹ The court explicitly links California's system of maintaining a "separate institution of domestic partnership," *id.* at 445 (emphasis added), with the (ostensibly) historic practice of "relegat[ing] . . . racial minorities to separate and assertedly equivalent public facilities and institutions," *id.* at 451 (emphasis added).

³² The court holds that California's system was unconstitutional on both a "fundamental right to marry" and equal protection grounds. *See id.* at 419, 433-34, 452.

There are many groundbreaking and interesting aspects to this decision. For example, this decision represented the first instance of a state's highest court applying a "strict scrutiny" standard to discrimination against gays and lesbians.³³ The decision was also noteworthy in its contemplation of the possibility that the State of California might create a relationship regime—available to everyone—that would use a rubric other than "marriage."³⁴ Finally, and without any sense of irony, the court seemed to agree with the same-sex marriage advocates litigating this case that there existed a fundamental "right to remain in the closet" in the State of California.³⁵

As important as all of the above features of the California decision are, this Part concentrates on an aspect of the court's decision that has remained under-examined in the academic literature, namely the court's discussion of the

³³ See *id.* at 441-42; see also Kenji Yoshino, *Magisterial Conviction: Why the California Supreme Court Did More than Legalize Gay Marriage*, SLATE, May 15, 2008, <http://www.slate.com/id/2191530/> (discussing uniqueness of California Supreme Court opinion with respect to applying strict scrutiny standard to sexual orientation discrimination).

³⁴ Wrote the court:

When a statute's differential treatment of separate categories of individuals is found to violate equal protection principles, a court must determine whether the constitutional violation should be eliminated or cured by extending to the previously excluded class the treatment or benefit that the statute affords to the included class, or alternatively should be remedied by withholding the benefit equally from both the previously included class and the excluded class. A court generally makes that determination by considering whether extending the benefit equally to both classes, or instead withholding it equally, would be most consistent with the likely intent of the Legislature, had that body recognized that unequal treatment was constitutionally impermissible.

. . . [T]here can be no doubt that extending the designation of marriage to same-sex couples, rather than denying it to all couples, is the equal protection remedy that is most consistent with our state's general legislative policy and preference.

In re Marriage Cases, 183 P.3d at 452-53; see also Melissa Murray, Remark, *Equal Rites and Equal Rights*, 96 CAL. L. REV. 1395 (2008) (discussing the California Supreme Court's opening the door to the possibility that the State of California may create a new type of officially-recognized relationship, equally available to all people, which is not called "marriage").

³⁵ The nomenclature for this right is mine, and it is a reaction to the court's sympathy for the plaintiffs' argument that "one consequence of the coexistence of two parallel types of familial relationships is that—in the numerous everyday . . . settings in which an individual is asked whether he or she 'is married or single'—an individual who is a domestic partner and who accurately responds to the question by disclosing that status will . . . be disclosing his or her homosexual orientation." *In re Marriage Cases*, 183 P.3d at 446. The court links this allegedly coercive disclosure to the fundamental right to privacy that is contained within California's state constitution. See *id.*

concept of “dignity” and its relationship to pluralistic family law systems. The court’s words on the subject of how dignity relates to family law pluralism are worth quoting at length:

One of the core elements of the right to establish an officially recognized family that is embodied in the California constitutional right to marry is a couple’s right to have their family relationship accorded dignity and respect equal to that accorded other officially recognized families, and assigning a different designation for the family relationship of same-sex couples while reserving the historic designation of “marriage” exclusively for opposite-sex couples poses at least a serious risk of denying the family relationship of same-sex couples such equal dignity and respect.

. . . .

. . . [R]etaining the designation of marriage exclusively for opposite-sex couples and providing only a separate and distinct designation for same-sex couples may well have the effect of perpetuating a more general premise—now emphatically rejected by this state—that gay individuals and same-sex couples are in some respects “second-class citizens” who may, under the law, be treated differently from, and less favorably than, heterosexual individuals or opposite-sex couples.³⁶

Like other parts of the court’s opinion, the court’s discussion of dignity here was groundbreaking, but perhaps in an unanticipated way. For many people outside of the United States (especially), the court’s equation of dignity and family law uniformity *is* revolutionary, but mainly because it seems so ahistorical and ungrounded in real-world experience. Part III will discuss these global family law experiences in more detail, and what they can tell us about the complicated relationship between dignity and family law pluralism.

That being said, the reality of family law around the globe did not completely escape the court’s attention in its opinion. For example, when discussing the California Attorney General’s arguments pertaining to the historical definition of marriage,³⁷ the court did observe that “until recently, there has been widespread societal disapproval and disparagement of

³⁶ *Id.* at 400, 402.

³⁷ Noted the court:

The Attorney General and the Governor maintain . . . that because the institution of marriage traditionally (both in California and throughout most of the world) has been limited to a union between a man and a woman, any change in that status necessarily is a matter solely for the legislative process.

Id. at 447-48.

homosexuality in many cultures” and that, as a result, the designation of marriage continues to apply only to a relationship between opposite-sex couples in the overwhelming majority of jurisdictions in the United States, and around the world.³⁸ Furthermore, the court ably made use of a Canadian Supreme Court opinion when describing how the history of discrimination against gay people cautions against thinking that any separate and parallel family law system for them can be anything but discriminatory.³⁹ Yet, as the next Part discusses, the court’s global vision in its decision was extremely partial, avoiding not only a deeper exploration of Canadian family law realities and debates, but similar ones pertaining to family law pluralism, dignity, and minority rights elsewhere.

Less than six months after the California Supreme Court’s decision, the Connecticut Supreme Court followed with its own path-breaking opinion on same-sex marriage. In *Kerrigan v. Commissioner of Public Health*,⁴⁰ the Connecticut Supreme Court decided whether—in the court’s own words—Connecticut’s practice of “segregat[ing] heterosexual and homosexual couples into [the] separate institutions” of (respectively) “marriage” and “civil union” violated the Connecticut Constitution’s protections as to substantive due process and equality.⁴¹ Similar to California’s system of parallel

³⁸ *Id.* at 451 n.70.

³⁹ Noted the court:

[P]articularly in light of the historic disparagement of and discrimination against gay persons, there is a very significant risk that retaining a distinction in nomenclature with regard to this most fundamental of relationships whereby the term “marriage” is denied only to same-sex couples inevitably will cause the new parallel institution that has been made available to those couples to be viewed as of a lesser stature than marriage and, in effect, as a mark of second-class citizenship. As the Canada Supreme Court observed in an analogous context: “One factor which may demonstrate that legislation that treats the claimant differently has the effect of demeaning the claimant’s dignity is the existence of pre-existing disadvantage, stereotyping, prejudice, or vulnerability experienced by the individual or group at issue It is logical to conclude that, in most cases, further differential treatment will contribute to the perpetuation or promotion of their unfair social characterization, and will have a more severe impact upon them, since they are already vulnerable.”

Id. at 445 (second and third alterations in original) (quoting *M. v. H.*, [1999] 2 S.C.R. 3, 54-55 [¶ 68] (Can.)).

⁴⁰ See generally *Kerrigan v. Comm’r of Pub. Health*, 957 A.2d 407 (Conn. 2008).

⁴¹ *Id.* at 412. The Connecticut Supreme Court understands the Connecticut Constitution’s due process guarantee to incorporate the “the fundamental right to marry the person of [one’s] choice.” *Id.* at 413.

relationship-recognition, Connecticut's civil union scheme "conferred on [civil] unions all the rights and privileges that are granted to spouses in a marriage."⁴²

As with the California opinion which shortly preceded it, there were many interesting aspects to the Connecticut opinion. For example, like the opinion from California, the Connecticut Supreme Court (following plaintiffs' example)⁴³ used language evocative of the struggle for African-American civil rights when characterizing the parallel relationship recognition regime in Connecticut as involving "segregation."⁴⁴ In addition, like the California court, the Connecticut court also decided to apply a heightened level of scrutiny to sexual orientation classifications contained in law. In this respect, the court found that any sexual orientation classifications that a law may use are "quasi-suspect" and deserve "intermediate scrutiny,"⁴⁵ i.e. more scrutiny than "rational basis" review but less than the "strict scrutiny" that the California Supreme

⁴² *Id.*

⁴³ *Id.* at 413.

⁴⁴ See *supra* text accompanying note 40. The plaintiffs also had a slightly different argument, based on *sex*-segregation. As the court characterized their claims:

[T]he plaintiffs maintained that, by limiting marriage to the union of a man and a woman, [the Connecticut] statutory scheme impermissibly segregates on the basis of *sex*. . . . The plaintiffs contended that [Connecticut's] statutes contravene the state constitutional prohibition against sex discrimination because these statutes preclude a woman from doing what a man may do, namely, marry a woman, and preclude a man from doing what a woman may do, namely, marry a man.

Kerrigan, 957 A.2d at 414 (emphasis added).

⁴⁵ See *Kerrigan*, 957 A.2d at 412. There exists quite a bit of irony in the constitutional methodology that the court deploys to justify its use of an intermediate level of scrutiny here. In this respect, while the court wields an age-old, unchanging, and overly-valorized institution of marriage throughout much of its opinion, the court holds much more flexible ideas about a constitution and its changing contours. For the court, when interpreting a constitution (such as Connecticut's), it is important to interpret it "in accordance with the demands of modern society" such that it will not remain "static [and] incapable of coping with changing times." *Id.* at 420-21 (quoting *McCulloch v. Maryland*, 17 U.S. (4 Wheat.) 316, 415 (1819); *State v. Dukes*, 457 A.2d 10, 19 (Conn. 1988)). The court needed to take this flexible approach to constitutionalism because the Connecticut state constitution does not explicitly forbid sexual orientation discrimination yet, with its opinion, the court intended to extend intermediate scrutiny to legislation specifically distinguishing gay and lesbian people. See *Kerrigan*, 957 A.2d at 425. As a result, however, marriage becomes more of a bedrock, foundational institution in Connecticut than even the Connecticut constitution itself. In this way, the Connecticut opinion is somewhat different than state court opinions that have used the "post-legal" (as opposed to "pre-legal") status of marriage in the process of arguing against the inclusion of same-sex couples within "marriage." For examples and discussion of such state court opinions, see generally *Abrams & Brooks*, *supra* note 4, at 20-28.

Court decided to exercise in relation to sexual orientation discrimination.

However, the strongest parallels between the Connecticut Supreme Court and the California Supreme Court's opinion were perhaps those found in the Connecticut court's holding that the denial of "real"⁴⁶ marriage to same-sex couples implicated the dignity interests of these couples, and also homosexual individuals more generally.⁴⁷ Indeed, the parallels could hardly be stronger, given that the Connecticut court largely relied on cutting-and-pasting from the California decision (and prior ones from Massachusetts) in the portions of its opinion dealing with the dignity question.⁴⁸

When speaking for itself on the dignity question, the Connecticut high court found that the basic equality that Connecticut had legislated between marriage and civil unions was constitutionally defective because these different institutions did not operate in a historical vacuum. According to the court, "[a]lthough marriage and civil unions do embody the same legal rights under our law, they are by no means 'equal.' . . . [T]he former is an institution of transcendent historical, cultural and social significance, whereas the latter most surely is not."⁴⁹

With respect to this asserted significance for marriage, and echoing plaintiffs' claim that marriage—more so than civil unions—is "special,"⁵⁰ the Connecticut Supreme Court's opinion explained in detail the unique and vital role that it believed marriage plays in the contemporary American polity. To do so, the opinion again relied heavily on quotations and citations

⁴⁶ *Kerrigan*, 957 A.2d at 417.

⁴⁷ *See id.* at 417-18, 465-74. It should also be noted that the Connecticut Supreme Court is also worried about how the withholding of "marriage" from same-sex couples affects the well-being of any children such couples have. The court wrote:

[T]he ban on same sex marriage is likely to have an especially deleterious effect on the children of same sex couples. A primary reason why many same sex couples wish to marry is so that their children can feel secure in knowing that their parents' relationships are as valid and as valued as the marital relationships of their friends' parents.

Id. at 474. For more on this harm to children, see also *id.* at 475 n.77.

⁴⁸ *See id.* at 417-18, 471-75 for the Connecticut Supreme Court's use of lengthy quotations from *In re Marriage Cases*, 183 P.3d 384 (Cal. 2008), *superseded by* CAL. CONST. art. I, § 7.5; *Opinions of the Justices to the Senate*, 802 N.E.2d 565 (Mass. 2004); and *Goodridge v. Dept. of Pub. Health*, 798 N.E.2d 941 (Mass. 2003).

⁴⁹ *Kerrigan*, 957 A.2d at 418.

⁵⁰ *Id.* at 416 ("[Plaintiffs] contend that [marriage] is an institution of unique and enduring importance in our society, one that carries with it a special status.").

from other U.S. courts. Following these other courts' lead, then, the Connecticut court alternatively characterized marriage as "fundamental to our very existence and survival,"⁵¹ "intimate to the degree of being sacred,"⁵² and, citing a more ancient yet less hyperbolic precedent, "one of the most fundamental of human relationships."⁵³

As a result of this remarkably (and perhaps uniquely) esteemed institutional history (for marriage), the withholding of the "marriage" nomenclature from same-sex couplings became acutely problematic for the court, especially given the fact that "historically [gays and lesbians have] been the object of scorn, intolerance, ridicule or worse."⁵⁴ Indeed, as a consequence of this historic stigmatization, the court believed that the separate legislation of civil unions could *only* be popularly perceived as "an official state policy that [civil unions are] inferior to marriage, and that the committed relationships of same sex couples are of a lesser stature than comparable relationships of opposite sex couples."⁵⁵

Given these concerns, it should come as no surprise that, using its intermediate level of scrutiny, the Connecticut Supreme Court ultimately determined that Connecticut's relationship-recognition scheme violated the Connecticut Constitution's equality protections. In doing so, the court stressed the "overriding similarities" between opposite-sex and same-sex couples,⁵⁶ with gay and lesbian people "shar[ing] the same interest in a committed and loving relationship as heterosexual persons who wish to marry, and . . . shar[ing] the same interest in having a family and raising their children in a loving and supportive environment."⁵⁷ Given this asserted⁵⁸ fundamental equivalence between same-sex and opposite-sex couples, it became inescapable that the court would declare that "firmly established equal protection principles lead[]

⁵¹ *Id.* (quoting *Loving v. Virginia*, 388 U.S. 1, 12 (1967)).

⁵² *Id.* (quoting *Griswold v. Connecticut*, 381 U.S. 479, 486 (1965)).

⁵³ *Id.* at 417 (quoting *Davis v. Davis*, 175 A. 574, 577 (Conn. 1934)).

⁵⁴ *Id.* at 418.

⁵⁵ *Id.* at 475.

⁵⁶ *Id.* at 424.

⁵⁷ *Id.*

⁵⁸ I use this word to indicate the unempirical nature of the court's findings in this respect. For evidence of significant differences between same-sex and opposite-sex marriages, see Scott James, *Many Successful Gay Marriages Share an Open Secret*, N.Y. TIMES, Jan. 29, 2010, at A17 (discussing widespread prevalence of non-monogamous marriages within the gay and lesbian community), available at <http://www.nytimes.com/2010/01/29/us/29sfmetro.html>.

inevitably to the conclusion that gay persons are entitled to marry the otherwise qualified same sex partner of their choice. . . . [S]ame sex couples cannot be denied the freedom to marry.⁵⁹

The Connecticut Supreme Court's opinion thus reached the same basic conclusion as that of the California Supreme Court, while also using some of the same tools that the California court used (e.g. heightened scrutiny, the segregation metaphor). One remarkable difference between the two decisions, however, was that the Connecticut opinion never discussed non-U.S. legal or political experience. As discussed above, the California opinion did discuss and utilize such experience, but in an ungrounded and distorted manner. The next Part engages in a different reading of transnational legal experience with respect to the issue of how dignity and family law pluralism can relate to each other.

II. DIGNITY AND FAMILY LAW PLURALISM, TRANSNATIONALLY-SPEAKING

Q: If you got married [in the United Kingdom], would you have a civil marriage as well as a *nikah* [Muslim religious marriage]?

A. *I would have a civil marriage; I don't know if it is more sort of a tradition thing that happens now; you have your nikah, and then you have your civil marriage as well.*

Q: Is there any other reason apart from the fact that it is what everyone else does?

A: No.

Q: Can you think of any reasons why you would want a civil marriage?

A: No.⁶⁰

—Interview by Sonia Nurin Shah-Kazemi
of a young Muslim woman in the United Kingdom (2001)

A. *Introduction*

California and Connecticut (and Massachusetts before them) clearly see family law pluralism—in particular,

⁵⁹ *Kerrigan*, 957 A.2d at 482.

⁶⁰ SONIA NURIN SHAH-KAZEMI, UNTYING THE KNOT: MUSLIM WOMEN, DIVORCE, AND THE SHARIAH 33 (2001).

pluralism with respect to the law of relationship-recognition—as implicating dignity concerns. For both state high courts, state-recognized “marriage” is the path to dignity, and gay and lesbian people are necessarily forced into second-class citizenship if the majority’s family law conventions are not opened up to them.

The choice by both the California and Connecticut Supreme Courts to invoke the language of dignity is important because, in the contemporary world, dignity is readily associated with the discourse of *human* rights. This is not to say that dignity has not featured in American constitutional discourse concerning *civil* rights—it most certainly has⁶¹—but it is to say that, in today’s world, “dignity” is more easily conjoined with “human” than it is with any particular subspecies of humanity.⁶² In other words, one speaks more easily of “*human* dignity” than one does “American dignity” or “European dignity” or “Indian dignity.”⁶³ Moreover, to say that

⁶¹ See, e.g., *Trop v. Dulles*, 356 U.S. 86, 100-01 (1958) (stating that the “dignity of man” underlies the Eighth Amendment and protects individuals from punishments that exceed current “civilized standards”). Dignity has also featured in American jurisprudential discussions of federalism. See, e.g., *Fed. Mar. Comm’n v. S.C. State Ports Auth.*, 535 U.S. 743 (2002). In this case, concerning the constitutionality of a U.S. government agency’s administrative hearing of a complaint by a private company against a South Carolina government agency’s decision, the U.S. Supreme Court wrote: “The *preeminent* purpose of state sovereign immunity is to accord States the *dignity* that is consistent with their status as sovereign entities.” *Id.* at 760 (emphasis added). This was not the first time that the Court recognized dignitary interests in upholding (a certain view of) states’ sovereignty rights. See, e.g., *Alden v. Maine*, 527 U.S. 706, 715 (1999); *Ex parte Ayers*, 123 U.S. 443, 505 (1887). However, it is one of the most recent and strongest statements as to those interests in the modern period. For an overview of how dignity has been deployed in U.S. Supreme Court jurisprudence, see generally Maxine D. Goodman, *Human Dignity in Supreme Court Constitutional Jurisprudence*, 84 NEB. L. REV. 740 (2005).

⁶² Some may disagree, but it is striking to note the regular invocation of the larger expression “human dignity” in any number of articles, rather than the simpler term “dignity.” For example, Maxine Goodman’s article, *supra* note 61, is entitled “Human Dignity in Supreme Court Constitutional Jurisprudence” when it might (conceivably) have been entitled simply “Dignity in Supreme Court Constitutional Jurisprudence.” Interestingly, even Neomi Rao finds the use of the term “human dignity” seemingly inescapable, even while trying to parochialize the concept. For example, she writes: “Perhaps we should direct our attention to developing an *American conception of human dignity* based on the Constitution as well as on our legal traditions.” See Neomi Rao, *On the Use and Abuse of Dignity in Constitutional Law*, 14 COLUM. J. EUR. L. 201, 255 (2008) (emphasis added) (proposing and arguing for an American legal definition of dignity that would differ from prevailing notions of dignity commonly deployed in European legal argumentation).

⁶³ *But see* Rao, *supra* note 62; James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1161 (2004) (finding a European conception of “personal dignity” which is tightly linked to the somewhat peculiar European ideas that one has “rights to one’s image, name, and reputation”) (emphasis omitted).

“dignity” is compromised by a particular law or legal framework suggests that one’s analysis in this respect can and should be extended to all of humanity.⁶⁴

As this Part demonstrates, however, it is difficult to claim that “separate but equal” family law regimes *necessarily* implicate the dignity interests of minorities or “second-class citizens”⁶⁵ if one looks globally at all of humanity. In particular, countries that implement family law via “personal law”⁶⁶ often do so either to affirmatively pursue multiculturalist legal policies, or do so in response to concerns (and resistance) from minorities about efforts to coerce them into majoritarian understandings of family, community, and the good life. Additionally, even in countries that legislate and enforce family law in a manner resembling more closely American-style family law, increasingly there are efforts to allow (religious) minorities to pursue alternative visions of family via non-state arbitration of family law matters. This Part will look at both kinds of countries, broadening the discussion of dignity and legal pluralism to take account of legal realities and developments in places as diverse as Canada, the United Kingdom, and India.

This Part’s selection of countries from which one can learn more about dignity and family law pluralism benefits from being diverse and broad-based, instead of narrow and unrepresentative of the world’s different cultural and legal traditions. The selection of Canada, the United Kingdom, and India as case-studies is also beneficial because, like the United States, each country is (proudly) a multi-ethnic, multi-religious democracy where debates over minority rights and cultural rights are common and longstanding. In other words, each of these three countries has a great deal of experience with “the dignity question,” and each of these countries has struggled with the reality of a diverse population that does not possess any single notion of “the good life.” Indeed, compared to these three countries, the United States is somewhat of a latecomer to discussions concerning dignity and family law pluralism.⁶⁷

⁶⁴ See generally Menachem Mautner, *From “Honor” to “Dignity”: How Should a Liberal State Treat Non-Liberal Cultural Groups?*, 9 THEORETICAL INQUIRIES LAW 609, 626 (2008) (discussing link between universal human rights and human dignity).

⁶⁵ *In re Marriage Cases*, 183 P.3d 384, 402 (Cal. 2008), *superseded by* CAL. CONST. art. I, § 7.5.

⁶⁶ See discussion *infra* Part II.C.

⁶⁷ I say “somewhat” here keeping in mind that it was American-style federalism itself that created the opportunity for both California and Connecticut to

It should be emphasized from the outset that the claim of this Part is *not* that the dignity argument with respect to American gays and lesbians is wrong per se.⁶⁸ The point, instead, is to highlight the fact that the dignity claim is a far-more-complicated one than it is typically made out to be by American lawyers and judges. This being the case, it is hoped that by stripping the dignity claim of its veneer of obviousness, it will be possible to see why the claim very much *might be* incorrect. Essentially, after peeling away some of the self-righteous rhetoric that provides both swords and shields for all sides in the same-sex marriage debates, this Part hopes to show how the dignity claim does not (conceptually or experientially) necessarily win the battle for same-sex marriage advocates. However, it does not lose it for them either. Finally, it should also be noted that many people feel that supporters of Proposition 8 acted in quite an undignified manner in the advertising campaign leading up to the California vote.⁶⁹ Both sides of the debate have their difficulties with dignity.

develop family law systems different than that found in New York, namely systems in which “marriage” and its homosexual sidekicks (i.e. “domestic partnerships” and “civil unions”) *differed minimally* in economic and legal benefits. There is a common inability, however, in American discussions of family law pluralism to conceive of this pluralism at a level different than that of the 50 states.

⁶⁸ Clearly, the fact that religious minorities around the world are not using dignity claims to argue for their amalgamation into majoritarian marital and family law does not necessarily preclude gays and lesbians in the United States from—correctly—doing so. There are real differences between other countries’ religious minorities and America’s sexual minorities, and also the histories of the family law systems that govern in each country. For one, many religions have had family law traditions that predate secular states and secular norms by centuries. Gays and lesbians, on the other hand, have often been excluded or excommunicated from the family altogether. It would not be surprising if each kind of community or cultural grouping sees different things in the family, and needs different things to feel “whole” or dignified. That being said, it would be a mistake to believe that American sexual minorities, to the extent that they do feel socially excluded, all necessarily view “marriage” as the antidote for that feeling of exclusion. It is also another question altogether whether such an antidote is necessarily the proper one for the future (as opposed to now). In this respect, it is my hope that the non-American family law examples discussed in this Article will incite a great deal of future exploration of dignified alternatives to majoritarian marriage. These possible alternatives are presently being ignored by mainstream actors in the American same-sex marriage debates.

For more on the cultural and legal obstacles that American gays and lesbians face with respect to imagining themselves like a (religious) “community” or “culture” with attendant legal rights and privileges, see *infra* Part III.

⁶⁹ See, e.g., Andy Birkey, *Kersten’s ‘Bullying Tactics’ Unhelpful to Gay Marriage Debate*, STAR TRIB. (Minneapolis), Jan. 26, 2010, available at <http://www.startribune.com/yourvoices/82763402.html?elr=KARKSUUODEY3LGDI07>; E.J. Schultz, *Prop. 8 TV Ad Raises Questions: Controversy Swirls Around the Teaching of Gay Marriage in*

This Part begins with a discussion of how the question of dignity has played out in recent debates over family law pluralism in Canada and the United Kingdom. It then moves to a discussion of “personal law” and dignity, in the paradigmatic (but not exhaustive) instance of India. The latter discussion is important because, whether known to Americans or not, the family law situation that is emerging in the United States (both before and after Proposition 8) strongly resembles a personal law system, i.e., a system of legal organization whereby different communities possess different laws within a given field of law (e.g. family law).⁷⁰ The lessons concerning dignity that the Indian system provides are thus quite instructive.

B. Private Ordering, Family Law Arbitration, and Dignity

This section will discuss two jurisdictions relatively familiar to the American lawyer—Canada⁷¹ and the United Kingdom—where religious minorities have used or are using non-state court arbitration (and “alternative dispute resolution” more broadly) to enforce family law norms that differ from those which are legislated by the state and enforced in state courts. In the academic literature, one commonly sees arbitration referred to as a type of “private ordering” of family law.⁷²

School Classrooms, THE FRESNO BEE, Oct. 15, 2008, available at <http://www.fresnobee.com/2008/10/14/937113/prop-8-tv-ad-raises-questions.html>; John Wildermuth, *Prop. 8 Supporters Fight Fierce TV Ad Battle*, S.F. CHRON., Oct. 11, 2008, available at http://articles.sfgate.com/2008-10-11/news/17134454_1_same-sex-marriage-ban-gay-marriage.

⁷⁰ Traditionally, personal law has been viewed as a kind of legal system that shares little with territorially-premised legal systems. I believe this view of things is wrong, however. See generally Jeffrey A. Redding, *Slicing the American Pie: Federalism and Personal Law*, 40 N.Y.U. J. INT'L L. & POL. 941 (2008). Indeed, in light of the pattern in U.S. state laws which is emerging with respect to the definition and enforcement of marriages versus domestic partnerships (or civil unions), it is time to question any easy conclusion about the existence of sharp differences between the American system of family law and Indian personal law. Indeed, just as Muslims and Hindus form families according to different laws in India, now so do homosexuals and heterosexuals utilize different family laws in some American states.

⁷¹ The particular jurisdiction within Canada that I will be focusing on here is that of Ontario. However, some of this discussion necessarily implicates discussion about Canada as a whole. Thus, depending on the situation, I will sometimes specifically refer to “Ontario,” and other times to “Canada” more generally.

⁷² See generally Ayelet Shachar, *Privatizing Diversity: A Cautionary Tale from Religious Arbitration in Family Law*, 9 THEORETICAL INQUIRIES LAW 573 (2008).

Arbitration, like personal law,⁷³ results in family law pluralism. However, arbitration differs from personal law in that the family law pluralism that results in a personal law system is (arguably) more dependent on, and more the creation of, the state. Arbitration, on the other hand, is imagined as existing “outside” of the state, and as providing an “alternative” to the state’s monolithic rules.⁷⁴ In this way, arbitration potentially allows for even greater family law pluralism than a personal law system does, as the potential variation in family law rules corresponds to the (larger) diversity found amongst cognizable *couples* (as opposed to cognizable *communities*) in society.

In 2003, Canadian politics become preoccupied with the issue of family law pluralism and, in particular, efforts by the Ontario-based Islamic Institute of Civil Justice (IICJ) to offer religiously-premised family law arbitration services to Muslims in Canada’s Ontario province. At the time, the president of this organization, Syed Mumtaz Ali, was said to have suggested that Canadian Muslims would not be “good Muslims” if they did not choose to have their family law issues decided outside of the secular Canadian legal system and according to Islamic law.⁷⁵ As one can imagine, coming as they did so soon after 9/11,

⁷³ See discussion *infra* Part II.C.

⁷⁴ As the Ontario “Boyd Report” described it:

[D]isputants may . . . give up on the quest for an agreed resolution to the[ir] dispute, and choose instead to have a neutral third party decide the[ir] dispute. When this is done by agreement of the parties to the dispute, it is known as arbitration. . . . [Arbitration is] private; [it does] not depend on “the law” to make [it] work, and [it does] not involve any governmental or state action.

MARION BOYD, DISPUTE RESOLUTION IN FAMILY LAW: PROTECTING CHOICE, PROMOTING INCLUSION 9-10 (2004) [hereinafter *Boyd Report*]; see also Shachar, *supra* note 72, at 580-81 (noting the difference between “calls for fair and just *inclusion* in the public sphere—the latter vividly captured by Iris Young’s image of a ‘heterogeneous public, in which persons stand forth with their differences acknowledged and respected’” and “claims for *opting out of*, or seceding from, the effects of the polity’s public laws and norms. Let us call the former pattern of multicultural inclusion *public accommodation*, and the latter, *privatized diversity*.”).

⁷⁵ *Boyd Report*, *supra* note 74, at 3 (interpreting a news report of Syed Mumtaz Ali’s comments at a conference); see also Judy Van Rhijn, *First Steps Taken for Islamic Arbitration Board*, LAW TIMES, Nov. 25, 2003, available at <http://www.freerepublic.com/focus/f-news/1028843/posts>. Prior to this 2003 conference, in a 1995 interview, Mr. Ali had also declared that

[a]s Canadian Muslims, you have a clear choice. Do you want to govern yourself by the personal law of your own religion, or do you prefer governance by secular Canadian family law? If you choose the latter, then you cannot

such efforts and statements struck a nerve in both secular and religious Canada, and much public controversy ensued.⁷⁶ While a great deal of this controversy was the result of Islamophobic and/or racist sentiment,⁷⁷ and overlooked the fact that Ontario Jews and Christians both had been using religiously-informed, legally-sanctioned arbitration to resolve their family law disputes for years,⁷⁸ it nonetheless represented a serious crisis for the Ontario government. As a result, a special report was commissioned by the provincial Government of Ontario in 2004, and this report, known as “the Boyd Report,” was issued at the end of 2004.⁷⁹ An examination of the Boyd Report’s discussion is instructive and important here, as this discussion demonstrates the existence of differing visions of the relationship between dignity and family law pluralism than those articulated by the California and Connecticut Supreme Courts.

At the time of the controversy, Ontario’s Arbitration Act⁸⁰ could be used to arbitrate a variety of family law (including inheritance) disputes outside of the courts, according to any body of law that the parties to the dispute chose. Certain

claim that you believe in Islam as a religion and a complete code of life actualized by a Prophet who you believe to be a mercy to all.

Interview by Rabia Mills with Syed Mumtaz Ali, President, Canadian Society of Muslims (Aug. 1995), <http://muslim-canada.org/pfl.htm>. That being said, Syed Mumtaz Ali’s organization, the Canadian Society of Muslims, also stated in 2003 that

[o]nce [a] matter comes to [Muslim arbitration,] the parties will be free to choose the law that they wish to rely upon. This model will not exclude application of Canadian laws if the parties wish to do so. It is expected that the Muslim Law and associated Case Law created through the old Anglo-Mohammadan Law precedents would be the model for Personal Law cases initially, but any other *Fiqh* could also be relied upon if the parties so desire.

Darul-Qada: Beginnings of Muslim Civil Justice System in Canada, CAN. SOC’Y MUSLIMS NEWS BULL., Apr. 2003, available at <http://muslim-canada.org/news03.html>.

⁷⁶ See *infra* note 97.

⁷⁷ The *Boyd Report* acknowledges this explicitly. See *Boyd Report*, *supra* note 74, at 68.

⁷⁸ See *id.* at 55-57. This report notes that representatives of one Jewish organization providing family law arbitration services told investigators for the report that Orthodox Jews are forbidden by their religion from bringing their legal disputes before “secular judges.” *Id.* at 55. The report also received a submission from one Christian organization (the Christian Legal Fellowship) representing hundreds of Christian lawyers, law professors, and law students, in which it was noted that “[m]any [faith] communities may feel that their core values, including the sanctity of the nuclear family are threatened by having their disputes resolved outside of their faith community by persons having no familiarity with their belief system.” *Id.* at 56.

⁷⁹ See *id.*

⁸⁰ Arbitration Act, R.S.O., ch. 17 (1991), available at http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_91a17_e.htm#BK3.

family law issues were outside of the power of an arbitrator to decide in a legally binding manner, including the basic status of a marriage (i.e. an arbitrator cannot declare a divorce; only a civil court can) and the custody of any children.⁸¹ However, disputes pertaining to spousal division of property, spousal support, child support, and inheritance could all be conclusively decided outside of the state's courts,⁸² in front of any kind of arbitrator (e.g. a Jewish rabbi, or a Muslim imam), according to any body of law (religious or otherwise).⁸³

In the recommendations it laid out with respect to how religious family law arbitration should proceed in Ontario in the future, the Boyd Report attempted to walk a careful path between the possibility of two different kinds of legal regimes, each of which the report found extreme and undesirable. The first of these regimes the Boyd Report called "secular absolutism," and it identified this type of legal system with the legal regime presently found in France.⁸⁴ Under a "secular absolutist" system, "the state must abstain from any involvement in religious matters, and religious authorities must be prohibited from having *any authority whatsoever* over matters that are regulated elsewhere by state law," including, presumably, family law.⁸⁵ Under such a (secular) system of law, the state is where the definition and enforcement of one family law, for everyone, both begins and ends.

The other extreme to be avoided, according to the Boyd Report, is a system whereby any group, such as Canadian Muslims, is allowed to establish a "separate" legal regime "distinct from [that of] the rest of Canadians, with the goal of political autonomy for the . . . community in this country."⁸⁶ Such a system is problematic because

Ontarians do not subscribe to the notion of "*separate but equal*" when it comes to the laws that apply to us. . . . A policy of compelling people to submit to different legal regimes on the basis of religion or culture would be counter to [Canadian] *Charter* values. . . . Equality before and under the law, and the existence of a *single legal regime*

⁸¹ See generally *Boyd Report*, *supra* note 74, at 14, 16.

⁸² See generally *id.* at 11-28.

⁸³ See generally *id.* at 12, for a discussion of parties' freedom to choose both the arbitrator and the body of law which would apply to the resolution of their dispute.

⁸⁴ *Id.* at 89.

⁸⁵ *Id.* (emphasis added).

⁸⁶ *Id.*

available to all Ontarians are the cornerstones of our liberal democratic society.⁸⁷

While invoking the talismanic vocabulary of “separate but equal” to decry any extreme form of family law pluralism, the Boyd Report’s observations as to the desirability of family law uniformity were clearly agonized, and perhaps ambivalent. The report, for example, was forced to acknowledge—as any contemporary Canadian discussion of Canadian legal pluralism would have to—that Canada has a rich tradition of “separate but equal” legal regimes, most notably in historically-francophone Quebec and also the aboriginal First Nations territories. With respect to the legal situation of Quebec, the report noted how

the historical context clarifies why Britain tolerated the use of the French civil law in Quebec after defeating the French and why that system of law was continued in our Constitution. Indeed, Canada is a delicate balancing act where protection of the religious, language and legal rights of both French and English have marked our ethos from the beginning.⁸⁸

With respect to the First Nations and their legal particularity in the Canadian set-up, the Boyd Report was even more adamant—and, as a result, also more tortured—about the inapplicability of this “separate but equal” legal situation for any claim to an autonomous, religiously-premised and religion-controlled⁸⁹ system of (family) law for Muslims, or any other non-First Nations group:

To compare any group of people, whether they are distinct on a cultural, ethnic or religious basis, to the First Nations of Canada in this country’s legal and historical context reveals a misunderstanding of the nature of the relationship between the Canadian state and the First Nations. From [this report’s] perspective, comparisons in this direction are erroneous at best.⁹⁰

Ultimately, the report’s legal conclusions here, to their detriment, rested on arguments about the First Nations’ singularity in Canada’s Constitution Act and other important

⁸⁷ *Id.* at 88 (emphasis added).

⁸⁸ *Id.* at 79.

⁸⁹ As the *Boyd Report* describes this model: “According to such a conception of minority rights, the Muslim community, and other communities arbitrating family law matters using religious principles, would be able to do so based on whatever internal rules they adopt and the state would have no right to intervene.” *Id.* at 90.

⁹⁰ *Id.* at 87-88.

legislation.⁹¹ Perhaps at the time of this report's writing, this kind of argument looked like an unimpeachable and ingenuous one. Now, however, in light of American same-sex marriage opponents' invocation of the U.S. Constitution's Fourteenth Amendment's historical rooting in anti-racism—and *only* anti-racism⁹²—the Boyd Report's similar mode of argumentation looks intellectually half-hearted at best, and desperate and Islamophobic at worst.

Ultimately, the Boyd Report ended up endorsing the basic system of optional arbitration for select family law matters that then existed in Ontario, while making suggestions on the margins for reforms to this system.⁹³ As the report saw it, the benefits of this existing system included that it was consistent with the basic Canadian commitment to multicultural policies, which “[a]llow[] and support[] communities' and individuals' links to cultures (including their

⁹¹ *Id.* at 87.

⁹² See, for example, Lynn Wardle's argument that

[w]hatever else may be said about the Fourteenth Amendment, it is undeniable from both its text and its history that it was intended to outlaw state action designed to foster *racism*—to outlaw government policies that manifest the demeaning notion of *racial* inferiority. Three constitutional amendments especially embrace the value of *racial* equality in our legal system. By contrast, nothing in the Fourteenth Amendment discloses a comparable intent to protect or promote the social or legal equality of homosexual relations.

Lynn D. Wardle, *A Critical Analysis of Constitutional Claims for Same-Sex Marriage*, 1996 BYU L. REV. 1, 78-79 (1996) (emphasis added).

⁹³ The *Boyd Report* noted that it

did not find any evidence to suggest that women are being systematically discriminated against as a result of arbitration of family law issues. Therefore the Review supports the continued use of arbitration to resolve family law matters *The Arbitration Act should continue to allow disputes to be arbitrated using religious law, if the safeguards currently prescribed and recommended by this Review are observed.*

Boyd Report, *supra* note 74, at 133. Many of the reforms suggested by the *Boyd Report* are relatively minor, such as requiring arbitrators to provide written reasons for their decisions and to keep and transmit to the government better written records of their decisions. *Id.* at 140. Some recommendations are more significant, such as the recommendation to require that the agreement to arbitrate a family dispute be reconfirmed at the time of the family law dispute instead of, say, allowing an agreement entered into at the time of the marriage to necessarily hold sway. *Id.* at 134. A potentially important recommendation is that the Arbitration Act should be amended to more concretely define what its requirement of a “fair and equal process” in arbitration means. *Id.* at 136.

religions) of origin,”⁹⁴ and that, at heart, this existing system supported “inclusion which takes account of difference.”⁹⁵

Despite the Boyd Report’s basic endorsement of the status quo, the Government of Ontario nonetheless rejected the report’s recommendations, and indeed went so far as to make illegal any arbitration conducted according to any body of law other than the law of Ontario or of another Canadian

⁹⁴ *Id.* at 90. Ayelet Shachar, another Canadian defender of the availability of some form of religious family law arbitration, has similarly stressed how religious law can “offer religious women a significant source of meaning and value,” Shachar, *supra* note 72, at 575, and as a result, can leave them feeling “obliged to have at least some aspects of their marriage and divorce regulated by religious principles and communal institutions,” *id.* at 604. Shachar has also argued the decision to ban religious arbitration is “not an ideal normative and jurisprudential solution,” given that the government’s

“out of sight, out of mind” approach [to religious arbitration] will probably not be of much assistance to vulnerable group members in blocking communal pressures to resolve family disputes by turning to “their” group’s authorities which, now legally unrecognized, remain free of *any* regulatory oversight, whether *ex ante* or *ex post*.

Id. at 604-05 (emphasis added).

⁹⁵ *Boyd Report*, *supra* note 74, at 89. The report continued onward to distinguish its endorsement of “inclusion which takes account of difference” from “exclusion based on difference.” *Id.* (emphasis added). Again, however, this statement about arbitration as inclusion, instead of “separate but equal” exclusion, is a curious one, and appears to be motivated by the Boyd Report’s need to distinguish religious arbitration from Quebecois or First Nation legal separatism. The Boyd Report’s distancing moves in this respect are somewhat dubious, however, especially when they result in the statement that Jews’, Muslims’, and others’ resort to religious arbitration—instead of the state’s courts—ultimately amounts to a vigorous *endorsement* by religious communities of the state and its legal norms and institutions:

By availing itself of provincial legislation that has been in place for over a decade, and that has been used by others, the Muslim community is drawing on the dominant legal culture to express itself. By using mainstream legal instruments minority communities openly engage in institutional dialogue. And by engaging in such dialogue, a community is also inviting the state into its affairs, particularly since the *Arbitration Act*, even in its present form, specifically sets out grounds for state intervention in the form of judicial oversight. Use of the *Arbitration Act* by minority communities can therefore be understood as a desire to engage with the broader community.

Id. at 93.

In fact, opposition to and hostility towards the state’s system of courts and legal administration was relatively strong amongst some groups. For example, the Orthodox Jewish non-state court in Toronto (*Beis Din*) even opposed the Boyd Report committee’s relatively timid exploration of enhanced training for and regulation of religious arbitrators. *See id.* at 116-17. With respect to aboriginal peoples, the Boyd Report also acknowledged the submission of the Ontario Federation of Indian Friendship Centres, and its concerns that state regulation of arbitrators working on aboriginal family law matters would “tend to ignore the wisdom and experience so important within [our] communities and tie the process to the ‘white man’s system of justice,’ from which the community seeks relief.” *Id.* at 117 (paraphrasing the submission by the Ontario aboriginal group).

jurisdiction.⁹⁶ This significant change in the law of arbitration was clearly the consequence of post-9/11 heightened anxiety concerning the loyalties and intentions of Canadian Muslims.⁹⁷ This dramatic post-Boyd Report turn of events notwithstanding, the Boyd Report's discussions and conclusions, as well as the politics to which they are a response, are instructive and important in that they demonstrate that alternative visions of the relationship between dignity and family law pluralism exist and are potentially viable in the modern, secular state.⁹⁸

While a certain sort of family law pluralism has been shut down in Canada post-Boyd, the Islamophobia that underlies this move is not necessarily instructive of how dignity-minded individuals and governments should themselves come out on the question of family law pluralism. As the present situation in the United Kingdom suggests, other

⁹⁶ See Family Arbitration Regulations (Arbitration Act), R.R.O./2007-134 (Ont.). After this amendment, the Arbitration Act in Ontario now reads:

Other third-party decision-making processes in family matters

2.2 (1) When a decision about a matter described in clause (a) of the definition of "family arbitration" in section 1 is made by a third person in a process that is not conducted exclusively in accordance with the law of Ontario or of another Canadian jurisdiction,

(a) the process is not a family arbitration; and

(b) the decision is not a family arbitration award and has no legal effect.
2006, c. 1, s. 1 (2).

See Arbitration Act, R.S.O., ch. 17 (1991), available at http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_91a17_e.htm#BK3.

⁹⁷ See Shachar, *supra* note 72, at 584; see also Haroon Siddiqui, Op-Ed., *Sensationalism Shrouds the Debate on Sharia*, TORONTO STAR, June 12, 2005, at A17.

⁹⁸ They are also witness to the fact that religious persons are in the forefront of efforts to reform secularism and the hegemonic political embodiments, such as the state, with which secularism has often been associated. This is not to say that religious people in Canada were united in challenging the preeminence of the Canadian state's role in regulating family relationships; they were not. In this respect, the *Boyd Report* was exemplary in its serious engagement with differences of opinion *amongst Muslims* (as well as amongst people of other religious faiths) about the proper goals of the community—including how best to obtain respect and dignity for this community. These differing views spanned the spectrum from a desire to establish a completely autonomous legal system for Canadian Muslims, see *Boyd Report*, *supra* note 74, at 88, to those of the Muslim Canadian Congress (MCC). The MCC is described as a private national organization that viewed itself as "progressive," and which also claimed that the Arbitration Act "does *not* cover family law disputes" and "that if indeed the government takes the position . . . that the Arbitration Act can deal with these matters, then the . . . Act is unconstitutional . . . in that . . . [it b]reaches the unwritten constitutional norms enunciated by the Supreme [C]ourt of Canada . . . namely the rule of law, constitutionalism, federalism, and respect for minorities." *Id.* at 29-30.

jurisdictions—equally afflicted by Islamophobia—might be on a different path.

In early 2008, the Archbishop of Canterbury, Rowan Williams, delivered a widely reported-upon and controversial talk in the United Kingdom on the topic of “Civil and Religious Law in England: A Religious Perspective.”⁹⁹ Conceived as a general talk about how to respond to “the presence of communities [in the United Kingdom] which, while no less ‘law-abiding’ than the rest of the population, relate to something other than the British legal system alone,”¹⁰⁰ the Archbishop’s words resonated widely and loudly in a country still recovering from the 2005 attacks on its capital’s public transportation system, and the fears of a Muslim “fifth-column” that these attacks engendered. Journalistic reporting of the lecture focused on its comments concerning the place of Islamic law¹⁰¹ in an ostensibly secular¹⁰² legal system. However, the Archbishop himself emphasized that he was trying to speak *generally* “about the right of religious believers . . . to opt out of certain legal provisions—[for example,] the problems around Roman Catholic adoption agencies which emerged in relation to the Sexual Orientation Regulations [the previous spring].”¹⁰³

While the Archbishop’s widely-publicized speech was a response to recent events and concerns, debates concerning the limits to legal pluralism in the United Kingdom have actually been ongoing for some time. For example, in the 1970s, U.K. Muslim organizations organized to demand the formal recognition of a separate system of family law in the United

⁹⁹ See Rowan Williams, Archbishop of Canterbury, Lecture at the Royal Courts of Justice, Civil and Religious Law in England: A Religious Perspective (Feb. 7, 2008), available at www.archbishopofcanterbury.org/1575.

¹⁰⁰ *Id.*

¹⁰¹ See, e.g., Ruth Gledhill & Phillip Webster, *Archbishop of Canterbury Argues for Islamic Law in Britain*, TIMES, Feb. 8, 2008, available at <http://www.timesonline.co.uk/tol/comment/faith/article3328024.ece>; Jonathan Petre & Andrew Porter, *Adopt Sharia Law in Britain, Says the Archbishop of Canterbury Dr. Rowan Williams*, DAILY TELEGRAPH, Feb. 8, 2008, available at <http://www.telegraph.co.uk/news/uknews/1578017/Adopt-sharia-law-in-Britain-says-the-Archbishop-of-Canterbury-Dr-Rowan-Williams.html>; *Sharia Law in UK is ‘Unavoidable’*, BBC NEWS, Feb. 7, 2008, available at http://news.bbc.co.uk/2/hi/uk_news/7232661.stm.

¹⁰² It is somewhat of a challenge to characterize the English legal system as “secular” when, as the Archbishop himself acknowledged, “the law of the Church of England is the law of the land.” Williams, *supra* note 99. The Archbishop went on to note, however, that the “daily operation” of that Church law “is in the hands of [non-Church] authorities to whom considerable independence is granted.” *Id.* That being said, later in his talk, the Archbishop spoke admirably of what he characterized as a necessary “theology of law.” *Id.*

¹⁰³ *Id.*

Kingdom for Muslims.¹⁰⁴ While these efforts to garner the state's official endorsement and enforcement of a separate family law system for Muslims in the United Kingdom were essentially unsuccessful, Muslim non-governmental organizations have developed a number of non-state Muslim legal institutions all over the United Kingdom in the past two decades.

These institutions, or "shari'a councils," use procedures and practices informed by Islamic legal and moral norms to provide mediation and family law dispute resolution services for disputes arising in Muslim families. They identify themselves with names like "Muslim Marriage Guidance Council," "Islamic Sharia Council," and "Muslim Arbitration Tribunal."¹⁰⁵ Most of these institutions see themselves as merely mediators in Muslim couples' mundane problems and disagreements, offering non-binding advice as to Islamic family norms. Some of these institutions also hear and decide individuals' petitions for religious divorce, and issue religious divorces.¹⁰⁶ However, these declarations of divorce have no civil law effect, since only a state court can declare an officially-married couple legally divorced.¹⁰⁷ Only one institution, the Muslim Arbitration Tribunal, has taken the steps to officially register itself under the state's Arbitration Act, so that it may

¹⁰⁴ See generally Sebastian Poulter, *The Claim to a Separate Islamic System of Personal Law for British Muslims*, in ISLAMIC FAMILY LAW 147 (Chibli Mallat & Jane Connors eds., 1990).

¹⁰⁵ See generally Sameer Ahmed, *Pluralism in British Islamic Reasoning: The Debate Over Official Recognition of Islamic Family Law in the United Kingdom* 50-60 (2006) (unpublished Ph.D. dissertation, Oxford University) (on file with author); see also John R. Bowen, *Private Arrangements: "Recognizing Sharia" in England*, BOSTON REV., March/April 2009, at 15 (providing a general overview of the functioning of the Muslim Arbitration Tribunal and Islamic Sharia Council).

¹⁰⁶ For example, John Bowen reports that at the February 2008 monthly meeting of scholars associated with the Islamic Sharia Council that, with respect to the seven cases that these scholars heard as a group that month, the scholars either dissolved the marriage in question or deferred a decision and asked for more information. Incidentally, all seven cases were requests by women to divorce their husbands. See Bowen, *supra* note 105, at 16. For a general overview of these institutions' functions, see Samia Bano, *In Pursuit of Religious and Legal Diversity: A Response to the Archbishop of Canterbury and the 'Sharia Debate' in Britain*, 10 ECCLESIASTICAL L.J. 283, 294-96 (2008). For a detailed scholarly study of one such institution, namely the Muslim Law (Shariah) Council, based in West London, see generally SHAH-KAZEMI, *supra* note 60.

¹⁰⁷ See Bowen, *supra* note 105, at 16; see also Lucy Carroll, *Muslim Women and 'Islamic Divorce' in England*, 17 J. MUSLIM MINORITY AFF. 97 (1997), available at <http://www.wluml.org/node/304>.

resolve civil (including intra-family) disputes¹⁰⁸ in a legally binding manner, using the tools of state-defined arbitration.

As in Canada,¹⁰⁹ Muslim opinion in the United Kingdom as to the desirability of establishing a distinct set of legal institutions for Muslims is not univocal; there are both Muslim supporters and Muslim detractors of efforts to establish non-state Muslim legal institutions. For example, as in Canada, some Muslims see the effort to establish officially-recognized and supported Islamic law in the United Kingdom as no different than—and as necessary as—the state’s recognition of sub-national territorial-cum-community laws. For example, one Muslim commentator has remarked that “[T]his country has already two laws—one law of inheritance applies to England and Wales and one law of inheritance applies to Scotland. How are these two laws able to coexist peacefully without disrupting the legal system of this country? Similarly, Islamic family law can coexist with this law without disrupting the whole legal structure.”¹¹⁰

Other Muslims, while supporting non-state Muslim legal institutions (such as shari‘a councils), believe that the effects on the Muslim community that could result from the state establishing or officially-recognizing Islamic legal institutions might be extremely detrimental. These possible effects include a potential exacerbation of intra-community communal tensions as groups vie with each other for the state’s patronage, or a corruption in the content of Islamic law as state concerns and priorities come to infiltrate previously autonomous religio-legal discussions.¹¹¹ Other Muslims worry explicitly about any sort of Muslim separateness, with these worries echoing those found in the U.S. about “separate but equal” legal regimes. For example, one commentator has argued that “Muslims should try to integrate themselves into society. . . . A separate system would create a *stigma* and lead people to discriminate against Muslims.”¹¹² Finally,

¹⁰⁸ Not all of these arbitration matters involve intra-family civil disputes. The website of the Muslim Arbitration Tribunal reports that they also handle “Commercial and Debt Disputes” and “Mosque Disputes.” See Muslim Arbitration Tribunal, <http://www.matribunal.com/cases.html> (last visited Feb. 10, 2010).

¹⁰⁹ See *supra* note 98.

¹¹⁰ Syed Aziz Pasha, Union of Muslim Org., Address in London (Aug. 22, 2004), in Ahmed at 79.

¹¹¹ See *id.* at 83-84.

¹¹² *Id.* at 85 (emphasis added); see also Samia Bano’s worry about the development of a “new normative discourse, which stigmatises Muslims as the

commentators have expressed worry that the welfare of Muslim women can be compromised by the “privatization” of family law enforcement and efforts to increasingly locate that enforcement in non-state, community-premised—and potentially patriarchal—bodies and organizations.¹¹³

This diversity of opinions being the case, there is evidence suggesting that the number of Muslims using the services of these non-state Muslim institutions might very well be on a steady rise.¹¹⁴ If that is actually the case, this would not be surprising in light of the finding by one recent poll of (500) British Muslims “that a clear majority [of those polled] want Islamic law introduced into this country in civil cases relating to their own community. Some 61% wanted Islamic courts—operating on shari’a principles—‘so long as the penalties did not contravene British law.’”¹¹⁵ Another recent study suggests that 37% of British Muslims aged 16-24 “would prefer to live under *sharia* law [as opposed to British law],” which is significantly higher than the 17% of British Muslims 55-years-old and older who would prefer the same.¹¹⁶

‘Other’—in conflict with, incompatible with and, most importantly, disloyal to the state.” Bano, *supra* note 106, at 287. Bano goes on to argue that “Muslim engagement with the law and sharia must be read within the broader social and political context in which [Muslims] operate [and] must . . . not fall in to the traps of cultural essentialism and homogeneity that reproduce the binaries that [one] seeks to dismantle and displace.” *Id.*

¹¹³ See, e.g., Bano, *supra* note 106, at 300-01. Bano is critical of ongoing discussions concerning shari’a councils which do not take into account the experiences and views of “Muslim women, who are the primary users of [shari’a councils].” *Id.* at 288. While Bano’s research reports a variety of views amongst Muslim women with respect to shari’a councils—with some women enthusiastically supporting these councils and other women far more skeptical—Bano herself is clearly troubled by efforts to enhance the powers and authority of shari’a councils. See *id.* at 309 (noting that Bano is writing “with the conviction that Muslim women remain extremely cautious of initiatives to accommodate sharia into English law”); see also SHAH-KAZEMI, *supra* note 60, at 70 for her research findings that “formal recognition of the shari’a system of laws in Britain would be problematic, and such recognition is not sought by . . . the majority of Muslim community organisations.”

¹¹⁴ The Islamic Sharia Council, one major such non-state Muslim legal institution, reports that from 1982-1995, 1500 cases were filed with it. From 1996-2009, however, at least 5500 cases were filed. Islamic Sharia Council, Islamic Sharia Council—About Us, <http://www.islamic-sharia.org/about-us/about-us-9.html> (last visited Feb. 10, 2010).

¹¹⁵ See Alan Travis & Madeleine Bunting, *British Muslims want Islamic law and prayers at work*, THE GUARDIAN, Nov. 30, 2004, available at <http://www.guardian.co.uk/uk/2004/nov/30/immigrationpolicy>.

¹¹⁶ See MUNIRA MIRZA, ABI SENTHIKUMARAN & ZEIN JA’FAR, LIVING APART TOGETHER: BRITISH MUSLIMS AND THE PARADOX OF MULTICULTURALISM 5 (2007), cited in Samia Bano, *Islamic Family Arbitration, Justice and Human Rights in Britain*, LAW, SOCIAL JUSTICE & GLOBAL DEVELOPMENT (2007), available at http://www.go.warwick.ac.uk/elj/lgd/2007_1/bano. In this piece by Samia Bano, Bano is critical of the

Clearly, people's responses to survey questions are more complicated and nuanced than any crude statistic can capture. However, these numbers in support of a separate legal system for a U.K. minority are nonetheless surprisingly robust, especially in light of the usual liberal claims that "separate" is necessarily "unequal." If that liberal claim is right, it appears that substantial numbers of British Muslims want to be stigmatized as unequal. While that is a possibility, what appears more probable is that substantial numbers of Muslims in the United Kingdom contest majority practices and values,¹¹⁷ including Islamophobia.¹¹⁸ Other Muslims worry less about majority ill-will than they do about majority (cultural) incompetence.¹¹⁹

While the future direction of the debate over official recognition (in some manner) of Islamic (family) law in the United Kingdom is entirely unpredictable, the fact that the head of the Church of England is making speeches speaking favorably of (some) Islamic legal institutions, and advocating more legal pluralism, suggests that monumental change is afoot. Whatever the outcome(s) of this debate, its existence, similar to the Ontario debate, demonstrates that alternative visions of the relationship between dignity and family law pluralism exist and are viable in the modern, secular state.

In both Canada and the United Kingdom, then, members of religious minorities have recently deployed arguments relating to dignity to argue *against* the universal application of majority-defined state family law norms. In Canada, these arguments ultimately proved unsuccessful in the face of a dignity-defying Islamophobia, and religious family

survey methodology used by Mirza, Senthikumaran & Ja'far and contests the accuracy of their findings. *See id.*

¹¹⁷ Speaking of Muslims in England, Ihsan Yilmaz writes that "[m]ost . . . see Western society as aimless and rootless, marred by increasing vandalism, crime, juvenile delinquency, the collapse of marriages, growing numbers of illegitimate children, and near constant stress and anxiety. They view Islam as the positive alternative." Ihsan Yilmaz, *Muslim Alternative Dispute Resolution and Neo-Ijtihad in England*, 2 ALTERNATIVES: TURKISH J. OF INT'L REL. 121-22 (2003).

¹¹⁸ Yilmaz, *supra* note 117, notes the disparity in how English Jews and Sikhs are protected under the Race Relations Act, but not Muslims. "As a result, there has been widespread alienation from the state among [Muslims]." *Id.* at 122.

¹¹⁹ See SHAH-KAZEMI, *supra* note 60, at 53-55, 71-77 for examples and discussion of incompetence on the behalf of British (non-Muslim) lawyers giving advice to their Muslim clients on both English and Islamic law. In one instance, one of these lawyers drew up a *talaqnama* for his female client, in which he had his client—a woman—attempt to divorce her husband by pronouncing "I TALAK YOU" thrice. *See id.* at 54-55.

law arbitration in one leading province of that country (Ontario) has been severely curtailed. Both Jews and Muslims—the two minorities who had been the most vocal in trying to protect the availability of religiously-informed non-state family law arbitration for their communities—have been forced to abide by the state’s legislated family law rules.¹²⁰ In the United Kingdom, the debate is gaining momentum. Following the important and widely-discussed speech by the Archbishop of Canterbury on legal pluralism and dignity, the issue of non-state arbitration for religious minorities’ family law issues is very much on the national radar, as is the question how the dignity of religious minorities can be enhanced by the existence of increased family law pluralism. In short, in both Canada and the United Kingdom, a positive relationship between dignity and legal pluralism has been discussed and made possible; any assumption of dissonance and incoherence between these two ideas is itself incoherent in these non-American contexts.

C. *Personal Law, “Separate But Equal” Family Laws, and Minority Rights*

The debates in Canada and the United Kingdom concerning family law pluralism are, in part, a debate about “private ordering,” or the ability of people to “privately” construct alternatives to the state’s monolithic family law rules, norms, and assumptions. However, another model of family law pluralism—namely, that of “personal law”—is also widely practiced and debated around the globe. In contrast to the *private* ordering model, this form of family law pluralism is one where *the state itself* is explicitly involved in defining and/or enforcing¹²¹ different family laws for different communities.

As a method of legislating and administering laws, personal law has a long history, dating back at least to the time of the Romans.¹²² However, personal law is still found all over

¹²⁰ See *supra* note 96.

¹²¹ States that have personal law systems will differ to the extent they will allow communities to legislate, administer, and otherwise enforce their particular personal laws. There is no single model of a personal law system, though there are commonalities between such systems. For a comparison of two widely-studied personal law systems, see Marc Galanter & Jayanth Krishnan, *Personal Law and Human Rights in India and Israel*, 34 *ISR. L. REV.* 101, 115 (2000).

¹²² See FRIEDRICH KARL VON SAVIGNY, *PRIVATE INTERNATIONAL LAW AND THE*

the modern world. While often, but not always, the product of European colonial rule, this kind of (family) law system has been retained in many post-colonial states, including Israel, Malaysia, Pakistan, and India.¹²³

At a very general level, a personal law system is a legal system in which laws or legal norms bind “different” people differently, sorting people into various legal regimes depending on the “type of person” involved.¹²⁴ The aspects of personhood that most contemporary personal law systems use to distinguish between people are those relating to religion and ethnicity.¹²⁵

As indicated, India is one prominent country where the administration of family law is organized around a personal law model. India’s personal law system is one that is premised on people’s religiously communal identifications.¹²⁶ When people refer to India’s personal law system, then, they mean the system of Indian family law whereby Hindus, Muslims, Christians, and others are governed by different family law codes, practices, and norms.¹²⁷ In this system of family law, one finds the “Hindu Marriage Act” (which also governs divorces

RETROSPECTIVE OPERATION OF STATUTES: A TREATISE ON THE CONFLICT OF LAWS AND THE LIMITS OF THEIR OPERATION IN RESPECT OF PLACE AND TIME 58-59 (William Guthrie trans., Lawbook Exchange 2003) (1880).

¹²³ See generally Redding, *supra* note 70.

¹²⁴ See generally *id.* The factors that are important to personhood may differ from society to society. As a result, any given personal law system might look unlike any other such system. However, what characterizes all personal law systems is that the law which applies to one in such systems depends on the “kind of person” one is, instead of on one’s generic membership in an undifferentiated polity. See generally *id.*

The terms “personhood” and “type of persons” are used here to emphasize that not every law that distinguishes between persons is a personal law, but only those laws that distinguish between socially and politically relevant “types” of people. This, obviously, will differ from society to society. For example, “high-caste” and “low-caste” people are relevant types of people in India, in a way that they are not for the vast majority of Americans. Race rather than caste, in this respect, is more central to the American discussion. See generally *id.*

¹²⁵ See generally *id.* That being said, personal law is not just law that distinguishes between people with different kinds of communal or kinship ties (religion and ethnicity being two prime examples of such ties). The term “personal law” has, in fact, not been strictly limited (either historically or contemporarily) in this way. See, e.g., RAMANI MUTTETUWEGAMA, PARALLEL SYSTEMS OF PERSONAL LAWS IN SRI LANKA 3-5 (1997) (discussing the quasi-territorial, quasi-ethnic aspects of Sri Lankan personal law).

¹²⁶ Galantar & Krishnan, *supra* note 121, at 103.

¹²⁷ *Id.* at 109. Presently in India, the central government (and, to a much lesser degree, state governments) legislates on different religious communities’ personal laws. Furthermore, there is a relatively unified, hierarchically-organized national judiciary in India that enforces and administers this legislation, as well as the large amount of uncodified religious personal law that is found in judicial precedents. *Id.* at 109.

between Hindu marital parties), and also the “Indian Christian Marriage Act.”¹²⁸ Furthermore, the “Indian Divorce Act” governs Christian divorces, while the “Dissolution of Muslim Marriages Act” governs some kinds of Muslim divorces.¹²⁹ There are also many other examples of these kinds of statutes in India, as well as a large body of religion-specific, judicially-developed common law that relates to the family.¹³⁰

While the motivations behind personal law systems are surely complex and dynamic over the course of history,¹³¹ today they are in very large part “intended to help ethnic groups and religious minorities express their cultural particularity and pride without it hampering their success in the economic and political institutions of the dominant society.”¹³² Looking at

¹²⁸ *Id.* at 109 n.42.

¹²⁹ The Indian Divorce Act, No. 4 of 1869, India Code, amended by The Indian Divorce (Amendment) Bill, Act No. 51 of 2001, India Code, available at <http://indiacode.nic.in/> (search “search Indiacode: Short Title” for “The Indian Divorce Act”; then follow “Download full act” hyperlink under search results); The Dissolution of Muslim Marriages Act, No. 8 of 1939, India Code, available at <http://indiacode.nic.in/> (search “search Indiacode: Short Title” for “The Dissolution of Muslim Marriages Act”; then follow “Download full act” hyperlink under search results).

¹³⁰ Galantar & Krishnan, *supra* note 121, at 109. There is also family law (for example, the recently-enacted “Protection of Women from Domestic Violence Act of 2005”) which is not administered along communitarian lines. See The Protection of Women from Domestic Violence Act, No. 43 of 2005, India Code, available at <http://indiacode.nic.in/> (indicating that this Act is applicable to “any woman”).

¹³¹ India’s present personal law system can be traced back at least to the 1772 decision by Warren Hastings, the British viceroy for India at the time, to “in all Suits regarding Marriage, Inheritance, Cast, and other religious Usages or Institutions, [apply] the Laws of the Koran with respect to [Muslims], and those of the Shaster with respect to [Hindus].” A Plan for the Administration of Justice (1772); see also WILLIAM H. MORLEY, THE ADMINISTRATION OF JUSTICE IN BRITISH INDIA; ITS PAST HISTORY AND PRESENT STATE: COMPRISING AN ACCOUNT OF THE LAWS PECULIAR TO INDIA 177, 177-78 (1858). For a discussion of this British policy, see Galanter & Krishnan, *supra* note 121, at 106; see also M.B. HOOKER, LEGAL PLURALISM: AN INTRODUCTION TO COLONIAL AND NEO-COLONIAL LAWS 60 (1975).

While one might have expected otherwise from such an ambitious announcement, ultimately Hastings’ decision was only fully implemented in the areas of marriage, divorce, inheritance, and adoption law, as well as in the management of religious endowments. After independence, and after much debate, the post-colonial Indian state decided to continue this basic split between universally oriented criminal law and personally oriented family law.

¹³² WILL KYMLICKA, MULTICULTURAL CITIZENSHIP: A LIBERAL THEORY OF MINORITY RIGHTS 31 (1995). Rina Verma Williams has characterized the post-colonial retention of personal law systems as “a way to avert ethnic unrest and preserve cultural autonomy in multiethnic societies.” RINA VERMA WILLIAMS, POSTCOLONIAL POLITICS AND PERSONAL LAWS: COLONIAL LEGAL LEGACIES AND THE INDIAN STATE 7 (2006). Finally, India’s post-Independence leader, Jawaharlal Nehru, himself remarked that “we do not dare touch the Moslems [with respect to their personal law] because they are a minority and we do not wish the Hindu majority to do it. These are personal laws and so they will remain for the Moslems, unless they want to change them.” See TIBOR MENDE, CONVERSATIONS WITH MR. NEHRU 57 (1956), cited in WILLIAMS, *supra*,

personal law systems in India and elsewhere, what is interesting to note is that “second class” citizens—for example, Muslims in the case of Hindu-majority India—often oppose any effort to amalgamate them into a common, unitary family law system.¹³³

Perhaps the best example of this kind of opposition to majoritarian absorption, in the Indian context at least, is a still-potent controversy which dates from the mid-1980s. This controversy, widely known as “the Shah Bano crisis,” resulted from a decision handed down by the Indian Supreme Court in the case of *Mohd. Ahmed Khan v. Shah Bano Begum*.¹³⁴ The question presented was whether the Indian Code of Criminal Procedure’s requirement that a man indefinitely financially maintain his ex-wife after a divorce if she is “unable to maintain herself”¹³⁵ was applicable to Muslim men, who arguably have more limited responsibilities¹³⁶ toward their ex-wives under classical Islamic family law. Ultimately, the Indian Supreme Court determined that 1) the Code of Criminal Procedure’s requirements superseded any contradictory Muslim personal law rules and requirements,¹³⁷ and 2) nothing in Muslim personal law forbade indefinite maintenance to a divorced wife “who is unable to maintain herself.”¹³⁸

at 116. *But see* MAHMOOD MAMDANI, *CITIZEN AND SUBJECT: CONTEMPORARY AFRICA AND THE LEGACY OF LATE COLONIALISM* 111 (1996) (arguing that colonial-era legal pluralism “was more an expression of power relations in a colonial society than a recognition and tolerance of any multicultural diversity”).

¹³³ It is important to note here that at India’s independence, conservative Hindu organizations also opposed the newly independent state’s (ultimately successful) attempts to reformulate Hindu personal law, using arguments about the inappropriateness of (secular) state “interference” in religious personal laws. *See* WILLIAMS, *supra* note 132, at 19, 104-14. Later, this particular brand of Hindu politics radically changed, such that while “[i]n the 1980s, religious identity for the Muslim community became virtually coterminous with the preservation of their personal law[, f]or some Hindus, . . . Indian national identity became virtually coterminous with forcing the Muslim community to give up their personal law.” *Id.* at 127.

¹³⁴ (1985) 3 S.C.R. 844.

¹³⁵ INDIA CODE CRIM. PROC. § 125(1)(a).

¹³⁶ Under most classical interpretations of Islamic divorce law, it is generally the rule that a man is required to financially maintain his (ex-)wife up until the time she has, post-divorce, menstruated three times. *See* DAVID PEARL & WERNER MENSKI, *MUSLIM FAMILY LAW* 182-84, 280-82 (3d ed. 1998).

¹³⁷ *Mohd. Ahmed Khan v. Shah Bano Begum*, (1985) 3 S.C.R. 844, 854-56.

¹³⁸ *Id.* at 859-62. Arguably, the first holding was sufficient to have settled the case, and it was gratuitous and provocative for the Indian Supreme Court to have interpreted the Muslim community’s personal law. This seems especially the case given that other portions of the court’s opinion took a patronizing tone in regards to the content of such personal law. The lead paragraph in this opinion, in fact, included the following remarks: “it is alleged that the fatal point in Islam is the degradation of woman. To the Prophet is ascribed the statement, hopefully wrongly, that Woman was

The opinion ignited large protests by conservative Muslims across India (and smaller counter-protests by a number of dissident Muslim women and their allies).¹³⁹ Eventually, then-Prime Minister Rajiv Gandhi and his government acquiesced to conservative Muslim demands to pass a law to eliminate Muslim (and only Muslim) women's rights to petition for and receive indefinite post-divorce maintenance from their ex-husbands.¹⁴⁰ While the legal effect of this relatively recent addition to India's personal law system has been whittled back over time, the law still remains on the books, and Muslim political and social organizations would most likely intensely resist its removal.¹⁴¹

This dispute over Muslim personal law is both cause and symptom of a larger social and political debate about the secular credentials of a post-colonial Indian state (as opposed to the "Islamic" post-colonial Pakistani state). There is no foreseeable end to this debate, but neither is there any foreseeable end to the enforcement of personal law. Amongst India's religious minorities, it is common to find antagonism to the idea that everyone in India should be bound to one uniform civil (family law) code. While religious feminists are working

made from a crooked rib, and if you try to bend it straight, it will break; therefore treat your wives kindly." *Id.* at 849-50 (internal quotation marks omitted).

¹³⁹ See Kirti Singh, *The Constitution and Muslim Personal Law*, in FORGING IDENTITIES: GENDER, COMMUNITIES, AND THE STATE 96, 101-03 (Zoya Hasan ed., 1994); WILLIAMS, *supra* note 132, at 145 (documenting smaller size of counter-protests by progressive Muslims).

¹⁴⁰ See The Muslim Women (Protection of Rights on Divorce) Act, No. 25 of 1986, available at <http://indiacode.nic.in/fullact1.asp?tfnm=198625>. In response to this legislation, cries of "appeasement" were effectively raised by Hindu nationalist quarters, which eventually helped lead to the national electoral successes of the Hindu-nationalist BJP political party. These successes, in turn, led to a severe polarization in Hindu-Muslim relations in India, a corresponding increase in violence between the two communities, and the drawing of new and sharper boundaries between the two communities. These communal problems, and the challenges they present for legislation and judicial decision-making in the area of personal law, persist today. See Redding, *supra* note 70, at 967-68.

¹⁴¹ For the results of different surveys of Muslim public opinion on the issue of personal law reform, see WILLIAMS, *supra* note 132, at 58. For example, a 1996 survey found that 67% of Muslims (and over 50% of Christians) favored the retention of India's personal law system, while only 42% of Hindus favored keeping this system. *Id.* Another 1995 survey of 200 Muslim women found that while 62% of respondents thought that Muslim personal law in India should be reformed in at least one aspect or another, only 14% would go so far as to eradicate the Indian method of organizing family law along a personal law model itself. See Sabeeha Bano, *Muslim Women's Voices*, 47 ECON. & POL. WKLY 2981, 2982 (1995). All of these results should be appropriately contextualized and qualified by noting both the enormous size of India's Muslim population—approximately 150 million—and the large number of class, caste, regional, and sectarian differences which internally differentiate this population.

for more women-friendly versions of personal law, such a project has goals different than delegating family law solely to patriarchal others, whether those “others” be religiously- or secularly-spirited.¹⁴² The result is that the Indian constitution’s declaration that India is a “sovereign, socialist, *secular*, democratic republic”¹⁴³ is read as including a commitment to enforcing “separate but equal” family law.¹⁴⁴

Ultimately, as this section’s (brief) discussion of India’s personal law system suggests, many people in India view family law pluralism as not only co-existing with the dignity of

¹⁴² For results of a poll of Muslim women which are consistent with this observation, see, for example, *supra* note 141. More generally, Madhavi Sunder has noted how

[i]ndividuals in the modern world [are] increasingly demand[ing] change *within* their religious communities in order to bring their faith in line with democratic norms and practices. Call this the New Enlightenment: Today, individuals [are] seek[ing] reason, equality, and liberty not just in the public sphere, but also in the private spheres of religion, culture, and family.

Madhavi Sunder, *Piercing the Veil*, 112 YALE L.J. 1399, 1403 (2003) (emphasis added). Finally, Kumkum Sangiri has remarked on the (perhaps) less-than-obvious patriarchal objectives of India’s ostensibly liberatory/secular state by noting that “[b]eneath the opposition between a state-imposed uniform civil code and personal laws that are sought to be reformed from ‘within’ a community . . . lies an unresolved but entirely patriarchal concern: who will control and regulate women . . .” Kumkum Sangiri, *Politics of Diversity: Religious Communities and Multiple Patriarchies*, ECON. & POL. WKLY. 3287, 3296 (1995).

¹⁴³ INDIA CONST. Preamble (emphasis added). The Constitution of India also includes a number of equality provisions. See, e.g., INDIA CONST. art. 14 (equality before law), art. 15 (sex equality), art. 16 (equality of opportunity in public employment), art. 17 (abolition of untouchability).

Many Indian feminists (and Hindu nationalists) have argued that the maintenance of different family laws for persons of different religious faiths is inconsistent with the Constitution (and its guarantees of religious and sexual equality). See generally FLAVIA AGNES, *LAW AND GENDER INEQUALITY: THE POLITICS OF WOMEN’S RIGHTS IN INDIA 192-202* (2000). However, many members of minority religious faiths have vociferously disagreed, basing their arguments on Article 26 of the Constitution, amongst other arguments. See generally *id.* at 100-23, 192-202. Article 26 guarantees that “[s]ubject to public order, morality and health, every religious denomination or any section thereof shall have the right . . . to manage its own affairs in matters of religion.” INDIA CONST. art. 26.

Both sides of this dispute utilize Article 44’s judicially-unenforceable plea for a “uniform civil code” in support of their constitutional and legal positions. See AGNES at 193. Article 44, part of the Constitution’s judicially-unenforceable “Directive Principles of State Policy,” reads as follows: “The State shall endeavour to secure for the citizens a uniform civil code throughout the territory of India.” INDIA CONST. art. 44.

¹⁴⁴ Members of the Bharatiya Janata Party (BJP), a Hindu nationalist party, would likely disagree. However, on this point, even these members would likely concede that their position that secularism is threatened by allowing minorities to be kept “separate and not equal” is a position swimming against the tide of history and practice. See WILLIAMS, *supra* note 132, at 171-72 (quoting a BJP publication and its use of “separate and not equal” phraseology).

minorities, but actually as somewhat of a *pre-requisite* for that dignity. Of course, there are intense disagreements among different members of any given minority group about the proper content of the family law that applies to the group.¹⁴⁵ Furthermore, these disagreements are often resolved at the expense of minority women. These (fortunate) debates and (unfortunate) abuses aside, the legal and political situation in many personal law systems nonetheless presents a very different take on the relationship between dignity and family law pluralism than that found in the California and Connecticut supreme courts' recent opinions. In these systems with communally-premised personal law systems, both refuge and dignity are found outside of the confines of majoritarian marriage and family law.

D. *Conclusion*

The California and Connecticut supreme courts viewed gay and lesbian dignity as *inextricably* bound up in formal equality and access to the (heterosexual) institution of majoritarian marriage. This account of dignity is not necessarily wrong,¹⁴⁶ but as the discussion in this Part (and Part I) has suggested, this account involves more assertion than analysis, and ignores the ways in which numerous people around the globe have felt that something other than mimicry of the majority creates a feeling of dignity in their lives. In this respect, religious people (amongst others) both inside and outside of the United States have attempted to exert *agency* over—and, hence, experience dignity with respect to—their family law.¹⁴⁷ In other words, these people have demonstrated how dignity inheres in being active authors of their law and, in this way, exercising both authority over and responsibility for this law.¹⁴⁸

¹⁴⁵ One might note that, at the very least, this intra-community disagreement is on full display in India, whereas legal and judicial discussions concerning same-sex marriage in the United States obscure and ignore debate within the gay and lesbian community about the desirability of marriage.

¹⁴⁶ See discussion *supra* note 68.

¹⁴⁷ For a more detailed discussion of how I am understanding and using the term “agency” in this Article, see *infra* Part III.

¹⁴⁸ This is the case even when (religious) people have sought particularized religious *exceptions* from otherwise generally-applicable law, as opposed to *affirmatively* drafting altogether-alternative legislation containing independently-authored norms. In this respect, arbitration of family law matters often involves *both*

The family law terrain in Canada, the United Kingdom, and India has been, and remains, contested. While great uncertainty exists in how these (and many other) states will ultimately resolve the competing interests and pressures present in these contemporary family law debates, what seems far more certain is that legal pluralism with respect to the state's regulation of the family will persist (and perhaps predominate) as a mode of contemporary governance. Majorities' intolerance of minorities threatens this pluralism, but this intolerance is also one of the major antecedents to the felt need for pluralism.

"Separate but equal" family law is thus here to stay, and arguments relating to minority rights, religious liberty, and human dignity will continue to support this kind of administration of family law, and also to put pressure on it. Ultimately, then, dignity is a much more complicated, contested, and dynamic concept than contemporary U.S. same-sex marriage advocates (including supportive courts) appear willing to acknowledge. Moreover, protecting gay and lesbian dignity may very well require something different than amalgamating gays and lesbians into a heterosexually-dominated majoritarian marriage regime, in which gays and lesbians will continually be democratically outmatched with respect to this regime's substantive content and norms. The next Part explores what a different approach to gay and lesbian dignity in the United States might look like.

III. SPECIAL RIGHTS, DIGNITY, AND THE FUTURE OF GAY AND LESBIAN RELATIONSHIP-RECOGNITION

Marriage is not the same thing as love. For their part, heterosexuals have shown us what marriage is worth and how long it lasts. . . . Rather than accept the narrowness under which heterosexuals themselves chafe, why not invite them to share in what we [homosexuals] know about the multiples ways in which relationships can form? If we come to heterosexuals and their institution, we valorize the mechanism of our oppression. Let them come to us.¹⁴⁹

—Steven K. Homer (1994)

exception from otherwise generally-applicable family law and the affirmative legislation of law (both procedural and substantive) to govern the issues at hand.

¹⁴⁹ Steven K. Homer, *Against Marriage*, 29 HARV. C.R.-C.L. L. REV. 504, 530 (1994).

At first I was calling it getting “civilized,” but that wasn’t going over very well. “Getting unionized,” that’s what the TV reporters are saying. Others are saying “united.”¹⁵⁰

—Jon Pominville
Middlebury, VT Town Clerk (2000)

Buried in the California Supreme Court’s decision in *In re Marriage Cases* was the following observation by the court about the need for same-sex couples to be able to “marry,” as opposed to enter into a (equally privileged) “domestic partnership”:

Because the constitutional right of privacy ordinarily would protect an individual from having to disclose his or her sexual orientation under circumstances in which that information is irrelevant, the existence of two separate family designations—one available only to opposite-sex couples and the other to same-sex couples—impinges upon this privacy interest, and may expose gay individuals to detrimental treatment by those who continue to harbor prejudices that have been rejected by California society at large.¹⁵¹

While playing a minor part in its overall decision, the court’s invocation of the proverbial “closet” to justify same-sex marriage rights is instructive more generally about some of the real injuries to gay and lesbian dignity, as well as other gay and lesbian interests, that the push for same-sex marriage is inflicting. And indeed, in addition to giving the closet a new door (and lock), gay and lesbian activists’ pursuit of same-sex marriage rights has resulted in a number of other curious tactics. These include 1) conveying to fellow (sexual) minorities that they are not welcome to join the struggle for gay and lesbian civil rights,¹⁵² 2) re-validating sexual shame and

¹⁵⁰ Carol Ness, *Couples Flock to Vermont, Only Legal Place to Get Hitched*, S.F. EXAMINER, Aug. 7, 2000, at A1 (quoting Jon Pominville, a town clerk in Middlebury, Vermont, on public confusion over how to refer to people getting a Vermont civil union).

¹⁵¹ *In re Marriage Cases*, 183 P.3d 384, 446 (Cal. 2008), *superseded by CAL. CONST. art. I, § 7.5*.

¹⁵² I am thinking here of how some same-sex marriage advocates have opposed the extension of rights that they are seeking to those wishing to enter polygamous marriages. For a sampling of academic literature that advocates this two-track approach to the right to marry, see, for example, Hema Chatlani, *In Defense of Marriage: Why Same-Sex Marriage Will Not Lead Us Down A Slippery Slope Toward the Legalization of Polygamy*, 6 APPALACHIAN J.L. 101 (2006) (arguing that the legalization of polygamous marriage would pose problems for social order and gender equality that same-sex marriage does not); Jaime M. Gher, *Polygamy and Same-Sex Marriage: Allies or Adversaries Within the Same-Sex Marriage Movement*, 14 WM. & MARY J. WOMEN & L. 559 (2008) (arguing that both polygamists and gays and lesbians have faced persecution in the U.S. but nonetheless suggesting, mostly for tactical

mockery as legitimate weapons in American political discourse,¹⁵³ and 3) running rough-shod over (intimate) logic and experience.¹⁵⁴

Some of these tactics have surely been born out of frustration, while others are the result of nakedly tactical considerations. With respect to tactical decisions about who to include in the movement, and who to exclude, given the difficult (if not dangerous) social climate in the United States with respect to gay and lesbian issues, it is not surprising that gay and lesbian activism has tried to weave a path of least resistance for itself, distancing itself publicly from politically unpopular allies in an attempt to mimic the majority.

reasons, that same-sex marriage activists distance themselves from pro-polygamy activists); Elizabeth Larcano, A "Pink" Herring: *The Prospect of Polygamy Following the Legalization of Same-Sex Marriage*, 38 CONN. L. REV. 1065, 1067 (2006) (arguing that there are a large number of differences both between polygamous and same-sex unions, and between polygamist and gay and lesbian persons); Maura I. Strassberg, *Distinctions of Form or Substance: Monogamy, Polygamy and Same-Sex Marriage*, 75 N.C. L. REV. 1501, 1617 (1997) (arguing that polygamous marriages are patriarchal while same-sex marriages are not).

¹⁵³ I am thinking here of the gay and lesbian protests which erupted around the country in late 2008 in the wake of the passage of Proposition 8 in California. See, e.g., *Gay-Marriage Rally Held at NYC Mormon Temple*, Associated Press, Nov. 12, 2008, available at <http://www.newsday.com/news/new-york/gay-marriage-rally-held-at-nyc-mormon-temple-1.885717?qr=1>; *Prop 8 Protest in New York City*, TOWLEROAD, Nov. 9, 2008, available at <http://www.towleroad.com/2008/11/prop-8-protes-1.html>; Chris Rovzar, *Gays Turn Anger, Snappy Sarcasm Toward Mormon Church*, DAILY INTEL, Nov. 13, 2008, available at http://nymag.com/daily/intel/2008/11/gays_turn_anger_snappy_sarcasm.html. A constant refrain at these protests involved the willful distortion and mocking disrespect of religious (and, most notably, Mormon) beliefs and practices. Some signs at these protests contained the following slogans and statements: "You want three wives, I want one husband," "I Don't Need 5 Wives Just 1 Husband," and "Keep Your Magic Undies Off My Civil Rights." For photos of signs at Proposition 8 protests outside of Mormon temple and elsewhere, see http://www.nbclosangeles.com/news/local/Prop_8_Protestors_March_LA_Streets.html (last visited Feb. 11, 2010), http://lh6.ggpht.com/_780ZZpC_ZNU/SRgdL1yCBwI/AAAAAAAAAjs/nIvwa8j4u4w/s400/Not5WivesCropx390.jpg (last visited Feb. 11, 2010), <http://www.utne.com/uploaded/Images/utne/blogs/Spirituality/Prop8protest.jpg> (last visited Feb. 11, 2010). For a statement from the Mormon religion's leadership firmly disavowing polygamy, see The Church of Jesus Christ of Latter-Day Saints, *Polygamy: Latter-day Saints and the Practice of Plural Marriage*, available at <http://www.newsroom.lds.org/ldsnewsroom/eng/background-information/polygamy-latter-day-saints-and-the-practice-of-plural-marriage> (last visited Feb. 11, 2010).

¹⁵⁴ For example, arguing for same-sex marriage rights, one attorney has remarked: "I used to say, 'Why do we want to get married? It doesn't work for straight people' But now I say we should care: They have the privilege of divorce and we don't. We're left out there to twirl around in pain." Kirk Johnson, *Gay Divorce: Few Markers in This Realm*, N.Y. TIMES, Aug. 12, 1994 (quoting Margaret M. Cassella). While for some people there might be something glamorous, and hence desirable, about the figure of the divorcée, I believe it is rather doubtful to argue that gays and lesbians need marriage because they want divorce. This would seem to be a case of putting the cart before the horse.

Nonetheless, such political and legal strategies are compromising gay and lesbian dignity, as are strategies that seek alignment with institutions that have been and will remain captive to majoritarian interests, i.e. institutions where gays and lesbians will be unable to exercise much agency with respect to laws and policies that will directly affect gay and lesbian lives and well-being. Marriage is one such majoritarian institution.

This Part aims to sketch a vision for gay and lesbian dignity that is different than the coercive one articulated by the California and Connecticut supreme courts, and also by leading gay and lesbian advocacy organizations which are attempting to legalize same-sex marriage. This alternative, and arguably more robust, vision of gay and lesbian dignity is one which is informed by the comparative experience discussed in Part II. It is also one that is informed by a close reading (below) of a desire expressed by many ordinary gays and lesbians post-Proposition 8, namely a desire for more *agency* with respect to laws and policies affecting gay and lesbian lives. Ultimately, the vision of dignity sketched here is one which cooperates neither with any homophobic desire to socially erase gay and lesbian existence, nor homophobic efforts to force gays and lesbians to conform with heterosexually-authored codes of behavior.

This Part will repeatedly invoke the idea of “agency,” so a few words of how this term is being used are in order. What constitutes agency is, obviously, a difficult question, which requires more discussion than space here permits. Briefly, however, this Part understands the existence of (individual or collective) “agency” to mean the ability of persons to engage in a complicated “calculus of action”¹⁵⁵ directed toward their “self-realization/self-fulfillment.”¹⁵⁶ However, this Part does not employ the term as a synonym for (personal or communal) “autonomy,” or any simplistic notion of (personal or communal) “sovereignty.” Similarly, this Part does not mean to equate “agency” with simplistic notions of “freedom” or “choice” or otherwise suggest that agency implies a socially and culturally unfettered ability to pick and choose with abandon what one desires in life. Such freedom (of choice) does not exist in this

¹⁵⁵ See PERVEEZ MODY, *THE INTIMATE STATE: LOVE-MARRIAGE AND THE LAW IN DELHI* 193 (2008).

¹⁵⁶ SABA MAHMOOD, *POLITICS OF PIETY: THE ISLAMIC REVIVAL AND THE FEMINIST SUBJECT* 13 (2005).

life. Instead, agency is more about the *authorship* of one's (individual or collective) path, given the opportunities, obstacles, language, and grammar that one's social, cultural, and political contexts continually (and somewhat unpredictably) provide.

Indeed, the "separate" system of gay and lesbian family law that this Part argues for—and which this Part explicitly links to the idea of gay and lesbian agency—will have to, under existing governmental structures, come into force through legislation passed by heterosexually-dominated state legislatures. This is unavoidable. That being said, the thought here is that gays and lesbians have the very real possibility of exercising a certain kind of political ownership over "domestic partnerships," "civil unions," or other forms of gay and lesbian relationship-recognition that any given state legislature might create. With this gay and lesbian ownership, significant gay and lesbian authorship of gay and lesbian law could follow—perhaps informed by practice elsewhere, such as India, where the national Parliament is responsible for legislating and otherwise enabling (the bulk of) religious communities' personal law, and these same communities have been able to exercise a great deal of say with respect to this legislation.¹⁵⁷ This would be agency, as this Part understands and uses this idea.

More specifically, this would be *American agency*, and indeed this Part does not understand or use "agency" in a way that is de-linked from local context, which includes local imaginations of the possible. With respect to these local imaginations, in some contexts—including perhaps the contemporary United States—"agency" is entailed not only in those acts that resist norms but also in the multiple ways in which one *inhabits* norms.¹⁵⁸ In other words, in some contexts, agency exists where one finds "*submission* to certain forms of (external) authority."¹⁵⁹ This being the case, this Part does not insist that gay and lesbian agency find expression in a system of relationship-recognition and family law that is completely different than majoritarian marriage and majoritarian family law. The "separate" system of gay and lesbian family law that this Part (following extant Californian practice) suggests, and

¹⁵⁷ See Galanter & Krishnan, *supra* note 121, at 109 (2000). *But see* WILLIAMS, *supra* note 132, at 98-99.

¹⁵⁸ MAHMOOD, *supra* note 156, at 15.

¹⁵⁹ *Id.* at 31 (emphasis added).

begins to sketch, is not intended to be different for difference's sake. In fact, no pluralist system of law anywhere in the world functions in this facile way. Instead, the separate system of family law that this Part suggests is intended to provide a space¹⁶⁰ from which gays and lesbians can argue for, and implement, a different set of norms than majoritarian ones *if and when* differences with the majority arise. Preserving this potential for difference is important for gay and lesbian agency and—as this Article understands the relationship between agency and dignity—gay and lesbian *dignity*.

This Part proceeds in three sections. The first section shows, very generally, how one might see Proposition 8 in a positive light, by demonstrating how this allegedly anti-gay ballot initiative resulted in something that anti-gay activists have feared and railed against for some time now, namely “special rights” for gays and lesbians.¹⁶¹ Indeed, the fact that Proposition 8 effectively resulted in a successful initiative for gay and lesbian “special rights” signifies an important reworking of anti-gay activists’ political and legal agendas. Consequently, gay and lesbian activists would be remiss in not grappling with—and capitalizing on—this important shift in the legal and political terrain and the unprecedented opportunities for gay and lesbian agency (and dignity) that have opened up as a result of Proposition 8.

Building on the first section’s re-reading of Proposition 8 through the lens of “what your enemies do *not* want for you might very well be what you *should* want,” the second section of this Part begins to provide a more affirmative account of how parallel relationship-recognition regimes defend important gay and lesbian interests. With respect to these interests, this section first engages seriously with what ordinary gays and lesbians expressed about their needs after Proposition 8. As this section reads those needs, they included *more agency* vis-à-vis the laws that govern gay and lesbian lives and families.

¹⁶⁰ Jim Bohman argues in a similar vein when he writes that “sometimes separate jurisdictions can serve a public function, to the extent that they provide the public *space* needed for groups like Native Americans to have a more coherent and effective voice in the larger, civic public sphere.” James Bohman, *The Moral Costs of Political Pluralism: The Dilemmas of Difference and Equality in Arendt’s “Reflections on Little Rock”*, in HANNAH ARENDT: TWENTY YEARS LATER 53, 73 (Larry May & Jerome Kohn eds., 1996) (emphasis added).

¹⁶¹ For a history of the anti-gay use of the “special rights” terminology, see generally TINA FETNER, *HOW THE RELIGIOUS RIGHT SHAPED LESBIAN AND GAY ACTIVISM* 84-100 (2008).

After diagnosing this gay and lesbian desire for more agency with respect to the laws affecting gay and lesbian lives, this section moves on to diagnose and discuss in detail how this agency is threatened by gay and lesbian amalgamation into existing majoritarian marriage regimes in the United States. Like Part II's discussion of the legal and political experiences of minorities around the globe and the desire for more agency that these experiences have given rise to, the discussion in this section similarly provides examples of (heterosexual) majoritarian indifference and/or hostility in the contemporary United States (as well as just across the border in Canada). Accordingly, in the same way that global minorities have sought both dignity and refuge via pluralist legal set-ups, this section suggests that such set-ups might serve as helpful templates for gay and lesbian action, and dignity, in the contemporary American same-sex marriage debates. In other words, this section argues that gay and lesbian relationship-recognition politics in the United States should be a great deal less sanguine about the dignity that majoritarian marriage slyly promises, and why American gay and lesbian politics should be more receptive to learning from the politics and practices of legal pluralism elsewhere.

The final section of this Part concludes by offering two specific suggestions of how gay and lesbian non-majoritarian relationship-recognition regimes might offer different—and better—alternatives to those provided by majoritarian marriage. While, as this Part has already discussed, it is not at all necessary for gay and lesbian agency that gay and lesbian relationship-recognition regimes be entirely different than majoritarian marriage, there are some distinct ways in which domestic partnership/civil union regimes might be structured in order to better demonstrate their distinct worth. In making these two particular suggestions—one concerning the nomenclature of gay and lesbian relationship-recognition regimes, and the other concerning the substance of such regimes—this concluding section will also be able to respond to two major concerns that contemporary same-sex marriage advocates will likely have about this Part's arguments and proposals. These concerns are: 1) the alleged inability of the "domestic partnership" or "civil union" nomenclature to provide anything more than an inferior and insulting neologism in the face of the magical-realism of the word "marriage," and 2) the restrictions on "choice" that are implicit in creating separate

relationship recognition regimes that are each available *only* to certain type of couples (e.g. same-sex versus opposite-sex).

The conventional terrain over which the same-sex marriage debates are transpiring is creating serious impediments to gay and lesbian dignity. As a result, this Part attempts to re-conceptualize and reframe basic terms of reference in the same-sex marriage debates in order to advance gay and lesbian dignity. As this Part demonstrates, what looks like homophilia can very much be homophobia, and what looks homophobic can prove homophilic. That being said, this Part focuses less on the conceptual, legal, and political missteps of advocates for same-sex marriage—the homophobia of their brand of homophilia—than it does on the homophilia in others’ homophobia. At one level, then, the goal of this Part is to find homophilic opportunity in some of the ironies that have been opened up in a world where (ostensibly) homophobic initiatives, such as Proposition 8, are par for the course. More particularly, this Part means to demonstrate how same-sex “domestic partnerships” or “civil unions”—separate from opposite-sex “marriage”—can be dignity-enhancing for gays and lesbians. Indeed, they are not “separate but equal” institutions, but potentially “separate *and better*” ones.

A. *Special Rights and the Anti-Homophobic Promise of Proposition 8*

To begin to see how measures like Proposition 8 (and the plural relationship-recognition system that it returned California to) might align with gay and lesbian interests, because of the way that this measure resulted in *special* recognition of same-sex relationships, one need only examine a provision attached (ironically) to recent legislation banning discrimination on the basis of sexual orientation in the State of Connecticut. According to this 2005 addition to the General Statutes of Connecticut, nothing contained in the Connecticut anti-discrimination legislation shall be

deemed or construed (1) to mean the state of Connecticut condones homosexuality or bisexuality or any equivalent lifestyle, (2) to authorize the promotion of homosexuality or bisexuality in educational institutions or require the teaching in educational institutions of homosexuality or bisexuality as an acceptable lifestyle, (3) to authorize or permit the use of numerical goals or quotas, or other types of affirmative action programs, with respect to homosexuality or bisexuality in the administration or enforcement of the [state’s antidiscrimination laws], (4) to authorize the recognition

of or the right of marriage between persons of the same sex, or (5) to establish sexual orientation as a specific and separate cultural classification in society.¹⁶²

The incredible fear that homosexuality might gain social credence as either a lifestyle or recognized cultural group is palpable in this recent legislative declaration.

The fear that gays and lesbians might find benefit from or even want “special rights” is older than this recent Connecticut legislation might indicate. Indeed, before there was this Connecticut law (and before there was Proposition 8), there was Amendment 2, the infamous 1992 amendment to the Colorado State Constitution that declared that

[n]either the State of Colorado, through any of its branches or departments, nor any of its agencies, political subdivisions, municipalities or school districts, shall enact, adopt or enforce any statute, regulation, ordinance or policy whereby homosexual, lesbian or bisexual orientation, conduct, practices or relationships shall constitute or otherwise be the basis of or entitle any person or class of persons to have or claim any minority status, quota preferences, protected status or claim of discrimination.¹⁶³

As is well known, Amendment 2 was challenged using the federal constitution in the U.S. Supreme Court, the result of which was the landmark *Romer v. Evans* decision.¹⁶⁴ The terrain over which the legality of Amendment 2 was fought, both inside and outside of the Supreme Court, concerned whether Amendment 2 was an appropriate response to the supposed menace of “special rights” for gays and lesbians (and bisexuals). As the Supreme Court described it, “[Colorado’s] principal argument in defense of Amendment 2 is that it puts gays and lesbians in the same position as all other persons. So, the State says, the measure does no more than deny homosexuals special rights.”¹⁶⁵

¹⁶² CONN. GEN. STAT. § 46A-81R (2005), *repealed by* R.B. 899, 2009 Gen. Assem., Jan. Sess. (Conn. 2009) (implementing the Connecticut Supreme Court’s *Kerrigan v. Comm’r of Pub. Health* decision, recognizing marriages and relationships providing “substantially the same rights, benefits, and responsibilities entered into in another state or jurisdiction,” and providing for the merger of “existing civil unions into marriages” in Connecticut).

¹⁶³ COLO. CONST. art. 2 § 30(b), *invalidated by* *Romer v. Evans*, 517 U.S. 620 (1996).

¹⁶⁴ *Romer v. Evans*, 517 U.S. 620 (1996).

¹⁶⁵ *Id.* at 626; see also Justice Scalia’s dissenting opinion, in which he writes:

[A]ssuming that, in Amendment 2, a person of homosexual ‘orientation’ is someone who does not engage in homosexual conduct but merely has a tendency or desire to do so, *Bowers* still suffices to establish a rational basis

In its opinion, the Supreme Court disagreed with the State of Colorado, holding that Amendment 2 violated the Equal Protection Clause of the U.S. Constitution.¹⁶⁶ In the process, the Court also found that it was not gay and lesbian people who were seeking legal peculiarity in Colorado, but the proponents of Amendment 2 themselves. Wrote the Court:

[T]he amendment imposes a special disability upon [homosexual] persons alone. Homosexuals are forbidden the safeguards that others enjoy or may seek without constraint. They can obtain specific protection against discrimination only by enlisting the citizenry of Colorado to amend the State Constitution or perhaps, on the State's view, by trying to pass helpful laws of general applicability. This is so no matter how local or discrete the harm, no matter how public and widespread the injury. We find nothing special in the protections Amendment 2 withholds.¹⁶⁷

The Connecticut legislature's recent efforts to pre-empt an (alleged) gay and lesbian effort to be viewed as "special" is just the latest installment, then, in what has been a recurring theme in anti-gay polemics in the United States. Similarly, same-sex "marriage" can be viewed as the latest instance of gay and lesbian advocates explicitly (and fearfully) rejecting any mark of special-ness or distinction. Given the history of majoritarian pillorying of gays and lesbians for their allegedly constant attempts to seek special legal accommodation, Proposition 8's creation of (or return to) a special relationship-recognition regime for same-sex couples is extremely noteworthy. Indeed, given anti-gay fears of how gays and lesbians might fruitfully capitalize upon any sort of potential special recognition by the law, the fact that Proposition 8 and other measures actually create "special" parallel relationship-recognition regimes for gay and lesbian persons deserves closer scrutiny and appreciation from advocates for gays and lesbians. Now may very likely be the time to re-examine the typical gay and lesbian urge to retreat into the majority.

Of course, this will not be easy for such advocates, given the particular course that anti-gay stigmatization has taken in

for the provision. If it is rational to criminalize the [homosexual] conduct [according to our *Bowers* precedent], surely it is rational to deny *special* favor and protection to those with a self-avowed tendency or desire to engage in the conduct.

Id. at 642 (Scalia, J., dissenting) (emphasis added).

¹⁶⁶ *Id.* at 635.

¹⁶⁷ *Id.* at 631.

the United States for so long. For gays and lesbians in the United States, there have long been negative consequences associated with the claim that gays and lesbians seek “special” and unique privileges in an otherwise egalitarian America, and also with the corollary description of homosexuality as mere “lifestyle”¹⁶⁸—the same “lifestyle” that the “rich and famous” always already enjoy.¹⁶⁹ In response to this particular brand of anti-gay baiting, gay and lesbian advocates have typically fled from anything associated with either term.¹⁷⁰ However, in the

¹⁶⁸ This pejorative use of “lifestyle” can be found in many places including, as Douglas NeJaime has documented, the educational context. See Douglas NeJaime, *Inclusion, Accommodation, and Recognition: Accounting for Differences Based on Religion and Sexual Orientation*, 32 HARV. J.L. & GENDER 303, n.139 (2009); see also ALA. CODE § 16-40A-2(c)(8) (LexisNexis 1992) (requiring sex education program materials to “emphasi[ze] . . . in a factual manner and from a public health perspective, that homosexuality is not a lifestyle acceptable to the general public and that homosexual conduct is a criminal offense under the laws of the state”); ARIZ. REV. STAT. ANN. § 15-716(c)(1)-(3) (1995) (prohibiting instruction that (1) “[p]romotes a homosexual life-style,” or (2) “[p]ortrays homosexuality as a positive alternative life-style”); S.C. CODE ANN. § 59-32-30(A)(5) (2004) (prohibiting health education programs from discussing “alternate sexual lifestyles from heterosexual relationships including, but not limited to, homosexual relationships except in the context of instruction concerning sexually transmitted diseases”); TEX. HEALTH & SAFETY CODE ANN. § 85.007 (Vernon 1999) (requiring education programs for persons eighteen-years-old and younger to “state that homosexual conduct is not an acceptable lifestyle and is a criminal offense”).

¹⁶⁹ See, e.g., *Romer*, 517 U.S. at 645-46 (Justice Scalia’s finding that homosexuals have “high disposable income”).

¹⁷⁰ For example, in the *Romer* litigation, it became *everyone’s* objective in the litigation to flaunt their mundane, “un-special” credentials. Justice Kennedy’s majority opinion, for example, found that

Amendment 2 confounds th[e] normal process of judicial review. It is at once too narrow and too broad. It identifies persons by a single trait and then denies them protection across the board. The resulting disqualification of a class of persons from the right to seek specific protection from the law is unprecedented in our jurisprudence. The absence of precedent for Amendment 2 is itself instructive It is not within our constitutional tradition to enact laws of this sort.

Id. at 633. Justice Scalia’s minority, dissenting opinion argued the elite nature of both American homosexuals and their supporters:

It is . . . nothing short of preposterous to call “politically unpopular” a group [e.g. homosexuals] which enjoys enormous influence in American media and politics. . . . When the Court takes sides in the culture wars, it tends to be with the knights rather than the villeins—and more specifically with the Templars, reflecting the views and values of the lawyer class from which the Court’s Members are drawn.

Id. at 652.

As a result, whatever victory for gay and lesbian people that *Romer’s* outcome represented, the opinion’s silences and lapses also tell a story of equally-important missed opportunities. Examining the history which led up to this state constitutional amendment, as well as the Supreme Court’s particular focus in this case, one finds the entire legal battle centered around the question of whether Amendment

process, they have arguably stymied consideration that they are akin to other sorts of “cultural” groups that might benefit from multiculturalist policies and the “protected status” (or even “quotas”¹⁷¹) that forms of multiculturalism can distribute to cultural groups. As a result of this pressure to culturally dissolve, and also politically disassociate from controversial social “re-engineering” plans, gay and lesbian activists have found it difficult to ask for (or even imagine the possibility of) *strong* remedies for discrimination that have been implemented (however unevenly or ineffectively) with respect to other discriminated-against groups.¹⁷² Affirmative action, for example, is one such remedy, and the surprise and debate—both within and without the gay and lesbian community—that greeted Middlebury College’s 2006 announcement (later disavowed) that it would affirmatively act to admit openly-homosexual students is but one example of this.¹⁷³

2’s eradication of Colorado municipal non-discrimination statutes (amongst other measures put into place in Colorado ensuring non-discrimination on the basis of sexual orientation) marked an end to “special-ness” and a return to equality, or whether this outcome itself created a state of exception and marked some people as legal “outlaws.” In other words, only Amendment 2’s silencing of gay and lesbian people’s “claims of discrimination” was dealt with in the case; the other parts of Amendment 2 which envisioned the possibility of giving gay and lesbian people “minority status, quota preferences, [and/or] protected status” were completely ignored. Indeed, most fundamentally, the debate in the case failed to ask, “What’s wrong with being ‘special?’”

¹⁷¹ See text accompanying *supra* notes 162 and 163.

¹⁷² This is not to say that gays and lesbians are necessarily in the *same* position as discriminated-against racial and ethnic minorities in the United States, nor that gays and lesbians should imbricate themselves in all of the tropes and technologies relating to countering racial and ethnic discrimination (e.g. “separate but equal”) in the United States, but it is to say that “despite the adoption of a goal of civil rights, gay collective identity is at present closer in form to that of the white ethnic groups than to those of racial minorities. Movement away from a political consciousness based on white ‘ethnicity’ . . . might increase the gay movement’s capacity to pose a more fundamental challenge to the socio-sexual order.” Steven Epstein, *Gay Politics, Ethnic Identity: The Limits of Social Constructionism*, in *FORMS OF DESIRE: SEXUAL ORIENTATION AND THE SOCIAL CONSTRUCTIONIST CONTROVERSY* 239, 291 (Edward Stein ed., 1990).

¹⁷³ See Heather Schwedel, *Pondering Affirmative Action for Gays*, *THE DAILY PENNSYLVANIAN*, Oct. 30, 2006, available at <http://media.www.dailypennsylvanian.com/media/storage/paper882/news/2006/10/30/News/Pondering.Affirmative.Action.For.Gays-2409198.shtml>, for an example of the confusion and debate that accompanied the supposed Middlebury College announcement; see also John Calapinto, *The Harvey Milk School Has No Right to Exist. Discuss*, *N.Y. MAG.*, May 21, 2005, available at <http://nymag.com/nymetro/news/features/10970/> (discussing liberal unease with educational admission policies that give preferential treatment to gay and lesbian students). See generally David Luc Nguyen, *Taking Affirmative Action: Do Gays Deserve the Same Boost Into College as Racial Minorities?*, Jan. 30, 2007, available at <http://www.thefreelibrary.com/Taking+affirmative+action:+do+gays+deserve+the+same+boost+into...-a0159593303>.

Thus, as much as the contemporary gay and lesbian civil rights movement links itself to the civil rights struggles of before, an important disjuncture emerges with respect to the issue of group identity and group cohesiveness. This has important ramifications for the question of what to ask for legally and politically. The next section argues that gays and lesbians should not abandon the prospect of “special rights,” and the legal agency they can result in, especially where an unlikely opportunity to get both has finally presented itself in the form of domestic partnerships, civil unions, and the like.

B. The Possibility of Claiming Special Rights and Dignity

Gay and lesbian advocates’ fear of “lifestyle” and “special rights” allegations is real.¹⁷⁴ However, this understandable fear need not be paralyzing. And, indeed, many ordinary gay and lesbian people viewed Proposition 8 not as a paralytic, total defeat, but as a spur for action. This section first demonstrates how gay and lesbian people in the United States, like other people around the globe, have recently been arguing for a great deal more agency vis-à-vis the laws that directly impact their lives and families. It then proceeds to show how the amalgamation of gays and lesbians into majoritarian marriage regimes threatens this agency. This discussion sets the stage for the next section’s exploration of how a more legally-pluralistic relationship-recognition system provides for gay and lesbian agency—and dignity—in ways that gay and lesbian advocates’ pursuit of majoritarian marriage has not, and cannot.

In the aftermath of the Proposition 8 vote, many gays and lesbians expressed the feeling that the vote left them feeling powerless with respect to their destiny, in at least three different ways. First, many gay and lesbian Californians lamented the control that *non-Californian, out-of-state* forces seemed to have over the outcome of the vote. For example, Lorri Jean, CEO of the Los Angeles Gay and Lesbian Center publicly stated:

We have been critical of all of the *out-of-state* conservative religious groups that made significant contributions to the campaign, including the Knights of Columbus National Headquarters in Connecticut and Focus on the Family in Colorado. But the truth is that the LDS church leadership in Utah specifically directed its

¹⁷⁴ See, e.g., Cruz, *supra* note 15, for an articulation of this common fear.

membership to get involved with the Yes campaign in an unprecedented way—both in terms of volunteer time and dollars.¹⁷⁵

Second, as this statement by Jean simultaneously reveals, many gay and lesbian people felt that Proposition 8's passage demonstrated how *religious* groups were dictating the laws of a *secular* state, which many gay and lesbian Americans clearly feel an especially strong (if secular) attachment to.¹⁷⁶ Finally, and similarly, there were many gay and lesbian laments that the civil rights of a *minority* should not be dictated by the votes of a *majority*.¹⁷⁷

Agency, then, has been an important issue for many ordinary gay and lesbian people, even if it has been neglected by lawyers, judges, and academics in their discussions of the same-sex marriage issue. Taking this concern for gay and lesbian agency seriously, the rest of this section will highlight the ways in which a unitary relationship-recognition system

¹⁷⁵ Lorri L. Jean, *No on Proposition 8 Frequently Asked Questions*, http://laglc.convio.net/site/PageServer?pagename=Prop_8_FAQ (last visited Jun. 12, 2009) (emphasis added).

¹⁷⁶ See, e.g., AMERICANS UNITED FOR SEPARATION OF CHURCH AND STATE, THE RELIGIOUS RIGHT'S WAR ON LGBT AMERICANS: CHURCH, STATE, AND YOUR FREEDOM AT RISK 1 (noting that Rev. Barry W. Lynn, executive director of Americans United for Separation of Church and State, had previously commented on Proposition 8 by stating: "Allowing powerful religious groups to take away minority rights by referendum is fundamentally at odds with what America is about."), available at <http://www.au.org/resources/brochures/the-religious-rights-war-on-lgbt-americans/lgbt-2009.pdf> (last visited Jan. 15, 2010). For a more quotidian example of this sentiment, see Linda Morgan, *Letter to the Editor, Church and State*, S.F. CHRON., Oct. 25, 2008, at B4 (arguing that "[t]he separation of church and state is meant to prevent the use of state power to enforce the religious views of any particular group on society as a whole. It is, in fact, the proponents of Proposition 8 who are seeking to compel all of us to abide by their vision of right and wrong.").

¹⁷⁷ See, e.g., Jennifer Harper, *Inside the Beltway*, WASH. TIMES, Nov. 6, 2009, at A7 for a quote by Geoff Kors, executive director of Equality California, stating his belief that "people's lives should never be put up for a popular vote. Civil rights for minority groups should be decided by the sound reason of the legislature and the courts—not by the will and whims of the majority."; see also Frank Rich, Op-Ed., *The Bigots' Last Hurrah*, N.Y. TIMES, Apr. 19, 2009, § WK, at 10, for this acerbic commentary:

Some [same-sex marriage] opponents grumbled anyway [after the Iowa Supreme Court decision legalizing same-sex marriage], reviving their perennial complaint, dating back to *Brown v. Board of Education*, about activist judges. But the judiciary has long played a leading role in sticking up for the civil rights of minorities so they're not held hostage to a majority vote.

Finally, Stuart Milk, nephew of Harvey Milk, has recently proclaimed that "[t]aking away a civil right we had is a violent act. . . . As Harvey would say, when you let the majority deprive the minority of their civil rights, you start a shopping list. . . . Who is next?" See Meredith May, *Rally in Castro on Eve of Prop. 8 Hearing*, S.F. CHRON., Mar. 5, 2009, at B1.

threatens this agency, in order to set the stage for the next section's specific (yet preliminary) suggestions for how a pluralist system might do things (somewhat) differently and (very likely) better.

Again, the best place to begin to understand how gay and lesbian agency is threatened by a unitary marriage regime for one-and-all is the California Supreme Court's recent same-sex marriage decision. In this respect, the California Supreme Court, while discussing the nomenclature politics of relationship-recognition, opined that

because of the long and celebrated history of the term "marriage" and the widespread understanding that this word describes a family relationship unreservedly sanctioned by the community, the statutory provisions that continue to limit access to this designation exclusively to opposite-sex couples—while providing only a novel, alternative institution for same-sex couples—likely will be viewed as an official statement that the family relationship of same-sex couples is not of comparable stature or equal dignity to the family relationship of opposite-sex couples.¹⁷⁸

Distilling the California Supreme Court's opinion here, then, one learns that, in California, there is apparently one community ("the community"), which for a long time has "unreservedly" endorsed an unchanging, universally understood (i.e., "well-understood") institution known as "marriage."

One might worry that the monolithic, "transcendent"¹⁷⁹ vision of marriage painted by the California Supreme Court here—and, later, by the Connecticut Supreme Court¹⁸⁰—is a decidedly un-secular one. Not only is there an undeniable shade of sectarian monotheism coloring this vision of marriage—the single, indivisible god here being "marriage" itself—but also, at times, an outright religiosity presents itself in these opinions. Discussing the nature of marriage, for example, the Connecticut Supreme Court wrote:

[T]he following observation of Connecticut Catholic Conference, Inc., which filed an amicus brief in support of the defendants, is relevant. "In our culture, there has been a consensus on . . . [the] unique ethical foundations [of marriage]: that the union should be for life (permanency), that the union should be exclusive (fidelity), and that

¹⁷⁸ *In re Marriage Cases*, 183 P.3d 384, 452 (Cal. 2008), *superseded by* CAL. CONST. art. I, § 7.5.

¹⁷⁹ *See* Kerrigan, 957 A.2d at 418.

¹⁸⁰ *Id.*

the love that sustains and nurtures the union should be characterized by mutual support and self-sacrifice (selflessness).” These ideals apply equally to committed same sex and committed opposite sex couples who wish to marry.¹⁸¹

Thus, here one finds a secular court quoting a religious brief in support of an antiquated vision of (heterosexual) marriage.¹⁸² It would seem to be a small step between this kind of religious influence on secular marriage and the type of religious influence on secular government that many gays and lesbians protested in the aftermath of Proposition 8.¹⁸³ To the extent that one is worried about gay and lesbian agency in one context, one might also be worried about it in the other.

¹⁸¹ *Id.* at n.76 (quoting Brief of Connecticut Catholic Conference, Inc. as Amicus Curiae in Support of Defendant-Appellees at 11, *Kerrigan v. Comm’r of Pub. Health*, 957 A.2d 407 (Conn. 2007) (No. 17716)). Of course, this is also a sectarian observation. With respect to divorce, there has been and remains intense disagreement between Catholics and Protestants over the availability of religious divorce, and both Christian traditions have serious objections with aspects of Muslim divorce law.

¹⁸² It should be noted that once one puts this Connecticut opinion side-by-side with the California Supreme Court’s opinion, one has two high courts describing marriage in a way that appears as monolithic and impervious to change as the description of marriage put forward by advocates working to keep the institution heterosexual. When this latter set of advocates cite “the historic and well-established nature of [the opposite-sex] limitation [for marriage] and the circumstance that the designation of marriage continues to apply only to a relationship between opposite-sex couples in the overwhelming majority of jurisdictions in the United States and around the world,” *In re Marriage Cases*, 183 P.3d at 450, they do so in order to accuse their adversaries of trying to “redefine” the (single possible) definition of “marriage.” *Id.* at 470 (Corrigan, J., concurring and dissenting). For both (ostensibly pro-gay) advocates and (anti-gay) opponents of same-sex marriage, then, there can only be one type of marriage for “the” single community that supposedly comprises the polity. Given these (unnecessarily-inflated) stakes, one can perhaps better appreciate the intensity of the conflict between the two sides.

See also these additional comments by Justice Baxter:

The bans on incestuous and polygamous marriages are ancient and deep-rooted, and, as the majority suggests, they are supported by strong considerations of social policy. *Our society* abhors such relationships, and the notion that our laws could not forever prohibit them seems preposterous. Yet here, the majority overturns, in abrupt fashion, an initiative statute confirming the equally deep-rooted assumption that marriage is a union of partners of the opposite sex. The majority does so by relying on its own assessment of contemporary community values, and by inserting in our Constitution an expanded definition of the right to marry that contravenes express statutory law.

Id. at 463 (Baxter, J., concurring and dissenting) (emphasis added). Of course, this statement ignores the fact that “incestuous” is a notoriously difficult term to define, and may or may not include first-cousin marriages. Given this reality, and the fact that there are surely people who are California citizens who, for religious or secular reasons, believe in polygamy (and “incest”), the assertion here of one society—“our society”—is truly a hegemonic move.

¹⁸³ See *supra* note 176.

Similar concerns about the possibility of gay and lesbian agency vis-à-vis majoritarian marriage can be raised in the aftermath of another same-sex marriage judicial decision, though one that did not uphold same-sex marriage rights. Specifically, in a recent (2006) opinion, *Hernandez v. Robles*, New York's highest court argued the existence of a persisting connection between marriage and hetero-sex. Explaining its decision to uphold the traditional legal definition of marriage in that state, the New York court emphasized that marriage was for heterosexuals, and heterosexuals only, because "[h]eterosexual intercourse has a natural tendency to lead to the birth of children; homosexual intercourse does not. . . . [Same-sex] couples can become parents by adoption, or by artificial insemination or other technological marvels, but they do not become parents as a result of accident or impulse."¹⁸⁴ In other words, marriage is important as a social prophylactic when the condom breaks.

This judicial decision, and ones like it,¹⁸⁵ is indicative of the hold that majoritarian (heterosexual) concerns and priorities presently have, and will likely maintain, over the institution of marriage in the United States.

While the New York court ultimately used these majoritarian concerns and priorities to deny gay and lesbian access to the institution of marriage, it seems likely that such majoritarian concerns will motivate the future direction (including potential regression) of marriage *even if* gays and lesbians are allowed to "marry" the intimate partner of their choice.

If such a concern seems preposterous, one only has to examine what happened in Canada *after* the introduction of same-sex marriage rights there in 2005. In two recent cases,¹⁸⁶

¹⁸⁴ *Hernandez v. Robles*, 7 N.Y.3d 338, 359, 855 N.E.2d 1, 7 (2006). See generally Kenji Yoshino, Op-Ed., *Too Good for Marriage*, N.Y. TIMES, Jul. 14, 2006, available at <http://www.nytimes.com/2006/07/14/opinion/14yoshino.html>.

¹⁸⁵ See *Standhardt v. Superior Court ex rel. County of Maricopa*, 77 P.3d 451 (Ariz. 2003); *Morrison v. Sadler*, 821 N.E.2d 15, 24-26 (Ind. Ct. App. 2005); *Conaway v. Deane*, 932 A.2d 571 (Md. 2007); *Lewis v. Harris*, 908 A.2d 196, 216-17 (N.J. 2005); *Andersen v. King County*, 138 P.3d 963, 982 (Wash. 2006).

¹⁸⁶ *P. (S.E.) v. P. (D.D.)*, [2005] 50 B.C.L.R.4th 358 (Can.); *Thébeau v. Thébeau*, [2006] 302 N.B.R.2d 190 (Can.). A focus on troubling, recent developments in Canada is especially appropriate here because of the way previous scholarly work has attempted to use Canadian experience to argue the unalloyed benefits of extending marital regimes to same-sex couples in the United States. See, e.g., Mark E. Wojcik, *The Wedding Bells Heard Around the World: Years From Now, Will We Wonder Why We Worried About Same-Sex Marriage?*, 24 N. ILL. U. L. REV. 589, 636-47 (2004); Renée M. Landers, *A Marriage of Principles: The Relevance of Federal Precedent and*

Canadian provincial high courts have held extra-marital same-sex conduct to constitute “adultery” for purposes of the Divorce Act.¹⁸⁷ Under the historic Divorce Act, both “adultery” and “cruelty” constituted the sole fault grounds for divorce.¹⁸⁸ However, neither term is defined in the Act and, given the historically opposite-sex nature of marriage, it might seem that the former term necessarily involves opposite-sex intimacy.¹⁸⁹ As the British Columbia Supreme Court summarized the then-present law of “adultery” in its 2005 opinion, *P. (S.E.) v. P. (D.D.)*, “[a]lthough there is some uncertainty in the common law as to the precise definition of adultery, until now the courts in Canada have generally said that the act of adultery is between persons of the opposite sex.”¹⁹⁰

Nonetheless, in this case, the British Columbia Supreme Court deemed it necessary to “incremental[ly] change” this definition of adultery. It did so, noting that it took

parliament’s [recent] enactment of the *Civil Marriage Act* to be a legislative statement of the current values of our society [that is] consistent with the *Charter* [and which we are] obliged to use as a guide to [our] consideration of the current common law definition of adultery. Individuals of the same sex can now marry and divorce and the common law would be anomalous if those same-sex spouses were not bound by the same legal and social constraints against extra-marital sexual relationships that apply to heterosexual spouses.¹⁹¹

While deciding to apply the pre-modern heterosexual offence of adultery to homosexuals in this case, the court declined to define what specific acts of same-sex intimacy would constitute “adultery,” given that historical case law on this point seemed to primarily concern penile-vaginal contact.¹⁹²

International Sources of Law in Analyzing Claims for a Right to Same-Sex Marriage, 41 NEW ENG. L. REV. 683, 703-05 (2007).

¹⁸⁷ Canada Divorce Act, R.S.C., ch. 3 (1985).

¹⁸⁸ No-fault divorce is also available if “the spouses have lived separate and apart for at least year immediately preceding the determination of the divorce proceeding and were living separate and apart at the commencement of the proceeding.” *Id.*

¹⁸⁹ This conclusion is strengthened by the fact that, historically as well, “engag[ing] in a homosexual act” provided a *separate* fault ground for divorce. *See P. (S.E.) v. P. (D.D.)*, [2005] 50 B.C.L.R.4th 34 (Can.) (citing Canada Divorce Act, R.S.C. ch. 24 (1967-68)). This provision was removed in 1985, leaving “adultery” and “cruelty” as the sole fault grounds for divorce. *See id.* at 4.

¹⁹⁰ *Id.*

¹⁹¹ *Id.* at 16-17.

¹⁹² But see *Orford v. Orford*, [1921] 45 O.L.R. 15 (Can.) for a case where “artificial insemination, without the consent of the husband” was held to constitute

What parts of the male anatomy might have similarities to the vagina (in the case of male-male “adultery”), and what parts of the female body might be considered a penis (in the case of female-female “adultery”), the court explicitly declined to say. Indeed, the bashful court noted that such graphic explicitness would be neither “necessary [n]or desirable.”¹⁹³

As Part II discussed, the inability of majoritarian institutions to take into account non-majoritarian interests—much less find it “necessary or desirable” to do so—is one major reason why non-majoritarian peoples around the world have sought refuge and dignity outside of such institutions. With this in mind, this Article has proposed that a new goal for gay and lesbian people in the United States should be the imagination and legislation of a separate, more-homosexually-centered family law and relationship-recognition system. Indeed, the goal should not be a “separate but equal” system, but a “separate *and better*” one, the latter determination derived in part from the democratic-pedigree of the process behind this system’s formulation and its tight responsiveness to the people who will be specifically bound by it.

“Domestic partnerships” or “civil unions” provide one way out of the majoritarian problem. This is not to say that they provide the only way out, or necessarily the best way out forever, but they do represent a crucial beginning of the solution for America’s odd (and ironic)¹⁹⁴ incapacity to envision more than one possibility of the good intimate life, or to engage with family law pluralism in a sustained and rigorous manner. The existence and continuing development of legal alternatives to majoritarian marriage should be encouraged. Domestic partnerships and civil unions can be conceived of, not as a way-station on the road to majoritarian marriage, but as a way to avoid majoritarian marriage altogether.

The next section concludes this Part by discussing how one might further develop and improve the separate and

adultery because it involved “the possibility of introducing into the family of the husband a false strain of blood.” *See also* P. (S.E.) v. P. (D.D.) at 8-10.

¹⁹³ P. (S.E.) v. P. (D.D.) at 18.

¹⁹⁴ The ironies here are manifold, but one of the most interesting is the disconnect between a general American obsession with ensuring freedom generally, yet American paranoia with respect to *sexual* freedom particularly. *See* JANET R. JAKOBSEN & ANN PELLEGRINI, *LOVE THE SIN: SEXUAL REGULATION AND THE LIMITS OF RELIGIOUS TOLERANCE*, at ix (2003), for an attempt to understand “why the high value set on freedom in the United States comes crashing to the floor when it comes to sex. If freedom is such an important value in American life, then why isn’t sexual freedom a mainstream American value too?”

(arguably) better system of domestic partnerships and civil unions that has gained traction in the United States, including in its most populous state (California). The following necessarily consists only of musings at the minimum, and suggestions at the most, as the particular features of this system should be left to the results of a future gay and lesbian community-oriented discussion and debate.¹⁹⁵ The proposal here, after all, is a self-consciously democratic one. That being the case, two relatively specific recommendations will be advanced, namely that 1) gay and lesbian relationship-recognition schemes could use a different—and better—nomenclature than “domestic partnership,” “civil union,” and (also) “marriage,” and 2) gay and lesbian relationship-recognition schemes should work to facilitate greater gay and lesbian *freedom* and *agency* (as opposed to something called “choice”) by avoiding further legal entrenchment of pre-modern (heterosexually-authored) “sex offenses” such as adultery, infidelity, fornication, and the like.

C. Suggestions/Concerns

Any proposal for homosexual-authored and homosexual-respecting family law is likely to face only tepid (if any) support by traditional gay and lesbian (same-sex marriage) advocates. Their reaction will likely come back to arguments rehearsed in the California and Connecticut supreme courts, focusing on the alleged indignity of “separate but equal.” Part II raised serious doubts about the correctness of these universally-oriented claims, however, using transnational experience. This transnational experience holds several potential (and perhaps conflicting) lessons, but this Part has focused on one that is particularly relevant in a post-Proposition 8 U.S., namely the dignity—read as including a robust notion of *agency*—that can blossom by building and maintaining different family law systems for different types of people.

This concluding section builds on this basic (but nonetheless neglected) observation by exploring what a dignity-enhancing family law system for gay and lesbian people—one that is distinct from the troubled marital (and divorce) system that heterosexuals have built for their own purposes and

¹⁹⁵ As well as future academic research and commentary by myself and others. I consider this Article to be at the beginning of a much longer engagement by myself with the issues and ideas raised herein.

needs—might look like. In doing so, this section responds, in a more concrete fashion, to a concern about *nomenclature* which sits at the heart of gay and lesbian advocacy organizations’ “separate but equal” claims. This section also responds to a somewhat more inchoate worry about the *bona fides* of restricting people from exercising a “choice” to enter into marriage, even if an equally (or even better) endowed alternative—for example, domestic partnership or civil union—is available to them.

Before beginning each of these particular discussions, several observations and clarifications are (again) in order. First, while in some respects this Article’s proposal of the creation of a “separate *and better*” system of family law for gays and lesbians in the United States is a radical proposal, in many other respects it is just what remains to be worked out in the aftermath of Proposition 8 and similar measures. Even before Proposition 8, and before the legalization of same-sex marriage in California, gay and lesbian advocates in California had successfully argued for and helped legislate a separate system of relationship-recognition for same-sex couples that was broadly protective of such couples. Something similar happened in Connecticut (now replaced by a marriage regime for both opposite- and same-sex couples), and something similar now exists in New Jersey,¹⁹⁶ Nevada,¹⁹⁷ Oregon,¹⁹⁸ and Washington state.¹⁹⁹ While many people in California and elsewhere have viewed such separate systems as stepping stones towards (same-sex) marriage, recent events have demonstrated that it is far from certain that these struggles will actually end in marriage. Seen in this light, this Article’s proposal is rather banal in its acknowledgment of present realities, though

¹⁹⁶ N.J. STAT. ANN. § 37:1-28(d) (West 2007) (“Those rights and benefits afforded to same-sex couples under the Domestic Partnership Act should be expanded by the legal recognition of civil unions between same-sex couples in order to provide these couples with all the rights and benefits that married heterosexual couples enjoy.”).

¹⁹⁷ See Human Rights Campaign, Nevada Marriage/Relationship Recognition Law, <http://www.hrc.org/issues/1285.htm> (last visited Jan. 21, 2010).

¹⁹⁸ H.B. 2007, 74th Leg. Assem., Reg. Sess. § 2(5) (Or. 2007) (“Sections 1 to 9 of this 2007 Act are intended to better align Oregon law with the values embodied in the Constitution and public policy of this state, and to further the state’s interest in the promotion of stable and lasting families, by extending benefits, protections and responsibilities to committed same-sex partners and their children that are comparable to those provided to married individuals and their children by the laws of this state.”).

¹⁹⁹ WASH. REV. CODE § 26.60.015 (2010) (“It is the intent of the legislature that for all purposes under state law, state registered domestic partners shall be treated the same as married spouses.”).

admittedly it is radical to the extent that its proposal views gays' and lesbians' contributions to American discourses of sex, friendship, and family to be profound, insightful, and more worthy of emulation than much of what else has come to pass for common sense in the United States.

Second, while the suggestions below map out *differences* from the majoritarian marital regime that a separate gay and lesbian relationship-recognition regime might adopt (after democratic deliberation), this is not to suggest that any such regime must be completely different than what has come to pass before in order for this regime to prove its dignity/agency credentials. A separate system of family law is agency-enhancing because it provides a space from which to argue for a different set of norms than majoritarian ones *if and when* differences with the majority arise. I repeat that a separate system is not intended to be different for difference's sake, and no pluralist system of law anywhere in the world functions in this facile way. To the same extent that American federalism retains its value even as the 50 different states often adopt the same laws and policies, and to the same extent that Christian personal law in India retains its value even as it shares a disavowal of polygamy with Hindu personal law, so too does the separate system of relationship-recognition for same-sex couples outlined here retain its value even as it overlaps with heterosexual norms and practices. Indeed, even if gays and lesbians (in a particular state) chose to call their relationship-recognition system something like "same-sex marriage," despite the arguably more-attractive nomenclature options presented below, this nomenclature overlap with opposite-sex "marriage" still preserves for the future—in the legal separateness of its regime—the possibility of difference, either with respect to nomenclature or other aspects of family law. In an era of increasingly strident right-wing American politics, this potential is not only worth fighting for, but very likely requisite.

Third, the proposal for "separate and better" family law for gays and lesbians presented here is different than proposals put forward by Nancy Polikoff and similarly-minded activists and scholars.²⁰⁰ Polikoff's work,²⁰¹ in which she has developed an

²⁰⁰ See, e.g., Ruthann Robson, *Assimilation, Marriage, and Lesbian Liberation*, 75 TEMP. L. REV. 709, 710-12 (2002); BeyondMarriage.org, *Beyond Same-Sex Marriage: A New Strategic Vision for All Our Families and Relationships*, http://beyondmarriage.org/full_statement.html (last visited Jan. 21, 2010).

approach to family law that she calls “valuing all families,”²⁰² is extremely important. Such an approach recognizes that

[i]n every area of law that matters to same-sex couples, such as healthcare decision making, government and employee benefits, and the right to raise children, [non-marital] laws already exist in some places that could form the basis for just family policies for those who can’t marry or enter civil unions or register their domestic partnerships, as well as for those who don’t want to or who simply don’t.²⁰³

For Polikoff, such non-marriage-premised laws could be expanded in number and scope, as an alternative to merely pursuing and further entrenching the current practice of handing out healthcare, employment, and parental rights solely through the institution of marriage—whether opposite-sex or same-sex.²⁰⁴ The desirable goal, under Polikoff’s approach, would be “[l]aws that value all families,” i.e. laws which “ensur[e] that every relationship and every family has the legal framework for economic and emotional security,” and not (more) laws which merely “legitimate[] gay relationships that mirror marriage.”²⁰⁵

Clearly, this Article shares in Polikoff’s desire to de-center the role that marriage attempts to play for all people in contemporary American life. However, this Article’s proposals differ from Polikoff’s in that its proposals are simultaneously more realistic than Polikoff’s, and more radical. This “realistic radicalism” recognizes, like Polikoff, the need to start someplace else than “the package of rights that marriage gives different-sex couples and [merely] work[ing] down from there.”²⁰⁶ Instead, the goal should be something like, as Polikoff describes it, “identifying the needs of all LGBT people and work[ing] up from there to craft legislative proposals to meet those needs.”²⁰⁷ However, this Article’s discussions are also motivated by a very realistic recognition that “marriage”—and, indeed, “family” itself—are extremely centrifugal terms in contemporary American political life, ones which have

²⁰¹ The fullest statement of Polikoff’s beliefs can be found in her most recent book. See generally NANCY D. POLIKOFF, *supra* note 6.

²⁰² *Id.* at 5.

²⁰³ *Id.* at 9.

²⁰⁴ See *id.*

²⁰⁵ *Id.* at 210.

²⁰⁶ *Id.* at 209.

²⁰⁷ *Id.*

effectively—time and time again—subverted the possibility of radical “family values.” In fact, they occasionally work to subvert the radical possibilities of Polikoff’s own work.²⁰⁸

This being the case, a *new public vocabulary*—including a new nomenclature for same-sex relationships specifically—might very well be required to achieve the re-imagination of family values that both Polikoff and this Article desires. It is with this hunch, and for this reason, that this Article suggests a very public, and very political, and very “separate” system of gay and lesbian “family” law. Indeed, while “private ordering” of family law does contribute to increased agency, the suggestions presented here are not oriented towards further privatizing family law. In fact, they are much more oriented towards the “personal law” systems of family law which are found presently in locations as diverse as India and California. Separate (and better) systems of law for homosexuals—differing (to some degree) from majoritarian marriage in both nomenclature and substance—would not only help highlight and politicize homosexual lives, homosexual families, and homosexual family law,²⁰⁹ but also heterosexual lives,

²⁰⁸ For example, Polikoff apparently considers it important, in order to “value all families,” to provide a mechanism for unmarried partners (whether homosexual or heterosexual) to inherit wealth and property from one another in an orderly and predictable manner upon a partner’s death. *See id.* at 184-89. One might wonder, however, whether a more-progressive move would be to altogether re-think unstated (yet powerful) norms that sanction the private transfer of (large amounts of) resources between dead and living members of a “family.” The practice of inheritance is such a fundamental part of the law of “marriage” and “family,” however, that it seems doubtful whether either institution—whether valued or de-valued—can really allow for any profound re-imagining (or eradication) of it.

²⁰⁹ For this reason, I am in disagreement with some of the positions expressed by Marc Poirier in his recent, thoughtful work on the nomenclature issue. *See* Marc R. Poirier, *Name Calling: Identifying Stigma in the “Civil Union” / “Marriage” Distinction*, 41 CONN. L. REV. 1425, 1425 (2009). Poirier writes that

to deploy “civil union” and “marriage” properly requires everyone involved in interactions where these names are to be used to identify the couple as same- or different-sex. The mere fact of imposing a nomenclature distinction is problematic. . . . The law’s provision of a separate name serves to perpetuate microperformances and microidentifications of [the previously stigmatized category of “gay.”]

Id. at 1437. Putting aside (for the time being), the issue of whether the law should be encouraging microperformances of “acting straight,” any time any person seeks any type of recognition from the state, this interaction inevitably results in some loss of privacy. To (voluntarily) identify as “married” almost inevitably raises the question for the state (as well as employers, friends, and acquaintances): “To whom?” In this way, one might say that everyone (homosexual or heterosexual) who identifies as “married” is engaging in, at least in part, a flamboyant “coming out.” As Kenji Yoshino has stated: “I’m sometimes asked . . . whether I consider same-sex marriage to be an act of covering or flaunting. I think it is both.” KENJI YOSHINO, *COVERING: THE HIDDEN*

heterosexual families, and heterosexual family law. Indeed, such a proposal firmly puts the “majoritarian” into that which now simply passes as “marriage.”²¹⁰

Finally, a demurrer: For reasons of space, this section will not discuss how to actually procedurally operationalize a democratically-minded, (gay and lesbian) community-oriented legislative scheme. There are many questions to ponder in this respect. For example: Who counts as part of “the community”? How does one assess the community’s sentiments on any given proposal? To do so, would one return to the same gay and lesbian advocacy organizations which have sought alignment with majoritarian practice in the first place? To legislate, would one have to rely on the unpredictable votes of state legislators who don’t belong to “the community”? These are difficult questions, and this Article raises them to provide no conclusive answers. However, that being said, they are also questions that can only be raised in the event that gays and lesbians see a dignified alternative to majoritarian marriage in the first instance. This Article’s primary goal is to raise the possibility of this alternative, as a starting point to a much longer gay and lesbian community-oriented discussion involving these additional questions, if and when they should arise.

1. Nomenclature

One aspect of any future same-sex relationship-recognition regime(s) that will generate public interest concerns what these same-sex relationships will be called by the state, and how to ensure dignity with this choice in

ASSAULT ON OUR CIVIL RIGHTS 91 (2006); *see also* Poirier, *supra*, at 1488 n.375 (briefly acknowledging the difficulty that same-sex partners will have in hiding their homosexuality from the public even if they were legally-entitled to identify themselves as legally “married”).

Moreover, the existence of a new kind of relationship such as “civil union” or “domestic partnership” puts a great deal—and unprecedented amount—of onus on *heterosexuals* to account for their decisions to “marry.” In this way, heterosexuality becomes (micro)politicized in a way which previously only homosexuality was (by heterosexuals). In other words, with the advent of distinct forms of state relationship-recognition for homosexuals, heterosexuals’ microperformances and microidentifications would now become greatly magnified.

²¹⁰ Here I am somewhat echoing the views of Cheshire Calhoun, when she argues that debates concerning the possibility of same-sex marriage are so disturbing for many heterosexuals because these debates shine light on the heterosexual desire for “heterosexual love, marriage, and family [to] have a uniquely *prepolitical*, foundational status in civil society.” CHESHIRE CALHOUN, *FEMINISM, THE FAMILY, AND THE POLITICS OF THE CLOSET: LESBIAN AND GAY DISPLACEMENT* 127 (2000) (emphasis added).

nomenclature. Indeed, nomenclature appears to be the most crucial issue for a great number of people (whether heterosexual or homosexual) involved in the American same-sex marriage debates, overshadowing even (seemingly) important discussions about the substantive rights and responsibilities attaching to any potential same-sex relationship-recognition regime. The California Supreme Court described the importance of the issue of a nomenclature in the following noteworthy passage:

[I]t . . . is significant that although the meaning of the term “marriage” is well understood by the public generally, the status of domestic partnership is not. While it is true that this circumstance may change over time, it is difficult to deny that the unfamiliarity of the term “domestic partnership” is likely, for a considerable period of time, to pose significant difficulties and complications for same-sex couples, and perhaps most poignantly for their children, that would not be presented if, like opposite-sex couples, same-sex couples were permitted access to the established and well-understood family relationship of marriage.²¹¹

Echoing this concern with same-sex relationship nomenclature, Ronald Dworkin has recently written, in a widely-noted essay, that with respect to marriage and the debate over same-sex “marriage” versus “civil unions”: “We can no more now create an alternate mode of commitment carrying a parallel intensity of meaning than we can now create a substitute for poetry or for love.”²¹²

Dworkin’s claim, many people have felt, is a powerful one.²¹³ It is also one that opens up a broader discussion about the possibility of re-signifying terms. This is a large topic, to be sure, but something must be said here and, most bluntly, it is that Dworkin’s claim and ones like it largely work to obscure

²¹¹ *In re Marriage Cases*, 183 P.3d 384, 445-46 (Cal. 2008), *superseded by* CAL. CONST. art. I, § 7.5.

²¹² Ronald Dworkin, *Three Questions for America*, N.Y. REV. BOOKS, Sept. 21, 2006, available at <http://www.nybooks.com/articles/19271> (last visited Feb. 25, 2010).

²¹³ A recent crop of law review articles specifically addressing the nomenclature issue, and concluding that a different nomenclature for same-sex relationships is problematic, is evidence of the interest in this issue and also that most liberal thinkers broadly agree with Dworkin’s conclusion here. See, e.g., Courtney Megan Cahill, *(Still) Not Fit to be Named: Moving Beyond Race to Explain Why ‘Separate’ Nomenclature for Gay and Straight Relationships Will Never Be ‘Equal,’* 97 GEORGETOWN L.J. 1155 (2009); Suzanne A. Kim, *Marital Naming/Naming Marriage: Language and Status in Family Law*, 85 IND. L.J. (forthcoming 2010), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1351133; Poirier, *supra* note 209, at 1437.

(and dishonor) the history of the gay and lesbian civil rights movement.

This movement is one that has spoken the “love that dare not speak its name,” one that has made “gay” synonymous with both homosexuality *and* happiness, and one which has organized marches around the world every summer attesting to the “pride” that (many) gays and lesbians possess despite living in a world that desires to shame them. More generally, the gay and lesbian civil rights movement has been one that has demonstrated vividly the protean quality of words and labels—including “family,” “queer,” and even “sex”—and the alchemic potential of any ambitious politics of nomenclature.²¹⁴ It is also a movement that has managed to convince many heterosexuals to stop using the gendered terms “husband” and “wife,” or even the term “spouse,” and instead use the term “partner” to describe their “significant others.” In other words, despite the claims of Dworkin and like-minded others, neologisms can take hold, and the “disempowered” can change the terms of power’s discourse—sometimes quite literally.

It would appear to be the case, then, that if gays and lesbians could seize the opportunities which now attach to having “their own” family law in jurisdictions as populated and influential as California, New Jersey, and elsewhere, they would have a great deal of potential to change not only the vocabulary surrounding their own relationships, but also that surrounding relationships more broadly.

It is the case that “domestic partnership” and “civil union” are old terms, from another era, and arguably boring. As the epigraph to this Part suggests, they are also too suggestive of some of the domesticating aspirations and requirements of these current institutions.²¹⁵ That being the

²¹⁴ Indeed, one can view the insistence by mainstream gay and lesbian civil rights organizations and activists that “marriage” is the proper province of secular states, instead of churches and temples, *see supra* note 176, and their insistence that “marriage” can incorporate fertile, same-sex couplings—just as readily as it can sterile, opposite-sex couples—as a further testament to the general tendency of the larger gay and lesbian civil rights movement to believe in the possibility of challenging not only the conventional meaning of conventional words, but also what words should be used conventionally in the first place.

²¹⁵ *See also* Mary Anne Case, *Marriage Licenses*, 89 MINN. L. REV. 1758, 1772-74 (2005) for the observation that

[f]or good, as well as for ill, marriage now licenses couples to structure their lives as best suits them without losing recognition for their relationship. . . .
[A] marriage certificate now allows heterosexual couples to have an open

case, (re)claiming gay and lesbian ownership of these separate, same-sex relationship-recognition regimes, and then acting upon that ownership, could create real opportunities for a more exciting and less sterile nomenclature. The possibilities with respect to this nomenclature are relatively boundless and do not have to remain static and stale like “marriage” itself. Given this Article’s focus on dignity, one useful suggestion to contemplate might be that if gay and lesbian people want dignity, then they should give up the indirect pursuit of that through “marriage” and, instead, directly pursue that dignity by working to rename their state-recognized relationships as “dignity.” Indeed, with this name chosen *by* gays and lesbians *for* a gay and lesbian-authored family law institution, the contrast could not be more clear or more poetic: gays and lesbians would now enter into “dignity” while heterosexuals would enter into “marriage.”²¹⁶

2. “Choice”

The above nomenclature suggestion not only highlights the exciting opportunities for new relationship nomenclatures that a more pluralistic system of relationship-recognition permits, but also the way in which “marriage” itself increasingly possesses an uncertain valence in the contemporary United States.²¹⁷ Quite a bit of the contemporary legal discussion on “marriage” versus “domestic partnership” and “civil union” has rhetorically and simplistically distorted the (ostensibly positive) valence that marriage holds in today’s United States. While it might be possible (if unlikely) that gay and lesbian Midas-like magic can re-signify and revive the flagging fortunes of the *term* “marriage,” it is very unlikely that

marriage, to live in different cities or in different apartments in the same city, to structure their finances as they please, without having their commitment or the legal benefits that follow from it challenged. . . . [In contrast, t]he requirements of actual cohabitation in a shared residence and commingled finances are quite typical of most domestic partner registries.

²¹⁶ One important objection to this suggestion might be that calling any sort of relationship “dignity” implies that people outside of that relationship—for example, single people who are gay or lesbian—are “undignified.” I believe this objection is a legitimate one, though one potential response might be that the contemporary notion of dignity is a relatively universal and capacious value/trait and that it does not admit, conceptually at least, of its (potential) converse. In other words, there might be no dignified way, in our contemporary world, to treat someone with indignity or view them as undignified.

²¹⁷ See Abrams & Brooks, *supra* note 4, at 1.

gay and lesbian votes will ever be able to reform the legally-defined *institution* of majoritarian marriage, or other areas of family law linked to it. And whatever gains in dignity that gays and lesbians may accrue (and impart to others) by entering into “marriage” will very likely be outweighed by the loss of agency that gay and lesbian people will experience with respect to the definition and democratic legislation of laws that govern gay and lesbian lives and families.

This loss of agency that attaches to any gay and lesbian absorption into majoritarian marriage highlights the odd character of arguments that have been made about an alleged right to *choose* to “marry.” These kinds of arguments, explicitly about something characterized as “choice,” pop up here and there in the contemporary debates over same-sex marriage. For example, in *Goodridge*, the Massachusetts Supreme Judicial Court characterized the right at stake in that case as “the right to marry—or more properly, the right to choose to marry.”²¹⁸ This concern for “choice” will also likely make itself heard as an objection to this Article’s suggestion of a “separate and better” relationship-recognition regime for same-sex couples, the objection being that same-sex couples should—no matter what—have the right to “marry.”

While this section has already addressed the (perhaps surprising) compatibility of same-sex “marriage” with “separate and better” family law for gays and lesbians, it is nonetheless worth addressing some of the troublesome implications of the particular kind of “choice” arguments that some same-sex marriage advocates are making presently, in the process drawing out some of the important differences between these implications and those emanating from this Article’s particular suggestions.

In this respect, it is worth stating again that “choice” is an odd terrain over which to argue marriage rights. As Nancy Polikoff has astutely observed, “marriage would be a real *choice*”²¹⁹ if it were not so completely bound up with so many personal and social necessities (e.g. family and medical leave to take care of a sick marital partner).²²⁰ As Polikoff even more

²¹⁸ *Goodridge v. Dept. of Public Health*, 798 N.E.2d 941, 957 (Mass. 2003).

²¹⁹ POLIKOFF, *supra* note 6, at 133.

²²⁰ For similar reasons, Ruthann Robson calls marriage “compulsory.” See Robson, *supra* note 200, at 777.

clearly states: “marriage is not a choice if it is the *only* way to achieve economic well-being and peace of mind.”²²¹

Nonetheless, contemporary same-sex marriage advocates are using this vocabulary of “choice” in their advocacy of same-sex marriage rights. And these advocates often seem to be understanding this “choice” to be embodying some sort of libertarian-utopia-like vision of a right to choose the nomenclature of one’s state-recognized relationship. It is for this reason, indeed, that this Article anticipates “choice”-premiered objections to its suggestion for a “separate and better” relationship-recognition regime for non-majoritarian unions. These objections would arise especially if some gay and lesbians who wanted to “marry” were not able to do so because the gay and lesbian community, as a whole, in a given state, had decided to use their delegated right to designate the nomenclature for their state-recognized unions in a manner such that gay and lesbian unions would be called something other than “marriage.”

This libertarian formulation of “choice,” however, is hard to understand, not least because the same advocates who endorse a right to choose “marriage” nomenclature for one’s relationship do not intend to extend that general right to all people—including, most notably, those involved in polygamous relationships.²²² While perhaps such advocates would respond that, while they believe it inappropriate to extend the *substantive* rights, privileges, and burdens of marriage to polygamous groupings, they actually have no problem with the state merely *recognizing* polygamous “marital” unions, this complete bifurcation between the nomenclature and the substance of state relationship-recognition would be odd, not only as a matter of extant law,²²³ but also with regards to same-sex marriage advocates’ own goals. At the very least, it would open the door for (perhaps anti-gay) proposals to allow same-sex “marriages” but also to restrict the substantive benefits that accompany this particular form of “marital” recognition.²²⁴

²²¹ POLIKOFF, *supra* note 6, at 133.

²²² See *supra* text accompanying note 152; see also Robson, *supra* note 200, at 771.

²²³ This is not to deny that a more libertarian-like right to choose “marriage” nomenclature might (or should) develop, but it is to say that this right—in this formulation—does not presently exist. In fact, states routinely criminalize the conducting of “marriage” ceremonies both of and by unauthorized persons. See, e.g., TEX. FAM. CODE ANN. § 2.202(c)-(d) (Vernon 2009).

²²⁴ See Homer, *supra* note 149, at 516 (envisioning how “[e]ach benefit associated with marriage is susceptible to an analysis of the public policy that

“Choice” arguments, as they presently stand, are then not well-formulated: they do not stand on any deep foundation of extant United States legal practice and, furthermore, some formulations of these arguments actually open the door to forms of legal mischief that same-sex marriage advocates would themselves find troubling.

Moreover, anti-gay mischief is certainly afoot when arguments for *more* “choice” to enter into majoritarian marriage result in *less agency* for gays and lesbians with respect to laws that will deeply influence gay and lesbian lives and families. Indeed, it would seem that the right of “choice” is something different than the right to choose the laws that will heavily influence one’s life path. In other words, it seems that “choice” is something that exists in a great deal of tension with agency.

As the Canadian example has demonstrated, more gay and lesbian “choice” can result in less gay and lesbian *freedom*, especially to the extent that gay and lesbian absorption into majoritarian marriage results in the application of pre-modern sexual morality norms—for example, the “sex offenses” of adultery, infidelity, fornication, and the like—to gays and lesbians. Canada has witnessed such misadventures with its recent application of the pre-modern (heterosexual) sex offense of adultery to (married) same-sex couples.²²⁵ In the far more conservative and increasingly reactionary United States, the consequences of extending heterosexual traditions to homosexuals could be far more devastating. As Steven Homer has noted, to the extent that the availability of same-sex marriage gets linked, like opposite-sex marriage, to the “right to have sex, . . . [sexual morality] may easily turn on the married-unmarried distinction, leaving unmarried gays and lesbians with no sexual privacy. This would introduce into gay culture, for the first time, the concept of pre-marital sex.”²²⁶

This cannot be what dignity absolutely requires.

underlies it. Thus, to the extent that a court can find that a particular benefit does not belong to the class of benefits that make a couple married but rather reflects state recognition of the idiosyncracies of heterosexuality, that benefit can be denied to same-sex couples.”).

²²⁵ See *supra* text accompanying notes 187-190.

²²⁶ Homer, *supra* note 149, at 513.

CONCLUSION

Dignity may be a universal human aspiration, but its attainment is complicated by the messiness of human history and the richly textured quality of both what *is* and what *can be imagined* in any given locality.

The issue of imagination is thus central to the debate over legal pluralism and dignity. As this Article has discussed, because of the way anti-gay discourse has been configured for so long in the United States, it is very difficult for gays and lesbians to view themselves, in any positive way, as comprising a relatively distinct lifestyle or cultural grouping. “Specialness” becomes conflated with “queer” and the history of homophobia with which that term is associated. Social and legal arguments then tend to congregate around claims that homosexuals are “just like” heterosexuals, and that the two groups must be treated exactly the same both for the purposes of equality and dignity.

This has certainly been the recent view of both the California and Connecticut Supreme Courts. For these courts, gay and lesbian dignity is compromised by family law pluralism. This Article has attempted to demonstrate, however, that an alternative way of imagining the connection between dignity, legal pluralism, and marriage is available. It has also hopefully ignited the imagination of those people who are interested in developing “*separate and better*” gay and lesbian alternatives to majoritarian (heterosexual) marriage.

These alternatives should be developed by gay and lesbian people through a truly democratic debate and process. The political and legal agency for gay and lesbian people that will accompany such a process is an important component of reinforcing the dignity of gays and lesbians. Proposition 8 was a difficult piece of legislation to swallow, but it does not have to spell the end of gay and lesbian dignity. That dignity was always there and, if anything, it just needs to be re-discovered. That being said, this re-discovery may have to happen by traveling to very unfamiliar places. This Article has hoped to facilitate that journey.

NOTES

Treating Section 303(b) of the Bankruptcy Code as Subject-Matter Jurisdictional

SOUND APPROACH OR INVOLUNTARY REFLEX?

INTRODUCTION

Bankruptcy is typically thought of as a “last resort,”¹ a process by which debtors can obtain relief from unmanageable debt² and escape the incessant and distressing collection attempts of creditors.³ Given this characterization of bankruptcy as “relief,”⁴ it is not surprising that the vast majority of bankruptcy cases are initiated by debtors, who voluntarily accept the downsides of bankruptcy in exchange for

¹ “Almost all who file for bankruptcy do so as a last resort” 151 CONG. REC. E753, E754 (daily ed. Apr. 25, 2005) (statement of Rep. McCollum). *But see* 151 CONG. REC. S2405, S2474 (daily ed. Mar. 10, 2005) (statement of Sen. Frist) (“For many people, bankruptcy has become a first step rather than a last resort.”); DAVID A. SKEEL JR., *DEBT’S DOMINION: A HISTORY OF BANKRUPTCY LAW IN AMERICA* 1 (2001) (“[I]ndividuals and businesses in the United States do not seem to view bankruptcy as the absolute last resort”).

² *See* *Burlingham v. Crouse*, 228 U.S. 459, 473 (1913) (One purpose of bankruptcy is “to give the bankrupt a fresh start.”); *In re Chalasani*, 92 F.3d 1300, 1304 (2d Cir. 1996) (“[O]ne of the principal purposes of the Bankruptcy Code [is] allowing the debtor to begin a new life free from debt”).

³ *See In re Meyers*, 344 B.R. 61, 66-67 (Bankr. E.D. Pa. 2006) (“One significant remedial purpose of a bankruptcy discharge order is to prevent the emotionally harmful conduct associated with debt collection tactics.”); *In re Gervin*, 337 B.R. 854, 863 (Bankr. W.D. Tex. 2005) (“A significant component of [bankruptcy] is being free of the kinds of harassment, threats, and anxiety that debtors were suffering before they filed.”).

⁴ *See, e.g.,* Charles G. Hallinan, *The ‘Fresh Start’ Policy in Consumer Bankruptcy: A Historical Inventory and an Interpretive Theory*, 21 U. RICH. L. REV. 49, 51 (1986) (“The central importance of debtor relief in consumer bankruptcies is a commonplace of legal discussion.”).

freedom from crushing financial obligations.⁵ In fact, in most consumer bankruptcies, creditors have little incentive to see their debtors file for bankruptcy, since there are usually no assets left to distribute after state and federal exemptions are applied to the debtor's estate.⁶ The United States Bankruptcy Code (the "Bankruptcy Code") does, however, provide a means for creditors to force unwilling debtors into bankruptcy,⁷ a potentially appealing option for a creditor who fears that the debtor's existing nonexempt assets will have been squandered by the time the debtor finally files a voluntary bankruptcy petition.⁸

As might be expected, however, a creditor cannot push an unwilling debtor into bankruptcy with the relative ease with which a debtor can do so to itself. Instead, Section 303(b) of the Bankruptcy Code provides a number of requirements⁹ that

⁵ 2 COLLIER ON BANKRUPTCY ¶ 303.01 (Alan N. Resnick & Henry J. Sommer eds., 15th ed. 2002); see also H.R. REP. NO. 108-110, at 2 & n.9 (2003) ("[F]ewer than 1 percent of all bankruptcy case filings are commenced involuntarily."); Richard M. Hynes, *Broke but Not Bankrupt: Consumer Debt Collection in State Courts*, 60 FLA. L. REV. 1, 9 n.43 (2008) ("Involuntary cases are very rare.").

⁶ Ed Flynn, Gordon Bermant, & Suzanne Hazard, *Bankruptcy by the Numbers: Chapter 7 Asset Cases*, AM. BANKR. INST. J., Dec. 2002-Jan. 2003, at 22 ("About 96 percent of chapter 7 cases are closed without any funds collected and distributed to creditors . . ."); SKEEL, *supra* note 1, at 8 ("Although creditors can push a debtor into bankruptcy by filing an involuntary bankruptcy petition, they have little incentive to do so.").

⁷ 11 U.S.C. § 303 (2006).

⁸ See, e.g., Evan D. Flaschen & Carrie A. Brodzinski, *Involuntary Petitions Under the Bankruptcy Code*, 547 PLI/COMM 93, 97-98 (1990) (detailing several situations that might prompt a creditor to file an involuntary bankruptcy petition); H.R. REP. NO. 108-110, at 2 ("[A]n involuntary bankruptcy petition can serve as a useful creditor collection tool. For example, it can preserve assets from further dissipation and provide for their orderly liquidation by a bankruptcy trustee, a fiduciary charged by statute to protect such assets and maximize their value for the benefit of creditors."); COLLIER, *supra* note 5, ¶ 303.01 ("[T]here are certain key situations in which the filing of an involuntary case remains a beneficial, and sometimes optimal, choice for creditors . . ."); *In re All Media Props., Inc.*, 5 B.R. 126, 134 (Bankr. S.D. Tex. 1980) ("[I]t is important that involuntary petitions be tried and resolved promptly because if the debtor is not paying its debts as they become due, then its creditors are entitled to the protection of their rights afforded by the Code and to prevent the debtor from wasting its assets.").

⁹ 11 U.S.C. § 303(b). This provision provides in relevant part:

An involuntary case against a person is commenced by the filing with the bankruptcy court of a petition under chapter 7 or 11 of this title—

(1) by three or more entities, each of which is . . . a holder of a claim against such person that is not contingent as to liability or the subject of a bona fide dispute as to liability or amount . . . if such noncontingent, undisputed claims aggregate at least [\$13,475] more than the value of any lien on property of the debtor securing such claims held by the holders of such claims; [or]

must be met in order for relief to be entered in an involuntary bankruptcy proceeding.¹⁰ The majority of courts and commentators have interpreted these requirements as *prima facie* elements of an involuntary bankruptcy case, which must be either disputed or waived by the debtor.¹¹ However, in *In re BDC 56*,¹² the Second Circuit Court of Appeals joined the minority of jurisdictions and interpreted certain of these requirements as being subject-matter jurisdictional in nature,¹³ meaning that they pertain to “the courts’ statutory or constitutional *power* to adjudicate the case.”¹⁴ The Second Circuit based its position on the argument that creditors should be forced to prove the sufficiency of the involuntary petition “at the earliest practicable point.”¹⁵ This Note argues that the Second Circuit’s treatment of the Section 303(b) requirements as subject-matter jurisdictional is contradictory to the provision’s implied goals of fairness and judicial efficiency, and is inconsistent with the jurisdictional structure of the United States bankruptcy system.

Part I of this Note surveys the history of the Bankruptcy Code, focusing on the origins and development of federal jurisdiction over bankruptcy cases. It details the jurisdictional structure of the bankruptcy system, laying a

(2) if there are fewer than 12 such holders . . . by one or more of such holders that hold in the aggregate at least \$10,000 of such claims . . .

Id.

¹⁰ Note that in the context of an involuntary bankruptcy, a debtor who opposes the petition is seeking dismissal rather than relief. *See In re Alta Title Co.*, 55 B.R. 133, 135-36 (Bankr. D. Utah 1985) (“An involuntary petition must end either in the entry of an order for relief against the debtor or dismissal of the creditors’ petition.”).

¹¹ *See, e.g., In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1101 (11th Cir. 2008) (collecting cases and authority supporting this proposition), *overruled by* 550 F.3d 1035 (11th Cir. 2008) (en banc).

¹² *In re BDC 56 LLC*, 330 F.3d 111 (2d Cir. 2003).

¹³ *Id.* at 118. In fact, although the Second Circuit shares its position with a small number of bankruptcy courts in other circuits, it is the only circuit court to have explicitly adopted this holding. *See, e.g., In re Paczesny*, 283 B.R. 715, 718 (Bankr. N.D. Ill. 2002) (“The absence of a bona fide dispute [as required by § 303(b)] is a jurisdictional prerequisite.”); *In re New Mexico Props., Inc.*, 18 B.R. 936, 939-40 (Bankr. D.N.M. 1982) (describing § 303(b) as a “[jurisdictional] hurdle for petitioning creditors to overcome”). The Ninth and Eleventh Circuits are the only other circuits to have explicitly ruled on this issue, both concluding that § 303(b) is not subject-matter jurisdictional in nature. *In re Trusted Net Media*, 550 F.3d at 1046; *In re Rubin*, 769 F.2d 611, 615 (9th Cir. 1985).

¹⁴ *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 89 (1998); *see infra* Part III.A.1 (describing the basic principles of subject matter jurisdiction).

¹⁵ *In re BDC 56 LLC*, 330 F.3d at 118.

foundation for the argument that the Section 303(b) requirements do not pertain to the bankruptcy courts' jurisdiction. Part II investigates the Section 303(b) circuit split, examining three cases in which courts have justified their treatment of the Section 303(b) requirements either as subject-matter jurisdictional or as "substantive matters which must be proved or waived for petitioning creditors to prevail in involuntary proceedings."¹⁶ Part III assesses the various rationales for, and implications of, both sides of the circuit split and contends that the Second Circuit's approach to the Section 303(b) requirements is inconsistent with the jurisdictional structure of the Bankruptcy Code, wastes judicial resources, and therefore should be abandoned. This section highlights two recent cases, one from the United States Supreme Court¹⁷ and another from the Bankruptcy Court of the Southern District of New York,¹⁸ both of which cast doubt on the Second Circuit's treatment of Section 303(b) as subject-matter jurisdictional and indicate that *BDC* should no longer be upheld as good law. This Note concludes that the Second Circuit should resolve the circuit split in favor of treating Section 303(b) as substantive, and not subject-matter jurisdictional, in nature.

I. DEVELOPMENT OF THE MODERN BANKRUPTCY CODE

In order to appreciate the merits and weaknesses of the arguments on either side of the Section 303(b) circuit split, it is first necessary to understand the overall structure of the Bankruptcy Code, including the role of involuntary bankruptcy proceedings and the rationale behind the modern jurisdictional structure of the U.S. bankruptcy system. Involuntary bankruptcy cases, though far less common today than they were at the inception of our nation's bankruptcy laws,¹⁹ were always within the jurisdiction of the bankruptcy courts.²⁰ While

¹⁶ *In re Rubin*, 769 F.2d at 614 n.3.

¹⁷ *Arbaugh v. Y & H Corp.*, 546 U.S. 500 (2006).

¹⁸ *In re MarketXT Holdings Corp.*, 347 B.R. 156 (Bankr. S.D.N.Y. 2006).

¹⁹ SKEEL, *supra* note 1, at 8 ("Under current law, the vast majority of debtors file for bankruptcy voluntarily In the nineteenth century, by contrast, involuntary bankruptcy figured quite prominently."); *supra* note 5 (discussing infrequency of involuntary bankruptcy petitions today).

²⁰ SKEEL, *supra* note 1, at 27 ("By 1867, it was evident that Congress could enact both voluntary and involuntary laws . . . [which] were administered through the federal district courts."); David S. Kennedy & R. Spencer Clift, III, *An Historical Analysis of Insolvency Laws and Their Impact on the Role, Power, and Jurisdiction of Today's United States Bankruptcy Court and its Judicial Officers*, 9 J. BANKR. L. &

the jurisdictional structure of bankruptcy courts is quite different today, its evolution has been shaped by the desire to make the bankruptcy process as fair, efficient, and cost-effective as possible.²¹ In order to further these objectives, Section 303(b) must be interpreted as substantive rather than jurisdictional because this approach better comports with the statutory structure of the bankruptcy system, creates greater predictability, and leads to more efficient resolution of bankruptcy cases.

A. *Involuntary Bankruptcy and the Roots of Modern U.S. Bankruptcy Law*

The English bankruptcy laws, from which our modern bankruptcy system evolved, were in fact remarkably different from the scheme that the United States has in place today.²² Most notably, the first English bankruptcy statutes, enacted under Henry VIII in 1582, treated the debtor as a criminal, did not release the debtor from debts remaining after liquidation and distribution, and could only be invoked on the initiative of the creditors.²³ Although later versions of English bankruptcy law decriminalized the proceedings and provided for discharge of unsatisfied obligations,²⁴ the process remained one that was commenced by creditors against potentially unwilling debtors.²⁵ In other words, the only bankruptcy proceeding available was involuntary.²⁶

With little debate or fanfare, the Founding Fathers granted to Congress the constitutional power to pass bankruptcy laws.²⁷ When Congress passed the first federal

PRAC. 165, 170-71 (2000) (the first Bankruptcy Act of the United States provided for only involuntary bankruptcy cases, and gave district courts jurisdiction to appoint “non-judicial, bankruptcy ‘commissioners’ to assist in administering proceedings under this Act”).

²¹ The goal of the Bankruptcy Code is to “secure the just, speedy, and inexpensive determination of every case and proceeding.” FED. R. BANKR. P. 1001.

²² Marcia S. Krieger, “*The Bankruptcy Court is a Court of Equity*”: *What Does That Mean?*, 50 S.C. L. REV. 275, 281-82 (explaining that “American social, economic, and philosophical influences modified English tradition to create American bankruptcy law”).

²³ Kennedy & Clift, *supra* note 20, at 169.

²⁴ *Id.* at 169-70.

²⁵ *See id.* at 169-70.

²⁶ *Id.* at 169.

²⁷ U.S. CONST. art. I, § 8, cl. 4 (granting Congress the power to pass “uniform Laws on the subject of Bankruptcies”); Jonathan C. Lipson, *Debt and Democracy: Towards a Constitutional Theory of Bankruptcy*, 83 NOTRE DAME L. REV. 605, 608

bankruptcy statute in 1800, it adopted the creditor-friendly involuntary approach used in England.²⁸ However, the Bankruptcy Act of 1800 was soon repealed and it was not until two more failed attempts by Congress that a workable bankruptcy system emerged in 1898.²⁹ The intervening years witnessed a continuous struggle between debtors and creditors to shape the law in their respective favor.³⁰ But by 1898, it became evident that debtors had definitively won the battle for voluntary bankruptcy proceedings, which first appeared in the Bankruptcy Act of 1841 and have remained a fixture of United States bankruptcy law ever since.³¹ Nevertheless, the 1898 Bankruptcy Act still included provisions for involuntary bankruptcy petitions,³² which remain substantially unchanged to this day.³³

(2008) (this constitutional provision was drafted “with surprisingly little debate”); SKEEL, *supra* note 1, at 23 (the provision was included “almost as an afterthought . . . and it was approved with little debate”).

²⁸ SKEEL, *supra* note 1, at 25; Kennedy & Clift, *supra* note 20, at 170-71.

²⁹ Kennedy & Clift, *supra* note 20, at 171-75 (“[The 1898] Act formed the basis of our modern bankruptcy laws.”).

³⁰ DAVID A. MOSS, WHEN ALL ELSE FAILS: GOVERNMENT AS THE ULTIMATE RISK MANAGER 136-37 (2004); Krieger, *supra* note 22, at 293 (“Until the middle of the nineteenth century, bankruptcy law was decidedly pro-creditor. Since then it has oscillated between provisions favoring debtors and those favoring creditors, depending on economic and political pressures at a given time.”).

³¹ Kennedy & Clift, *supra* note 20, at 171-75; MOSS, *supra* note 30, at 136-38.

³² Bankruptcy Act of 1898, ch. 541, § 59, 30 Stat. 544, 561-62 (1898) (repealed 1978). The provision governing involuntary cases under this Act stated:

[t]hree or more creditors who have provable claims against any person which amount in the aggregate, in excess of the value of securities held by them, if any, to five hundred dollars or over; or if all of the creditors of such persons are less than twelve in number, then one of such creditors whose claim equals such amount may file a petition to have him adjudged a bankrupt. . . . If it be averred in the petition that the creditors of the bankrupt are less than twelve in number, and less than three creditors have joined as petitioners therein, and the answer avers the existence of a larger number of creditors, there shall be filed with the answer a list under oath of all the creditors, with their addresses, and thereupon the court shall cause all such creditors to be notified of the pendency of such petition and shall delay the hearing upon such petition for a reasonable time, to the end that parties in interest shall have an opportunity to be heard

Bankruptcy Act of 1898 § 59(b), (d).

³³ Compare Bankruptcy Act of 1898 § 59(b) (requirements for creditors to file an involuntary bankruptcy petition), with 11 U.S.C. § 303(b), (c) (2006), and FED. R. BANKR. P. 1003(b) (same).

B. Development of the Present-Day Jurisdictional Structure of the Bankruptcy Code

The 1898 Bankruptcy Act remained in place until the passage of the 1978 Bankruptcy Code, which was enacted after lengthy studies by both the Brookings Institution³⁴ and the congressionally established Commission on the Bankruptcy Laws of the United States.³⁵ By 1970, it had become apparent that in addition to being outdated in a variety of respects, the 1898 Act caused confusion and inefficiency in bankruptcy suits because of serious jurisdictional deficiencies.³⁶ In particular, bankruptcy court jurisdiction under the 1898 Act was limited to the bankruptcy proceeding itself plus a narrow class of controversies that arose during the course of the bankruptcy proceeding.³⁷ All other disputes that arose during the case had to be litigated separately in either state or federal district court.³⁸ The resulting “bifurcated jurisdiction” led to great expense and delay due to both the “threshold jurisdictional litigation” as well as the practical inconvenience of litigating in multiple forums.³⁹ The 1978 Bankruptcy Code sought to remedy

³⁴ Kennedy & Clift, *supra* note 20, at 177. “The Brookings Institution is a nonprofit public policy organization based in Washington, D.C. . . [whose] mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations [to] [s]trengthen American democracy; [f]oster the economic and social welfare, security and opportunity of all Americans and [s]ecure a more open, safe, prosperous and cooperative international system.” Brookings Institution, <http://www.brookings.edu/about.aspx> (last visited Jan. 12, 2010).

³⁵ Judith A. McKenna & Elizabeth C. Wiggins, *Alternative Structures for Bankruptcy Appeals*, 76 AM. BANKR. L.J. 625, 638 (2002) (“Congress established [the Commission] in 1970 . . . to ‘study, analyze, evaluate, and recommend changes’ in the Bankruptcy Act. The ensuing report and hearings ultimately led to the passage of the Bankruptcy Reform Act of 1978.”) (footnote omitted) (quoting COMM’N ON THE BANKRUPTCY LAWS OF THE U.S., REPORT ON THE BANKRUPTCY LAWS OF THE U.S., H.R. DOC. NO. 137 93d Cong., 1st Sess., part I, at 1-2 (1973)); Kennedy & Clift, *supra* note 20, at 177.

³⁶ Kennedy & Clift, *supra* note 20, at 177, 188. The term “jurisdiction” in this discussion, and in this Note generally, relates to subject matter jurisdiction rather than personal jurisdiction. Personal jurisdiction is relatively easy to obtain in bankruptcy proceedings, at least where the defendant is located in the United States, because the bankruptcy courts can effect nationwide service of process. FED. R. BANKR. P. 7004; Leonard Gerson, *Class Proofs of Claim and Class Actions in Bankruptcy: Clarifying the Law, Improving the Process, and Expanding the Use of Class Actions*, 17 J. BANKR. L. & PRAC. 6 Art. 2, at n.208 (“[C]ourts . . . have determined that the minimum contacts required for a bankruptcy court to have personal jurisdiction over an entity is satisfied by the entity’s presence in the U.S. as a whole rather than in any particular state.”).

³⁷ Kennedy & Clift, *supra* note 20, at 187.

³⁸ *Id.* at 187-88.

³⁹ *Id.* at 188; Ralph Brubaker, *On the Nature of Federal Bankruptcy Jurisdiction: A General Statutory and Constitutional Theory*, 41 WM. & MARY L. REV. 743, 792 (2000) (“The primary vice of the 1898 Act’s jurisdictional regime was that it

this shortcoming by granting expansive subject matter jurisdiction to the bankruptcy courts.⁴⁰ Thus, the Code created independent bankruptcy courts that were instructed to exercise “original and exclusive jurisdiction of all cases under title 11” and “original but not exclusive jurisdiction of all civil proceedings arising under title 11 or arising in or related to cases under title 11.”⁴¹

Within four years, however, this broad jurisdictional grant to bankruptcy courts failed a constitutional challenge in the United States Supreme Court.⁴² In 1982, the Supreme Court in *Northern Pipeline Construction v. Marathon Pipeline Company* held that the jurisdictional grant of the 1978 Bankruptcy Code was unconstitutional because it allowed non-Article III judges⁴³ to adjudicate matters governed by state law that were merely “related to” a bankruptcy case.⁴⁴ Justice Brennan’s plurality opinion explained that Article III of the United States Constitution was designed to ensure the separation of powers and to protect the independence of the judiciary.⁴⁵ While Congress has the authority to assign certain judicial functions to non-Article III “adjunct tribunals,”⁴⁶ the

engendered an excessive amount of preliminary litigation over jurisdictional issues surrounding the bifurcation of bankruptcy jurisdiction.”)

⁴⁰ Brubaker, *supra* note 39, at 791; Kennedy & Clift, *supra* note 20, at 188; *see also In re Hospitality Ventures/LaVista*, 358 B.R. 462, 478 (Bankr. N.D. Ga. 2007) (“One of [the] primary objectives [of the Bankruptcy Reform Act of 1978] was to expand bankruptcy jurisdiction and eliminate disputes over what bankruptcy judges could hear in order to avoid costly and time-consuming arguments over jurisdiction.”).

⁴¹ 28 U.S.C. §§ 151, 1471 (Supp. III 1980).

⁴² *See Northern Pipeline Constr. Co. v. Marathon Pipeline Co.*, 458 U.S. 50, 50 (1982).

⁴³ *Id.* at 61 (“[T]here is no doubt that the bankruptcy judges created by the [1978] Act are not Art. III judges.”).

⁴⁴ *Id.* at 88; 1 HON. WILLIAM L. NORTON, JR. & WILLIAM L. NORTON, III, NORTON BANKRUPTCY LAW AND PRACTICE § 4:24 (3d ed. 2009); THOMAS J. SALERNO & JORDAN A. KROOP, BANKRUPTCY LITIGATION AND PRACTICE § 3.05 (2006) (“The Supreme Court’s principal concern was that bankruptcy judges, who were appointed for fixed terms and had salaries subject to reduction by Congress, had jurisdiction under the law to hear and decide all matters, even those based solely on state law and having only a tangential nexus to the bankruptcy estate.”).

⁴⁵ *Northern Pipeline*, 458 U.S. at 57-60. The bulk of this protection comes from the fact that Art. III judges are given life-tenure and guaranteed a non-diminishing salary, thus ensuring that concerns about their compensation do not color their judgment. *Id.* at 59; U.S. CONST. art. III, § 1. Under the 1978 Act, bankruptcy judges served only 14-year terms, they were subject to removal for “incompetency, misconduct, neglect of duty, or physical or mental disability,” 28 U.S.C. § 1471 (1978), and were not provided a guaranteed salary. *Northern Pipeline*, 458 U.S. at 53.

⁴⁶ *Northern Pipeline*, 458 U.S. at 77, 80-82. “The [1978] Act designate[d] the bankruptcy court in each district as an ‘adjunct’ to the district court.” *Id.* at 63 n.13 (quoting 28 U.S.C. § 151(a) (1976)).

1978 Act vested bankruptcy judges with all the “essential attributes’ of the judicial power of the United States.”⁴⁷ Finding that this broad jurisdictional grant exceeded “Congress’ power to create adjuncts to Art. III courts,” a plurality of the Court held that the jurisdictional provision of the 1978 Act was unconstitutional.⁴⁸

In order to keep the bankruptcy system afloat, the federal courts adopted the “Emergency Rule,” which was effectively a return to bifurcated jurisdiction.⁴⁹ When Congress finally passed the Bankruptcy Amendments and Federal Judgeship Act of 1984, it essentially maintained this bifurcated approach.⁵⁰ Under the act, Congress granted jurisdiction over bankruptcy proceedings to the federal district courts via 28 U.S.C. § 1334.⁵¹ It also designated bankruptcy courts as “unit[s] of the district court[s]”⁵² and authorized district courts to refer Title 11 cases to the bankruptcy court in their judicial districts via 28 U.S.C. § 157(a).⁵³ As a result of the 1984 Code’s demarcation between “core” and “non-core” (or “related to”) proceedings,⁵⁴ an approach that was adopted in order to implement the lessons learned in *Marathon Pipeline*, bifurcated jurisdiction became entrenched.⁵⁵ Core proceedings are those matters having a sufficiently close nexus to the pending bankruptcy so as to make final determination by the bankruptcy court proper.⁵⁶ Non-core matters, on the other hand, cannot be finally determined by the bankruptcy court without the consent of the affected parties.⁵⁷ Without such consent, the bankruptcy court can only make a recommendation, which

⁴⁷ *Id.* at 84-85.

⁴⁸ *Id.* at 87.

⁴⁹ Kennedy & Clift, *supra* note 20, at 189-90.

⁵⁰ *Id.* at 190-91 (“Broadly and briefly stated, another bifurcated jurisdictional approach was adopted by Congress in the 1984 amendments.”).

⁵¹ The statute provides in relevant part that “the district courts shall have original and exclusive jurisdiction of all cases under title 11.” 28 U.S.C. § 1334(a) (1982).

⁵² *Id.* § 151.

⁵³ This provision states that “[e]ach district court may provide that any or all cases under title 11 and any or all proceedings arising under title 11 or arising in or related to a case under title 11 shall be referred to the bankruptcy judges for the district.” *Id.* § 157(a).

⁵⁴ *Id.* § 157; Kennedy & Clift, *supra* note 20, at 191-92; 1 NORTON & NORTON, *supra* note 44, at §§ 4:10 & 4:28.

⁵⁵ Kennedy & Clift, *supra* note 20, at 180, 191.

⁵⁶ 1 NORTON & NORTON, *supra* note 44, at § 4:28 (“A nonexhaustive listing of ‘core’ proceedings is set forth in 28 [U.S.C.] § 157(b)(2).”).

⁵⁷ *Id.*

must go up to the district court for entry of a final order.⁵⁸ Moreover, Congress mandated that, upon timely motion of a party, federal courts must abstain from hearing a state law claim that could not have otherwise been commenced in the federal court had it not been introduced in a bankruptcy proceeding.⁵⁹ This jurisdictional structure has been preserved through subsequent amendments to the Bankruptcy Code and is still in effect today.⁶⁰

While numerous disputes as to the proper scope of bankruptcy jurisdiction remain,⁶¹ two observations emerge that are relevant to the analysis of Section 303(b). First, the Bankruptcy Code is divided, both conceptually and organizationally, into separate substantive and jurisdictional sections, with the substantive sections establishing the various types of bankruptcy cases available⁶² and the jurisdictional provisions granting district courts and bankruptcy courts the authority to hear those cases.⁶³ It has never been doubted that involuntary bankruptcy cases, like any other bankruptcy proceeding, are “cases under title 11” for the purposes of Congress’ grant of subject matter jurisdiction to the bankruptcy courts.⁶⁴ Thus, to the extent that the provisions of Title 11 are

⁵⁸ *Id.*

⁵⁹ 28 U.S.C. §§ 1334(c)(2), 157 (2006); SALERNO & KROOP, *supra* note 44, at § 3.10[B].

⁶⁰ 1 NORTON & NORTON, *supra* note 44, at § 4:10.

⁶¹ *See, e.g.*, Jackie Gardina, *The Bankruptcy of Due Process: Nationwide Service of Process, Personal Jurisdiction and the Bankruptcy Code*, 16 AM. BANKR. INST. L. REV. 37, 54 & n.85 (2008) (pointing out disagreement between circuit courts as to whether a bankruptcy court may retain “their ‘related to’ jurisdiction . . . after the bankruptcy has been dismissed”); Radha A. Pathak, *Breaking the “Unbreakable Rule”: Federal Court, Article I, and the Problem of “Related To” Bankruptcy Jurisdiction*, 85 OR. L. REV. 59, 61 (2006) (“The United States Supreme Court appears to have accepted the constitutionality of ‘related to’ bankruptcy jurisdiction, but it has never explicitly articulated the constitutional basis for such jurisdiction.”); 1 NORTON & NORTON, *supra* note 44, at § 4:63 (identifying “split of authority on whether a particular type of proceeding is ‘core’ or ‘related to’ the bankruptcy case”).

⁶² The substantive provisions reside in Title 11. *See, e.g.*, 11 U.S.C. Ch. 7 (concerning liquidation cases); 11 U.S.C. Ch. 11 (concerning reorganization cases); 11 U.S.C. Ch. 13 (concerning adjustment cases).

⁶³ The jurisdictional provisions reside in Title 28. *See, e.g.*, 28 U.S.C. §§ 157 & 1334.

⁶⁴ 28 U.S.C. §§ 157 & 1334; *see also In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1044 (11th Cir. 2008) (“As a class of cases, involuntary bankruptcy cases unquestionably arise under Title 11 (the Bankruptcy Code), and thus fall within the congressional grant of subject matter jurisdiction to the bankruptcy courts.”); *In re Bowshier*, 313 B.R. 232, 238 (Bankr. S.D. Ohio 2004) (“There is no dispute that bankruptcy courts have jurisdiction over involuntary bankruptcy proceedings.”).

construed as substantive in nature, Section 303 should be similarly interpreted.⁶⁵

Second, while bankruptcy jurisdiction is notoriously complex,⁶⁶ this complexity stems more from the constitutional uncertainty surrounding the broad congressional grant of jurisdiction to non-Article III judges⁶⁷ than from any significant disagreement as to the finer contours of bankruptcy court jurisdiction. The area of greatest uncertainty in the context of this jurisdiction relates to the bankruptcy courts' ability to entertain cases and proceedings other than the bankruptcy case itself.⁶⁸ While this issue is just as likely to arise in the context of an involuntary bankruptcy as in a debtor-initiated bankruptcy, the issue of whether Section 303(b) relates to subject matter jurisdiction is largely unrelated to this particular area of uncertainty. Rather, it has more to do with general notions of subject matter jurisdiction and statutory construction. Consequently, despite falling within the broader and more complex realm of bankruptcy court jurisdiction, the question of how to best interpret Section 303(b) can be addressed using the same rules of statutory analysis as are used in other contexts.

II. THREE CASES ADDRESSING THE SECTION 303(b) REQUIREMENTS

Section 303(b) of the Bankruptcy Code requires that an involuntary petition be brought by creditors holding claims that “aggregate at least [\$13,475]”⁶⁹ and that are “not contingent as to liability or the subject of a bona fide dispute as

⁶⁵ An involuntary bankruptcy does not, in fact, arise under its own distinct chapter of the Bankruptcy Code, but rather is a bankruptcy case of the type established by 11 U.S.C. Ch. 7 or 11 U.S.C. Ch. 11. 11 U.S.C. § 303(a) (“An involuntary case may be commenced only under chapter 7 or 11 of this title . . .”). Thus, § 303 is best viewed as supplementing those substantive provisions that establish liquidation and reorganization cases, so as to allow for their commencement by a creditor as opposed to the debtor.

⁶⁶ See, e.g., Brubaker, *supra* note 39, at 746 (“[T]he jurisdiction in bankruptcy remains one of the most enduring puzzles of our federal court system.”); Lipson, *supra* note 27, at 645 (“At least as a conceptual matter, bankruptcy jurisdiction is exceedingly—perhaps needlessly—complex . . .”).

⁶⁷ See, e.g., Lipson, *supra* note 27, at 645-46.

⁶⁸ See *supra* note 61.

⁶⁹ 11 U.S.C. § 303(b) (2006). This amount was increased from \$5,000 to \$10,000 in 1994, and the Bankruptcy Code requires that as of 1998, automatic adjustments take place every three years. Bankruptcy Reform Act of 1994, Pub. L. 103-394, § 108(b)(1) (Oct. 22, 1994); 11 U.S.C. § 104; 2 NORTON & NORTON, *supra* note 44, at § 22:7.

to liability or amount.”⁷⁰ Moreover, if the debtor has more than twelve creditors, the petition cannot be brought by fewer than three of those creditors.⁷¹ Few courts have bothered to perform a rigorous analysis of how to best characterize the Section 303(b) requirements, either because the procedural postures of the involuntary cases before them have not required it,⁷² or because they simply chose to apply a precedent that mandated a particular conclusion.⁷³ Nevertheless, those courts that have addressed this issue have reached conflicting results, with the overwhelming majority of them finding that the Section 303(b) requirements are not subject-matter jurisdictional but rather substantive, as recently held by the Eleventh Circuit in *In re*

⁷⁰ 11 U.S.C. § 303(b). The noncontingency requirement prevents claim-holders from being counted toward the requisite number of petitioning creditors if their claims are dependent on the occurrence of a future uncertain event, such as “the liability of a guarantor when the principal has not defaulted.” 2 NORTON & NORTON, *supra* note 44, at § 22:3. The undisputed claim requirement is meant to keep creditors from forcing a debtor into bankruptcy when there is a “legitimate disagreement over whether money is owed, or, in certain cases, how much.” *In re Vortex Fishing Sys., Inc.*, 277 F.3d 1057, 1064 (9th Cir. 2004). It “prevent[s] creditors from using the bankruptcy courts as a club in collecting claims that [are] disputed, although not contingent.” 2 NORTON & NORTON, *supra* note 44, at § 22:3. Although the phrase “bona fide dispute” is not defined in the Bankruptcy Code, the circuit courts have defined it as “an objective basis for either a factual or a legal dispute as to the validity of the debt.” *In re Byrd*, 357 F.3d 433, 437 (4th Cir. 2004) (quoting *Matter of Busick*, 831 F.2d 745, 750 (7th Cir. 1987)). “The bankruptcy court need not resolve the merits of the bona fide dispute, but simply determine whether one exists.” *Id.* (citation omitted).

⁷¹ 11 U.S.C. § 303(b).

⁷² An involuntary bankruptcy petition is often “timely controverted” by the debtor, 11 U.S.C. § 303(h), in which case the characterization of § 303(b) as substantive or jurisdictional loses its significance because there is no longer any question of the debtor’s possible waiver of the § 303(b) requirements as a defense. *See, e.g., In re Reg’l Anesthesia Assocs. PC*, 360 B.R. 466, 470 (Bankr. W.D. Pa. 2007) (dismissing involuntary petition due to “bona fide dispute” after debtor timely controverted the petition); *In re Euro-American Lodging Corp.*, 357 B.R. 700, 730 (Bankr. S.D.N.Y. 2007) (ordering relief against involuntary debtor who filed a timely answer because petitioning creditor adequately demonstrated that the petition satisfied the § 303(b) requirements).

⁷³ *See, e.g., In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1101 n.5 (11th Cir. 2008) (collecting cases in which courts concluded that § 303(b) is subject-matter jurisdictional without providing an explanation of why they did so), *overruled by* 550 F.3d 1035 (11th Cir. 2008) (en banc); *In re Quality Laser Works*, 211 B.R. 936, 941 (B.A.P. 9th Cir. 1997) (stating simply that “[i]t is well settled that the filing of an involuntary petition invokes the subject matter jurisdiction of the bankruptcy court if the petition is sufficient on its face and contains the essential allegations”); *In re Taylor & Assocs., L.P.*, 191 B.R. 374, 377 (Bankr. E.D. Tenn. 1996) (citing precedent from other bankruptcy courts and secondary authorities to conclude that “Courts have long recognized that the elements of Bankruptcy Code 303(b) are not prerequisites to establishing a bankruptcy court’s subject matter jurisdiction over proceedings arising from an involuntary petition”).

Trusted Net Media Holdings, LLC.⁷⁴ This part summarizes three cases in which courts have provided rationales for their differing conclusions. First, it looks at *In re Rubin*,⁷⁵ in which the Ninth Circuit concluded that Section 303(b) is not jurisdictional. Next, it presents *In re BDC 56 LLC*,⁷⁶ in which the Second Circuit held that Section 303(b) is subject-matter jurisdictional. Finally, it examines *In re Trusted Net Media Holdings*,⁷⁷ in which the Eleventh Circuit recognized the circuit split, performed a thoughtful analysis of both sides, and overruled an earlier case to hold that Section 303(b) does not pertain to subject matter jurisdiction.

A. *The Ninth Circuit's Analysis in In re Rubin*

*In re Rubin*⁷⁸ came before the Ninth Circuit Court of Appeals shortly after Congress passed the 1984 amendments to the Bankruptcy Code,⁷⁹ which added to Section 303 the new requirement that petitioning creditors' claims in an involuntary bankruptcy case not be "the subject of a bona fide dispute."⁸⁰ The debtor in the case, Rubin, submitted to the bankruptcy court a timely answer to an involuntary petition filed by ten creditors, in which he asserted that the claims alleged by the petitioning creditors were contingent and that the petition was filed in bad faith.⁸¹ A protracted and extensive discovery process ensued, during which the debtor repeatedly postponed depositions, produced thirty-three boxes of allegedly "disorganized and nonsensical" documents, and provided schedules of disputed claims that the bankruptcy court

⁷⁴ *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1041 (11th Cir. 2008) (en banc) (stating that "[m]ost other courts to consider the issue likewise have concluded that § 303(b)'s filing requirements are not subject matter jurisdictional," and listing relevant cases).

⁷⁵ 769 F.2d 611 (9th Cir. 1985).

⁷⁶ 330 F.3d 111 (2d Cir. 2003).

⁷⁷ 550 F.3d 1035 (11th Cir. 2008) (en banc), *overruling* 525 F.3d 1095 (11th Cir. 2008).

⁷⁸ 769 F.2d 611 (9th Cir. 1985).

⁷⁹ *See supra* Part I.B (chronicling the development of the Bankruptcy Code).

⁸⁰ *Rubin*, 769 F.2d at 614 & n.2; Bankruptcy Amendments and Federal Judgeship Act of 1984, Pub. L. No. 98-353, § 426(b), 98 Stat. 333, 369 (1984).

⁸¹ *In re Rubin*, 37 B.R. 232, 233 (B.A.P. 9th Cir. 1984). At the time that Rubin filed his answer with the bankruptcy court in 1982, the new bona fide dispute provision was not yet in effect. *Id.* at 232 (decided on February 29, 1984); Bankruptcy Amendments and Federal Judgeship Act of 1984, Pub. L. No. 98-353, 98 Stat. 333, 369, 392 (codified as amended at 11 U.S.C. § 303(b) (2006)) ("The amendments made [to § 303(b)] shall become effective upon the date of enactment of this Act[, July 10, 1984].").

repeatedly found “insufficient.”⁸² As a result, the bankruptcy court imposed sanctions on Rubin, “striking Rubin’s answer and entering an order for relief.”⁸³ The Bankruptcy Appellate Panel affirmed this order, and Rubin further appealed to the Ninth Circuit.⁸⁴

In his appearance before the Ninth Circuit, Rubin argued for the first time that the new “bona fide dispute” provision of the Bankruptcy Code was a jurisdictional requirement of an involuntary proceeding, and that his case should therefore be remanded to the bankruptcy court for a determination as to whether it had subject matter jurisdiction.⁸⁵ Although the Ninth Circuit ultimately did reverse and remand for a trial on the sufficiency of the involuntary petition, it did so based on a finding that the bankruptcy court’s discovery sanctions were an abuse of discretion—not on the jurisdictional basis that Rubin asserted.⁸⁶ Nevertheless, the court did engage in a jurisdictional analysis in order to establish its authority to reach the abuse of discretion issue.⁸⁷ The Ninth Circuit held that the undisputed-claims requirement of Section 303(b) was not jurisdictional in nature, but rather went “to the merits—an element that must be established to sustain an involuntary proceeding.”⁸⁸ In so doing, the court analogized this requirement to others in Section 303, which had been labeled as “jurisdictional” in prior cases, but were in fact treated as “substantive matters which must be proved or waived.”⁸⁹ The

⁸² *Rubin*, 769 F.2d at 613.

⁸³ *Id.*

⁸⁴ *Id.* at 613-14.

⁸⁵ *Id.* at 614. The bona fide dispute provisions became effective after the proceedings in the bankruptcy court but before those in the Ninth Circuit. *See supra* note 81.

⁸⁶ *Rubin*, 769 F.2d at 619.

⁸⁷ *Id.* at 614-15. The circuit court held that the bankruptcy court had jurisdiction independent of § 303(b), and that “[t]he bankruptcy court’s order striking Rubin’s answer and entering an order for relief was a final decision,” and thus that the circuit court had appellate jurisdiction to hear this case. *Id.* at 615.

⁸⁸ *Id.* at 614-15.

⁸⁹ *Id.* at 614 n.3. In *In re Mason*, the Ninth Circuit did not explicitly state that the § 303(b) requirements were not jurisdictional in nature, but it did hold that the petition’s failure to meet one of those requirements “did not deprive the bankruptcy court of jurisdiction to enter a valid order for relief,” when the debtor “waived his right to present this defense by failing to raise it in an answer to the petition.” 709 F.2d 1313, 1318-19 (9th Cir. 1983). In *In re Visioneering Construction*, a case similar to *Rubin* involving a debtor that allegedly obstructed discovery proceedings, the court held that the bankruptcy court did not abuse its discretion in imposing sanctions on the debtor by striking the debtor’s answer and ordering relief against the debtor. 661 F.2d 119, 123-24 (9th Cir. 1981). Discovery in that case was intended to help the

court also suggested that the nature of subject matter jurisdiction is such that a failure to satisfy the Section 303(b) requirements could not deprive the bankruptcy court of its already-existing power to hear the case.⁹⁰

B. The Second Circuit's Analysis in In re BDC 56 LLC

In *In re BDC 56 LLC*,⁹¹ the debtor, owner of the Chambers Hotel in Manhattan, successfully moved for dismissal of an involuntary petition filed by three construction companies that claimed to be owed money for work performed on the hotel.⁹² The bankruptcy court dismissed the petition based on BDC's assertion that two of the creditors' claims were subject to bona fide disputes⁹³ and that the third lacked

bankruptcy court determine whether it had subject matter jurisdiction over the case. *Visioneering*, 661 F.2d at 121. As a result of striking the debtor's answer, the bankruptcy court treated the allegations in the petition as admitted by the debtor, and found that those allegations "were sufficient to confer subject matter jurisdiction." *Id.* at 122. The *Rubin* court, in its analysis of *Visioneering*, observed that the notion of "conferring subject matter jurisdiction" by admission of the parties is entirely inconsistent with the Supreme Court's holding that "parties cannot confer subject matter jurisdiction on a federal court by their consent." *Rubin*, 769 F.2d at 614 n.3 (quoting *Insurance Corp. of Ireland v. Compagnie des Bauxite de Guinee*, 456 U.S. 694, 702 (1982)). Thus, despite the use of the term "subject matter jurisdiction" in *Mason* and *Visioneering*, the *Rubin* court held that those cases did not in fact establish that the § 303(b) requirements were anything other than "substantive matters which must be proved or waived." *Rubin*, 769 F.2d at 614 n.3.

⁹⁰ *Rubin*, 769 F.2d at 614 ("Subject matter jurisdiction deals with a court's competence to hear and determine cases of the general class to which the proceedings in question belong and the power to deal with the general subject involved in the action."). The court cited *In re Earl's Tire Service, Inc.*, in which a nonpetitioning creditor sought to have an involuntary petition against its debtor dismissed in order to prevent the trustee from voiding the creditor's collection activities. *Id.* (citing *In re Earl's Tire Serv., Inc.*, 6 B.R. 1019, 1020 (D. Del. 1980)). In order to get around its lack of standing to object to the petition, the creditor in *Earl* characterized its objection that there were an insufficient number of petitioning creditors as an attack on the court's subject matter jurisdiction. *Earl*, 6 B.R. at 1021. The *Earl* court observed that, since "Earl's Tire was qualified to be a debtor under the Bankruptcy Code, it is difficult to perceive how an arguable defect in the procedural mechanism for commencing a bankruptcy action would deprive the court of its subject matter jurisdiction." *Id.* at 1022. It went on to warn that courts' sometimes inaccurate use of the word "jurisdictional" does not provide a basis for "'jurisdictional' challenges raised by disgruntled creditors." *Id.* at 1023.

⁹¹ 330 F.3d 111 (2d Cir. 2003).

⁹² *Id.* at 114.

⁹³ *Id.* at 115 (debtor had "a longstanding dispute with [the first creditor] concerning its performance under the contract," and "contended that [the second creditor's] right to payment had not yet arisen under its contract"); see also 11 U.S.C. § 303(b); *supra* note 70.

standing.⁹⁴ After the creditors moved unsuccessfully for reconsideration and lost an appeal in the district court, they appealed to the Second Circuit.⁹⁵ At the circuit court, the parties argued for different standards of review—with BDC seeking review for “clear error” and the creditor-appellants urging “de novo review.”⁹⁶ Instead, the Second Circuit construed the Section 303(b) requirements as subject-matter jurisdictional and therefore applied plenary review.⁹⁷ In support of its holding that the Section 303(b) requirements were jurisdictional, the court stated that “[w]hether an alleged debtor is properly before the bankruptcy court in an involuntary case is a threshold determination that should be made at the earliest possible stage of the proceedings.”⁹⁸ The court cautioned that a result of failing to treat Section 303(b) as subject-matter jurisdictional would be that “creditors could, on the basis of relatively untested claims, haul a solvent debtor with whom they have legitimate disputes into bankruptcy court and force it to defend an involuntary proceeding while the bankruptcy court leaves for later merits determination whether the debtor is even properly before it.”⁹⁹ In addition to its own analysis, the court also relied on two previous holdings within the Second Circuit in which the courts repeatedly referred to Section 303(b) challenges as “jurisdictional.”¹⁰⁰

⁹⁴ *BDC 56*, 330 F.3d at 115 (the third creditor was a subcontractor of another party with whom debtor had contracted directly and to whom debtor had tendered complete payment).

⁹⁵ *Id.* at 116.

⁹⁶ *Id.* at 118.

⁹⁷ “When reviewing a district court’s determination of its subject matter jurisdiction, we review factual findings for clear error and legal conclusions *de novo*.” *Id.* at 119 (quoting *In re Vogel Van & Storage, Inc.*, 59 F.3d 9, 11 (2d Cir. 1995)); see also *Zappia Middle East Constr. Co. v. Emirate of Abu Dhabi*, 215 F.3d 247, 249 (2d Cir. 2000) (per curiam).

⁹⁸ *BDC 56*, 330 F.3d at 118.

⁹⁹ *Id.* at 118-19.

¹⁰⁰ *In re Elsa Designs, Ltd.*, 155 B.R. 859, 863 (Bankr. S.D.N.Y. 1993) (stating that the undisputed-claim requirement of § 303(b) “is both an element of the condition upon which a controverted order for relief may be entered and a necessary prerequisite for the bankruptcy court’s jurisdiction”); *In re Onyx Telecomm., Ltd.*, 60 B.R. 492, 495 (Bankr. S.D.N.Y. 1985) (stating that in a 12(b)(1) facial attack on an involuntary bankruptcy petition, “Section 303(b)(1) of the Bankruptcy Code is the applicable jurisdictional provision”). These cases do not explain the reasoning behind their conclusions that § 303(b) is jurisdictional. However, it is clear from both cases’ detailed discussions of the § 303(b) challenges that the court was indeed analyzing these challenges as subject-matter jurisdictional, and not merely making a careless “drive-by jurisdictional ruling[.]” *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 511 (2006) (quoting *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 91 (1998)); *Elsa*, 155 B.R. at 863, 864 n.2; *Onyx*, 60 B.R. at 493-97 (discussing at great length the difference between a

C. *The Eleventh Circuit Addresses the Split in In re Trusted Net Media Holdings, LLC*

In 2008, in *In re Trusted Net Media Holdings, LLC*,¹⁰¹ the Eleventh Circuit convened en banc to rehear an appeal from the lower court's denial of a motion to dismiss an involuntary bankruptcy petition.¹⁰² An involuntary Chapter 7 petition was filed in 2002 by a single creditor of Trusted Net.¹⁰³ After the debtor failed to respond, the bankruptcy court entered an order for relief.¹⁰⁴ Two years later, David W. Huffman, an officer of Trusted Net, moved to dismiss the petition on the basis that the single-creditor petition failed to meet the Section 303(b) requirements for subject matter jurisdiction because the claim was subject to a bona fide dispute and the debtor had twelve or more creditors.¹⁰⁵ Although the bankruptcy court denied this motion, no appeal was taken.¹⁰⁶ Two more years passed, at which point a number of Trusted Net's creditors reached a settlement with the trustee.¹⁰⁷ Huffman, whose deferred salary also made him a creditor of Trusted Net,¹⁰⁸ was not included in the settlement agreement and his objections to the settlement were overruled by the bankruptcy court.¹⁰⁹ Huffman then filed another motion to dismiss the case, in which he raised the same argument that he raised in 2004—namely, the lack of subject matter jurisdiction due to the use of a disputed claim and an insufficient number of petitioning creditors.¹¹⁰ In denying the motion, the bankruptcy court ruled that the requirements of Section 303(b) were not subject-matter jurisdictional and that the objection to the petition, raised more than four years after the commencement of the proceeding, had been waived by the

facial attack and a factual attack under 12(b)(1) before applying legal principles to the facts of that case).

¹⁰¹ 550 F.3d 1035 (11th Cir. 2008) (en banc).

¹⁰² *Id.* at 1037-38.

¹⁰³ *Id.* at 1037.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.* at 1037-38.

¹⁰⁶ *Id.* at 1038.

¹⁰⁷ *Id.*

¹⁰⁸ *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1097 (11th Cir. 2008), *rev'd*, 550 F.3d 1035 (11th Cir. 2008) (en banc).

¹⁰⁹ *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1038 (11th Cir. 2008) (en banc).

¹¹⁰ *Id.*

debtor.¹¹¹ When the district court affirmed, Trusted Net appealed to the Eleventh Circuit solely on the basis that the “requirements in Section 303(b) are jurisdictional, and thus cannot be waived.”¹¹² Despite concluding that Section 303(b) is properly construed as substantive rather than jurisdictional in nature, the Eleventh Circuit, found itself to be bound by contrary precedent.¹¹³ Consequently, with more than a bit of hesitation, the court reversed the lower court’s denial of Trusted Net’s motion to dismiss;¹¹⁴ however, the court subsequently vacated its decision¹¹⁵ and convened en banc to rehear the appeal.¹¹⁶

On rehearing, the Eleventh Circuit undertook a systematic analysis of the issue, looking not only at the “statutory framework for bankruptcy court jurisdiction and the commencement of involuntary bankruptcy cases,” but also the split between the Ninth and the Second Circuits.¹¹⁷ In its decision, the court concluded that Section 303(b) should not be treated as subject-matter jurisdictional for four main reasons: (1) there is no indication in the language of the provision that Congress intended to condition the court’s jurisdiction on satisfaction of the Section 303(b) requirements;¹¹⁸ (2) other

¹¹¹ *Id.*

¹¹² *Trusted Net*, 525 F.3d at 1097.

¹¹³ *Id.* at 1107 (finding itself bound by the precedent of *In re All Media Prop., Inc.*, 646 F.2d 193 (5th Cir. 1981), *aff’d* 5 B.R. 126 (Bankr. S.D. Tex. 1980)). In *All Media*, the former Fifth Circuit analyzed § 303 in one of the first cases to apply the then-new Bankruptcy Code in an involuntary proceeding. *All Media*, 5 B.R. at 131. Rather than explicitly stating that § 303(b) was jurisdictional, the *All Media* court made repeated reference to it as such. *Id.* at 133, 134, 138, 140, 142 (referring to various subsections of § 303 as jurisdictional). The *Trusted Net* court found that this treatment nevertheless qualified as a holding, because “a determination that § 303(b) is subject matter jurisdictional was a necessary predicate for the court’s consideration of [the debtor’s] argument—which was raised neither in the pleadings nor at trial—that the creditor . . . did not satisfy the statutory requirement of having an unsecured or undersecured claim.” *Trusted Net*, 525 F.3d at 1106-07.

¹¹⁴ *Trusted Net*, 525 F.3d at 1107.

¹¹⁵ *In re Trusted Net Media Holdings, LLC*, 530 F.3d 1363 (11th Cir. 2008).

¹¹⁶ *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1042 (11th Cir. 2008) (en banc) (“Because this Court sitting *en banc* is not bound by prior decisions of a panel of this Court or its predecessor, we need not revisit *All Media*. Instead, we reach our own conclusions as to the proper interpretation of § 303(b).” (internal citation omitted)).

¹¹⁷ *Id.* at 1038.

¹¹⁸ The court stated that “the language of § 303(b) does not evince a congressional intent to implicate the bankruptcy courts’ subject matter jurisdiction.” *Id.* at 1043. This conclusion was based, in part, on the Supreme Court’s opinion in *Arbaugh v. Y & H Corp.* See *infra* Part III.A.3. The *Trusted Net* court also noted that not only is there “no indication from the text of § 303 that Congress intended bankruptcy courts to consider *sua sponte* at any point in the proceedings whether the

similar provisions of the Bankruptcy Code have been interpreted as substantive rather than as jurisdictional;¹¹⁹ (3) this conclusion is consistent with “the bankruptcy-related jurisdictional grant in Title 28, as well as the basic nature of subject matter jurisdiction[;]”¹²⁰ and (4) this conclusion is consistent with the other provisions of Section 303.¹²¹ After the Eleventh Circuit’s en banc holding in *Trusted Net*, the Second Circuit stands alone in treating Section 303(b) as subject-matter jurisdictional in nature.¹²²

III. THE SECOND CIRCUIT ERRED IN FINDING SECTION 303(b) JURISDICTIONAL

The Section 303(b) requirements are best viewed as substantive rather than subject-matter jurisdictional. First, treating the Section 303(b) requirements as substantive better comports with the “basic nature of subject matter jurisdiction,”¹²³ the specific jurisdictional structure of the Bankruptcy Code, and the language and purpose of Section 303 itself. Additionally, the 2006 United States Supreme Court case, *Arbaugh v. Y & H Corp.*, definitively resolves this issue by establishing a test for determining whether a statute is jurisdictional or substantive in nature.¹²⁴ Second, a comparison between Section 303 and analogous provisions of the Bankruptcy Code that have been treated as either substantive or jurisdictional demonstrates that Section 303(b) should also be construed as nonjurisdictional for the sake of consistency. Third, the Second Circuit’s treatment of Section 303(b) as jurisdictional actually undercuts the court’s implied goals of fairness and judicial efficiency. In fact, a subsequent case

involuntary petition filing requirements have been met,” but that “the statutory language strongly suggests the opposite.” *Trusted Net*, 550 F.3d at 1044.

¹¹⁹ “[T]his Court has interpreted similar ‘commencement of the case’ language, found elsewhere in the Bankruptcy Code, to be non-jurisdictional.” *Trusted Net*, 550 F.3d at 1043.

¹²⁰ *Id.* at 1044.

¹²¹ *Id.* at 1044-45 (referring to § 303(c), (h)).

¹²² *See supra* note 13.

¹²³ *Id.* at 1044.

¹²⁴ *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 515-16 (2006) (holding that a statutory requirement should be treated as subject-matter jurisdictional only when Congress evinces a clear intent to make it so, and relying also in part on questions of fairness and judicial efficiency).

within the Second Circuit¹²⁵ demonstrates that the circuit's jurisdictional treatment of Section 303(b) is unworkable.

A. *Subject Matter Jurisdiction in the Bankruptcy Code:
Does Section 303 Belong?*

Treating Section 303(b) as jurisdictional conflicts with general notions of subject matter jurisdiction as well as the specific jurisdictional structure of the Bankruptcy Code, and prevents the other provisions of Section 303 from operating as Congress intended. Furthermore, this interpretation is a direct violation of the Supreme Court's holding in *Arbaugh v. Y & H Corporation*.¹²⁶

1. The Nature of Subject Matter Jurisdiction and Its
Place in the Bankruptcy Code

Subject matter jurisdiction refers to “the courts’ statutory or constitutional *power* to adjudicate the case.”¹²⁷ Because it goes to the fundamental ability of a court to entertain and adjudicate a proceeding, it is never too late to raise an objection based on lack of subject matter jurisdiction, even if the issue has not been introduced until appeal.¹²⁸ Nothing that the parties do in the course of litigation can serve to create jurisdiction that would otherwise be lacking.¹²⁹ A lack

¹²⁵ *In re MarketXT Holdings Corp.*, 347 B.R. 156 (Bankr. S.D.N.Y. 2006).

¹²⁶ *Arbaugh v. Y & H Corp.*, 546 U.S. 500 (2006).

¹²⁷ *Steel Co. v. Citizens for a Better Env't*, 523 U.S. 83, 89 (1998) (internal citation omitted) (emphasis in original); *In re Rubin*, 769 F.2d 611, 614 (9th Cir. 1985) (“Subject matter jurisdiction deals with a court’s competence to hear and determine cases of the general class to which the proceedings in question belong and the power to deal with the general subject involved in the action.” (internal citation omitted)); Howard M. Wasserman, *Jurisdiction, Merits, and Procedure: Thoughts on a Trichotomy*, 102 NW. U. L. REV. 1547, 1547-48 (2008) (“[Subject matter jurisdiction] can broadly be defined as the court’s raw, baseline power and legitimate authority to hear and resolve the legal and factual issues in a class of cases.”).

¹²⁸ FED. R. CIV. P. 12(b)(1) & 12(h); *Arbaugh*, 546 U.S. at 506 (“The objection that a federal court lacks subject-matter jurisdiction . . . may be raised by a party, or by a court on its own initiative, at any stage in the litigation, even after trial and the entry of judgment.”) (citing FED. R. CIV. P. 12(b)(1)); *Kontrick v. Ryan*, 540 U.S. 443, 455 (2004) (“A litigant generally may raise a court’s lack of subject-matter jurisdiction at any time in the same civil action, even initially at the highest appellate instance.” (internal citations omitted)). *But see Kontrick*, 540 U.S. at 455 n.9 (“Even subject-matter jurisdiction, however, cannot be attacked collaterally”); RESTATEMENT (SECOND) OF JUDGMENTS § 12 (1982) (listing the few circumstances in which subject matter jurisdiction may be attacked post-judgment).

¹²⁹ *Kontrick*, 540 U.S. at 456 (“Characteristically, a court’s subject-matter jurisdiction cannot be expanded to account for the parties’ litigation conduct”); *Ins.*

of subject matter jurisdiction can be raised by any party or the court sua sponte.¹³⁰ Once a federal court is found to lack subject matter jurisdiction, it is not within the court's discretion to retain the case.¹³¹

Like all federal courts, bankruptcy courts have limited subject matter jurisdiction,¹³² the scope and extent of which is defined by Congress.¹³³ It is well-established that Congress defined the subject matter jurisdiction of bankruptcy courts in sections 1334 and 157 of Title 28.¹³⁴ These sections provide that

Corp. of Ireland v. Compagnie des Bauxites de Guinee, 456 U.S. 694, 702 (1982) (“[N]o action of the parties can confer subject-matter jurisdiction upon a federal court.”).

¹³⁰ FED. R. CIV. P. 12(b)(1) & 12(h)(3) (“When it appears by suggestion of the parties or otherwise that the court lacks jurisdiction of the subject matter, the court shall dismiss the action.”); *Arbaugh*, 546 U.S. at 506 (“The objection that a federal court lacks subject-matter jurisdiction . . . may be raised by a party, or by a court on its own initiative, at any stage in the litigation, even after trial and the entry of judgment.”) (citing FED. R. CIV. P. 12(b)(1)); *Peretz v. United States*, 501 U.S. 923, 953 (1991) (Scalia, J., dissenting) (“One of the hoariest precepts in our federal judicial system is that a claim going to the court’s subject-matter jurisdiction may be raised at any point in the litigation *by any party*.”) (emphasis added).

¹³¹ *Compagnie des Bauxite*, 456 U.S. at 702 (“[T]he rule, springing from the nature and limits of the judicial power of the United States is inflexible and without exception, which requires this court, of its own motion, to deny its jurisdiction, and, in the exercise of its appellate power, that of all other courts of the United States, in all cases where such jurisdiction does not affirmatively appear in the record.” (quoting *Mansfield, C. & L.M. Ry. Co. v. Swan*, 111 U.S. 379, 382 (1884)); *Morrison v. Allstate Indem. Co.*, 228 F.3d 1255, 1261 (11th Cir. 2000) (“[L]ower federal courts are empowered to hear only cases for which there has been a congressional grant of jurisdiction, and once a court determines that there has been no grant that covers a particular case, the court’s sole remaining act is to dismiss the case for lack of jurisdiction.” (internal citation omitted))). Moreover, the question of whether subject matter jurisdiction is available is one for the court, and not a jury. 13 CHARLES A. WRIGHT, ARTHUR R. MILLER, EDWARD H. COOPER, & RICHARD D. FREER, *FEDERAL PRACTICE AND PROCEDURE* § 3522 (2008); Wasserman, *supra* note 127, at 1547-48 (“[T]he court resolves any factual issues on which jurisdiction turns.”).

¹³² *Morrison*, 228 F.3d at 1260-61.

¹³³ *Kline v. Burke Const. Co.*, 260 U.S. 226, 234 (1922) (“Only the jurisdiction of the Supreme Court is derived directly from the Constitution. Every other court created by the general government derives its jurisdiction wholly from the authority of Congress.”); *see also* 28 U.S.C. §§ 157 & 1334 (2006).

¹³⁴ *Kontrick*, 540 U.S. at 452-53 (“Only Congress may determine a lower federal court’s subject-matter jurisdiction Congress did so with respect to bankruptcy courts in Title 28”) (citation omitted); *In re Banks*, 235 Fed. Appx. 943, 944 (3d Cir. 2007) (“Two statutes, 28 U.S.C. §§ 1334 and 157, provide the source of a bankruptcy court’s jurisdiction.” (citation omitted)); *Valley Historic Ltd. P’ship v. Bank of New York*, 486 F.3d 831, 839 n.3 (4th Cir. 2007) (“Whether a bankruptcy court may exercise subject matter jurisdiction over a proceeding is determined by reference to 28 U.S.C. § 1334.”); *In re U.S. Brass Corp.*, 301 F.3d 296, 303 (5th Cir. 2002) (“The source of the bankruptcy court’s jurisdiction is 28 U.S.C. §§ 1334 and 157”) (quoting *United States Tr. v. Gryphon at the Stone Mansion, Inc.*, 216 B.R. 764, 769 (W.D. Pa. 1997), *aff’d*, 166 F.3d 552 (3d Cir. 1999)); 1 NORTON & NORTON, *supra* note 44, at § 4:4 (“The present Bankruptcy Code does not confer subject-matter jurisdiction, which is established solely by provisions of Title 28.”); *see also supra* Part I.B.

bankruptcy courts may exercise jurisdiction over “any or all cases under title 11,” pursuant to referral from the district court.¹³⁵ In contrast to the jurisdictional provisions of Title 28, Title 11 “contains the body of substantive law governing the federal bankruptcy regime.”¹³⁶ The practical effect of this statutory structure is that a bankruptcy court unquestionably has the jurisdiction to entertain an involuntary bankruptcy case, which by definition falls under Title 11.¹³⁷ In the course of exercising that jurisdiction, the bankruptcy court’s task is to determine whether the substantive requirements for bankruptcy relief are satisfied. The question of whether the substantive requirements of Title 11 are satisfied does not, in any case, affect the threshold determination that the court has the jurisdiction to hear and resolve the case.¹³⁸ The mere reference to Title 11 in the statutory provision that establishes bankruptcy jurisdiction¹³⁹ is not a ground for translating all of Title 11’s substantive requirements into jurisdictional requirements.¹⁴⁰

Given this jurisdictional and statutory framework, it is more sensible to conclude that Section 303(b) is unrelated to

¹³⁵ 28 U.S.C. § 157(a).

¹³⁶ Pathak, *supra* note 61, at 66 & n.21. *Cf. In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1098 (11th Cir. 2008) (describing Chapter 7, which defines liquidation, as the substantive provisions, and Chapter 3, which contains § 303, as “the procedural statute at issue”); *see also supra* Part I.B.

¹³⁷ 28 U.S.C. §§ 157 & 1334; 11 U.S.C. § 303; *In re Trusted Net Media Holdings, LLC*, No. 07-13429, 2008 WL 5069824, at *8 (11th Cir. Dec. 2, 2008) (“As a class of cases, involuntary bankruptcy cases unquestionably arise under Title 11 . . .”).

¹³⁸ *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 89 (1998); *Bell v. Hood*, 327 U.S. 678, 682 (1946). In *Bell*, the Supreme Court stated:

Jurisdiction, therefore, is not defeated . . . by the possibility that the averments might fail to state a cause of action on which petitioners could actually recover. For it is well settled that the failure to state a proper cause of action calls for a judgment on the merits and not for a dismissal for want of jurisdiction. Whether the complaint states a cause of action in which relief could be granted is a question of law and just as issues of fact it must be decided after and not before the court has assumed jurisdiction over the controversy. If the court does later exercise its jurisdiction to determine that the allegations in the complaint do not state a ground for relief, then dismissal of the case would be on the merits, not for want of jurisdiction.

Bell, 327 U.S. at 682.

¹³⁹ *See, e.g.*, 28 U.S.C. § 1334(a) (“[T]he district courts shall have original and exclusive jurisdiction of all cases under title 11.”); 28 U.S.C. § 157(a) (“Each district court may provide that any or all cases under title 11 and any or all proceedings arising under title 11 or arising in or related to a case under title 11 shall be referred to the bankruptcy judges for the district.”).

¹⁴⁰ *See, e.g., In re Bowshier*, 313 B.R. 232, 238 (Bankr. S.D. Ohio 2004) (“[I]t is important to note that not every statutory requirement is a matter of jurisdiction.”).

subject matter jurisdiction. First, Section 303(b) is codified within Title 11, which contains the substantive body of bankruptcy law, rather than in Title 28, which is the jurisdictional grant to federal courts.¹⁴¹ Second, Section 303 makes no reference to jurisdictional requirements.¹⁴² Third, treating Section 303(b) as jurisdictional could lead to the illogical result of incentivizing an involuntary debtor's default.¹⁴³ Consider, for example, a hypothetical case in which a single creditor files an involuntary bankruptcy petition against a debtor with more than twelve qualified creditors. If the debtor files an answer asserting that the petition fails to satisfy Section 303(b), he will then be required to supply the petitioning creditor with a list of his creditors' names and addresses and a description of their claims, in order for notice to be sent.¹⁴⁴ This allows the petitioning creditor to alert the other claimholders to the involuntary petition and gives those creditors an opportunity to join the petition with the same effect as if they were original petitioning creditors.¹⁴⁵ More likely than not, the requisite number of creditors will join the petition to ensure that they receive some part of the distribution of assets, and the debtor will lose his Section 303(b) jurisdictional defense.

Now consider a situation in which the same debtor fails to file a timely answer to the petition. Akin to a default judgment in a civil case,¹⁴⁶ if the debtor does not answer, the court must allow the bankruptcy case to proceed pursuant to

¹⁴¹ 11 U.S.C. § 303(b); *see supra* notes 62-63.

¹⁴² 11 U.S.C. § 303; *In re* Trusted Net Media Holding, 525 F.3d 1095, 1102 (11th Cir. 2008) ("Section 303(b) does not contain any explicit reference to its requirements being jurisdictional in nature."); *cf.* *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 502, 516 (2006) (holding that the employee-numerosity requirement of 42 U.S.C. § 2000e(b) is not jurisdictional, in part because "the 15-employee threshold appears in a . . . provision that 'does not speak in jurisdictional terms or refer in any way to the jurisdiction of the district courts.'") (quoting *Zipes v. Trans World Airlines, Inc.*, 455 U.S. 385, 394 (1982)).

¹⁴³ Admittedly, treating § 303(b) as substantive could lead to the undesirable result of allowing a creditor to force a debtor into bankruptcy on the basis of a single claim that would be better resolved through state collection procedures, or on the basis of a disputed claim. *See infra* Part III.C.2. However, this result is consistent with the judicial policy that a litigant's default may work to its detriment, and is more sensible than the alternative. *Id.*

¹⁴⁴ FED. R. BANKR. P. 1003(b).

¹⁴⁵ 11 U.S.C. § 303(c); FED. R. BANKR. P. 1003(b) and advisory committee note (d). This arrangement is sensible given that the debtor is the party with the most knowledge about his or her own financial affairs. *See In re Coppertone Commc'ns, Inc.*, 96 B.R. 233, 236 (Bankr. W.D. Mo. 1989).

¹⁴⁶ *See* FED. R. CIV. P. 55.

Section 303(h).¹⁴⁷ However, if Section 303(b) is treated as subject-matter jurisdictional, a default would actually be in the debtor's best interest since he could then move to dismiss the petition *after* the court has entered relief against him—still early enough to raise lack of subject matter jurisdiction as a defense but too late for additional creditors to join the petition.¹⁴⁸ In order to incentivize full disclosure by the debtor, it makes far more sense to treat the requirement of three or more petitioning creditors as a waivable affirmative defense, i.e., substantive rather than jurisdictional.¹⁴⁹ Under this approach, a debtor that fails to disclose the existence of claimholders gives up his Section 303(b) defense¹⁵⁰ and may be left with undischargeable debts if his creditors are not notified of the bankruptcy.¹⁵¹

2. A Non-Jurisdictional Interpretation Ensures that Section 303 Operates Effectively

When examined in conjunction with the other subsections of Section 303, it is plain that Section 303(b) must be treated as nonjurisdictional in order for the statute to operate sensibly.¹⁵² First, not only does a jurisdictional reading

¹⁴⁷ 11 U.S.C. § 303(h) (“If the petition is not timely controverted, the court shall order relief against the debtor in an involuntary case under the chapter under which the petition was filed.”); 2 NORTON & NORTON, *supra* note 44, at § 22:12.

¹⁴⁸ See 11 U.S.C. § 303(c) (“After the filing of a petition under this section *but before the case is dismissed or relief is ordered*, a creditor . . . may join in the petition with the same effect as if such joining creditor were a petitioning creditor”) (emphasis added). Moreover, while bankruptcy courts have broad powers to remedy bad faith conduct by litigants, see 11 U.S.C. § 105(a); 2 NORTON & NORTON, *supra* note 44, at § 13:4, they cannot use their equitable powers to expand the scope of their jurisdiction. 1 NORTON & NORTON, *supra* note 44, at § 4:5 (“The grant of equitable power to a bankruptcy court does not create, confer, or supply subject-matter jurisdiction if it is otherwise lacking.”). Thus, if § 303(b) was jurisdictional, the bankruptcy court would be unable to rely on its equitable powers to allow joinder of creditors after the entry of relief, as this would amount to a unilateral expansion of its jurisdiction. *Cf. Kontrick v. Ryan*, 540 U.S. 443, 454 (2004) (explaining that bankruptcy courts cannot use the Federal Rules of Bankruptcy Procedure to expand the scope of their jurisdiction); *In re Granger Garage, Inc.*, 921 F.2d 74, 77 (6th Cir. 1990) (“[Section] 105(a) [is not a] jurisdictional provision[.]. The subject matter jurisdiction of the bankruptcy court is limited to that which congress specifically grants.”).

¹⁴⁹ *In re Coppertone Commc'ns, Inc.* 96 B.R. 233, 236 (Bankr. W.D. Mo. 1989).

¹⁵⁰ *Id.*

¹⁵¹ See, e.g., 3 NORTON & NORTON, *supra* note 44, at § 57:20 (“Under Code § 523(a)(3), creditors who are neither listed by the debtor in the schedule of creditors filed with the court, nor who have otherwise learned of the bankruptcy case within a limited period of time, may have their claims excepted from discharge.”).

¹⁵² *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1102 (11th Cir. 2008) (“Interpreting § 303(b) as non-jurisdictional . . . results in a harmonious

of Section 303(b) incentivize a debtor's default when combined with Section 303(c),¹⁵³ but it also causes Section 303(c) to operate in contradiction of the general principles of subject matter jurisdiction.¹⁵⁴ If the bankruptcy court were to allow a nonpetitioning creditor to "join in the petition with the same effect as if such joining creditor were a petitioning creditor under [Section 303(b)],"¹⁵⁵ it would essentially be acting so as to confer jurisdiction upon itself, thus violating a rule that the Supreme Court has described as "inflexible."¹⁵⁶ By contrast, if a Section 303(b) defect is merely a substantive failure, then permitting a creditor to join the petition would effectuate Section 303(c) while adhering to the rules of subject matter jurisdiction.

Next, a jurisdictional reading of Section 303(b) would also cause Section 303(d) to violate general principles of subject matter jurisdiction. Section 303(d)'s limitation on who may file an answer to an involuntary petition¹⁵⁷ has been interpreted as an exhaustive list.¹⁵⁸ This list indicates that creditors, including

operation of the statutory subsections."), *overruled by* 550 F.3d 1035 (11th Cir. 2008) (en banc).

¹⁵³ See *supra* Part III.A.1.

¹⁵⁴ See *id.* (discussing general principles of subject matter jurisdiction); *In re* Trusted Net Media Holdings, LLC, 550 F.3d 1035, 1044-45 (11th Cir. 2008) (en banc) ("[I]t seems anomalous at best to conclude that a bankruptcy court, which lacks jurisdiction over an involuntary case because the petition was defectively filed, subsequently may create jurisdiction for itself by permitting additional creditors to join the petition [under § 303(c)].").

¹⁵⁵ 11 U.S.C. § 303(c).

¹⁵⁶ *Ins. Corp. of Ireland v. Compagnie des Bauxite de Guinee*, 456 U.S. 694, 702 (1982) ("[T]he rule, springing from the nature and limits of the judicial power of the United States is inflexible and without exception, which requires this court, of its own motion, to deny its jurisdiction, and, in the exercise of its appellate power, that of all other courts of the United States, in all cases where such jurisdiction does not affirmatively appear in the record.") (quoting *Mansfield, C. & L.M. Ry. Co. v. Swan*, 111 U.S. 379, 382 (1884)); *Morrison v. Allstate Indem. Co.*, 228 F.3d 1255, 1261 (11th Cir. 2000) ("[L]ower federal courts are empowered to hear only cases for which there has been a congressional grant of jurisdiction, and once a court determines that there has been no grant that covers a particular case, the court's sole remaining act is to dismiss the case for lack of jurisdiction.").

¹⁵⁷ 11 U.S.C. § 303(d) ("The debtor, or a general partner in a partnership debtor that did not join in the petition, may file an answer to a petition under this section.").

¹⁵⁸ See, e.g., FED. R. BANKR. P. 1011(a), (e) ("The debtor named in an involuntary petition . . . may contest the petition. . . . No other pleadings shall be permitted . . ."); *In re* MarketXT Holdings Corp., 347 B.R. 156, 160 (Bankr. S.D.N.Y. 2006) (interpreting § 303(d) to mean that "only [t]he debtor, or a general partner in a partnership debtor that did not join in the petition, may file an answer") (emphasis added); *In re* Taylor & Assocs., L.P., 191 B.R. 374, 378-79, 381 (Bankr. E.D. Tenn. 1996) (this rule "prohibit[s] creditors from contesting an involuntary petition in order to prevent creditors from protecting a preference or retaining some other unfair

petitioning creditors who are undoubtedly parties to the litigation, are not authorized to raise objections to an involuntary petition based on a Section 303(b) deficiency.¹⁵⁹ It is fundamental, however, that any party may raise an objection based on lack of subject matter jurisdiction.¹⁶⁰ Thus, if Section 303(b) is jurisdictional in nature, one of the most basic and longstanding features of subject matter jurisdiction would not apply. Rather than creating an unprecedented exception to the well-accepted principles of subject matter jurisdiction, the more logical approach is to interpret Section 303(b) in a manner that is consistent with both the principles of subject matter jurisdiction and the statutory structure to which Section 303(b) belongs.¹⁶¹ This approach leads to the conclusion that Section 303(b) is nonjurisdictional.

3. Distinguishing Jurisdictional and Substantive Statutory Provisions Under *Arbaugh*: Is the Split Over Section 303(b) Moot?

Part of the difficulty in characterizing any statutory provision as jurisdictional or substantive stems from the frequent and longstanding misuse of the word “jurisdiction” by courts.¹⁶² In 2006, the Supreme Court addressed this issue in *Arbaugh v. Y & H Corp.*,¹⁶³ when the Court faced the question of whether an employee-numerosity requirement in Title VII was

advantage”); *In re Westerleigh Development Corp.*, 141 B.R. 38, 40 (S.D.N.Y. 1992); *In re New Era Co.*, 115 B.R. 41, 44-45 (Bankr. S.D.N.Y. 1990) (listing cases in support of this proposition).

¹⁵⁹ FED. R. BANKR. P. 1011. The rationale for this rule is that “a creditor may have an incentive to protect a preference or to gain some unfair advantage at the expense of other creditors, contrary to the policy of requiring an equitable distribution of the debtor’s assets among all creditors.” *New Era*, 115 B.R. at 45.

¹⁶⁰ See *supra* note 130.

¹⁶¹ *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1102 (11th Cir. 2008) (“Interpreting § 303(b) as non-jurisdictional, on the other hand, results in a harmonious operation of the statutory subsections.”).

¹⁶² *Kontrick v. Ryan*, 540 U.S. 443, 454 (2004) (“Courts, including this Court, it is true, have been less than meticulous in this regard; they have more than occasionally used the term ‘jurisdictional’ to describe emphatic time prescriptions in rules of court.”); *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 90 (1998) (“‘Jurisdiction,’ it has been observed, ‘is a word of many, too many, meanings’”); *Da Silva v. Kinsho Intern. Corp.*, 229 F.3d 358, 361 (2d Cir. 2000) (“Court decisions often obscure the issue by stating that the court is dismissing for ‘lack of jurisdiction’ when some threshold fact has not been established, without explicitly considering whether the dismissal should be for lack of subject matter jurisdiction or for failure to state a claim.”); *United States v. Wey*, 895 F.2d 429, 431 (7th Cir. 1990) (“As for ‘jurisdiction’: the word is a many-hued term”).

¹⁶³ 546 U.S. 500 (2006).

subject-matter jurisdictional or substantive.¹⁶⁴ In *Arbaugh*, the plaintiff brought a Title VII suit against her employer alleging sexual harassment.¹⁶⁵ Two weeks after the trial court entered judgment for the plaintiff, the defendant moved to dismiss for lack of subject matter jurisdiction on the basis that the word “employer,” as defined under Title VII, included only those people who have “fifteen or more employees.”¹⁶⁶ The defendant asserted that he employed fewer than fifteen people and, therefore, the court lacked jurisdiction over the case.¹⁶⁷ In its decision, the Supreme Court held that courts should construe statutory requirements as nonjurisdictional unless Congress makes it clear that the requirement is intended to function as a jurisdictional limitation.¹⁶⁸ The Court also addressed the tendency of lower courts to carelessly label dismissals as “jurisdictional” when they were in fact based on a party’s failure to establish a substantive element of its claim,¹⁶⁹ and characterized “such unrefined dispositions as ‘drive-by jurisdictional rulings’ that should be accorded ‘no precedential effect’ on the question whether the federal court had authority to adjudicate the claim in suit.”¹⁷⁰ In addition to the main legislative intent test, the *Arbaugh* Court also mentioned the “‘unfair[ness]’ and ‘waste of judicial resources’” that would result from construing the employee numerosity requirement as jurisdictional, as factors in its decision.¹⁷¹

When the Eleventh Circuit first addressed the question of how to characterize Section 303(b) in *Trusted Net*, it held that it was not governed by *Arbaugh*.¹⁷² The original *Trusted Net* court found that because *All Media* explicitly treated

¹⁶⁴ *Id.* at 503.

¹⁶⁵ *Id.* at 503-04.

¹⁶⁶ 42 U.S.C. § 2000e(b); *Arbaugh*, 546 U.S. at 503-04.

¹⁶⁷ *Id.* at 504.

¹⁶⁸ *Id.* at 509, 515-16. The *Arbaugh* Court stated:

If the Legislature clearly states that a threshold limitation on a statute’s scope shall count as jurisdictional, the courts and litigants will be duly instructed and will not be left to wrestle with the issue. . . . But when Congress does not rank a statutory limitation on coverage as jurisdictional, courts should treat the restriction as nonjurisdictional in character.

Id. at 515-16. The Court also provided a nonexhaustive list of statutes in which Congress clearly stated its intent that a requirement be jurisdictional. *Id.* at 516 n.11.

¹⁶⁹ *Id.* at 511.

¹⁷⁰ *Id.* (quoting *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 91 (1998)).

¹⁷¹ *Id.* at 515.

¹⁷² *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1104 n.11 (11th Cir. 2008), *overruled by* 550 F.3d 1035 (11th Cir. 2008) (en banc).

Section 303(b) as jurisdictional, as opposed to simply labeling it as such, the decision was not a “drive-by jurisdictional ruling”¹⁷³ and therefore must be treated as binding precedent.¹⁷⁴ When the Eleventh Circuit subsequently reheard *Trusted Net* en banc, however, it overruled *All Media* after applying the *Arbaugh* factors and finding that the Section 303(b) requirements are in fact not subject-matter jurisdictional.¹⁷⁵ Consistent with the main test articulated in *Arbaugh*, the en banc panel focused primarily on the failure of Section 303(b) to “speak in jurisdictional terms.”¹⁷⁶ Although the *Trusted Net* court did not address the question of “‘unfair[ness]’ and ‘waste of judicial resources,’”¹⁷⁷ such considerations would have also militated in favor of its conclusion.¹⁷⁸ Indeed, a jurisdictional reading of Section 303(b) would allow the debtor to strategically default so as to prevent the petitioning creditors from curing a defective petition,¹⁷⁹ a blatantly unfair strategy. It would also allow creditors to squirrel away jurisdictional objections to be used in the event that the involuntary bankruptcy case does not appear to be progressing in their favor,¹⁸⁰ thus wasting the court’s time and depleting the debtor’s estate.¹⁸¹ In the face of *Trusted Net*’s well-reasoned application of *Arbaugh* and the fairness and efficiency considerations

¹⁷³ *Id.*; *Steel Co.*, 523 U.S. at 91.

¹⁷⁴ *Trusted Net*, 525 F.3d at 1104 n.11. In fact, this panel could have overruled the *All Media* precedent without having to convene en banc. *Id.* at 1104 n.7 (“We have . . . held that when an earlier panel of this court has adopted a lower court’s order, that order is binding precedent *unless and until overruled by the Supreme Court* or this Court sitting *en banc*.”) (first emphasis added). While the Supreme Court did not expressly hold that § 303(b) was nonjurisdictional, it did issue a clear directive to courts that, “when Congress does not rank a statutory limitation on coverage as jurisdictional, courts should treat the restriction as nonjurisdictional in character.” *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 516 (2006).

¹⁷⁵ *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1042-43 (11th Cir. 2008) (en banc). “Applying the Supreme Court’s . . . recent *Arbaugh* test, § 303(b)’s requirements are not subject matter jurisdictional,” because “the language of § 303(b) does not evince a congressional intent to implicate the bankruptcy courts’ subject matter jurisdiction.” *Id.* at 1046.

¹⁷⁶ *Trusted Net*, 550 F.3d at 1043 (quoting *Arbaugh*, 546 U.S. at 515).

¹⁷⁷ *Arbaugh*, 546 U.S. at 515.

¹⁷⁸ The emphasis on efficiency, demonstrated in FED. R. BANKR. P. 1001, also applies in the context of involuntary bankruptcy cases. *See supra* note 21; FED. R. BANKR. P. 1013 (“The court shall determine the issues of a contested [involuntary] petition at the earliest practicable time . . .”).

¹⁷⁹ *See supra* Part III.A.1.

¹⁸⁰ *See supra* notes 158-159.

¹⁸¹ 11 U.S.C. § 503 (providing for certain bankruptcy related expenses to be paid out of the debtor’s estate).

suggested by the *Arbaugh* Court, the Second Circuit's position is now even more tenuous.

B. *Analogous Provisions in the Bankruptcy Code*

Like Section 303(b), other provisions of the Bankruptcy Code have been the subject of debate regarding whether they are jurisdictional or substantive in nature. Significantly, most of these provisions have been deemed to be nonjurisdictional. A brief look at the rationales provided for some of these provisions suggests that the debate over Section 303(b) should be similarly resolved.¹⁸²

1. 11 U.S.C. §§ 546(a) and 549(d): Time Limit on Adversary Proceedings

Sections 546(a) and 549(d) of the Bankruptcy Code establish a time limit after which a trustee can no longer bring certain adversary proceedings to recover property of the debtor's estate that has been transferred away.¹⁸³ In *In re Pugh*, the Eleventh Circuit held that these sections were not jurisdictional, but rather waivable statutes of limitations.¹⁸⁴ In *Pugh*, the debtors did not raise the untimeliness of the trustee's adversary proceeding in their response since they asserted that it was jurisdictional and therefore could be raised at any time.¹⁸⁵ In refuting this interpretation, the court relied on the "plain language of the provisions themselves," the overall statutory

¹⁸² Cf. Wasserman, *supra* note 127, at 1547-49 ("[I]f only some jurisdictional grants are bound up with the merits, there is no explanation or justification for why some merits issues should be jurisdictional and others not.")

¹⁸³ 11 U.S.C. §§ 546(a) & 549(d). The trustee's ability to void transfers of a debtor's property is referred to as his "avoiding power." See, e.g., 1 NORTON & NORTON, *supra* note 44, at § 3:11. The trustee uses this power in order to maximize the value of the debtor's estate for distribution to creditors, and to prevent preferential treatment of favored creditors. *Id.* at § 22:12.

¹⁸⁴ *In re Pugh*, 158 F.3d 530, 530 (11th Cir. 1998). The court characterized the issue in this case as

whether these code provisions constitute grants of subject matter jurisdiction that leave a court without any authority to hear certain proceedings—i.e., that extinguish the right of action itself by divesting a court of its subject matter jurisdiction over certain proceedings—after the limitations period has elapsed, or whether they are true statutes of limitations that restrict the power of a court to grant certain remedies in a proceeding over which it has subject matter jurisdiction.

Id. at 533-34.

¹⁸⁵ *Id.* at 532.

scheme, decisions of other courts, and the legislative history in concluding that the limitations were not jurisdictional in nature.¹⁸⁶ In its rejection of the alternative view, the *Pugh* court noted that the key precedent in support of that approach was “devoid of analysis”¹⁸⁷ and relied on the faulty assumption that a limitation on a cause of action is automatically a limitation on a court’s subject matter jurisdiction over that cause of action.¹⁸⁸ As the Eleventh Circuit noted in *Trusted Net*, “[t]he reasons in *Pugh* apply equally to Section 303(b).”¹⁸⁹

2. 11 U.S.C. § 109(e): Limits on Amount of Debt to Qualify for Chapter 13 Relief

Under Section 109(e), the Code places a statutory cap on the amount of debt an individual can owe and still be a Chapter 13 debtor.¹⁹⁰ In *Rudd v. Laughlin*,¹⁹¹ the bankruptcy trustee alleged that the debtors had abused the bankruptcy system by filing six Chapter 13 petitions in a six-year period, despite their inability to qualify as Chapter 13 debtors under Section 109(e) due to the amount of their unsecured debt.¹⁹² In response to the trustee’s attempt to convert their case into a

¹⁸⁶ *Id.* at 538. This is the majority view. See, e.g., *In re Outboard Marine Corp.*, 299 B.R. 488, 496 (Bankr. N.D. Ill. 2003) (“The clear weight of recent authority bolsters the conclusion that § 546(a) is [nonjurisdictional].”); *In re Commercial Fin. Servs., Inc.*, 294 B.R. 164, 174 (Bankr. N.D. Okla. 2003) (“The court finds the analysis in *Pugh* to be persuasive.”); *In re Rodriguez*, 283 B.R. 112, 120 (Bankr. E.D.N.Y. 2001) (“Based on the *Pugh* case and the decisions cited in *Pugh*, the Court finds that Section 546(a) is [nonjurisdictional].”); *In re Klayman*, 228 B.R. 805, 806 (Bankr. M.D. Fla. 1999) (“The case of [*Pugh*] . . . follows the majority view.”).

¹⁸⁷ *Pugh*, 158 F.3d at 535-36 (“[T]he few other courts that have adopted this jurisdictional view offer little analysis to support their position.”). The Eleventh Circuit made the same observation with regard to § 303(b) in *Trusted Net*, where it noted that “[a]lthough some bankruptcy courts earlier had reached the same conclusion as the Second Circuit [in *BDC*], that § 303(b) is subject matter jurisdictional, they did so without explanation.” *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1101 n.5 (11th Cir. 2008), *overruled by* 550 F.3d 1035 (11th Cir. 2008) (en banc). This pattern lends credence to the Supreme Court’s concern over “drive-by jurisdictional rulings.” *Steel Co. v. Citizens for a Better Env’t*, 523 U.S. 83, 91 (1998); *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 511 (2006).

¹⁸⁸ *Pugh*, 158 F.3d at 535-36.

¹⁸⁹ *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1044 (11th Cir. 2008) (en banc). Significantly, at least one bankruptcy court within the Second Circuit has adopted *Pugh*’s reasoning. *In re Rodriguez*, 283 B.R. 112, 119-20 (Bankr. E.D.N.Y. 2001).

¹⁹⁰ 11 U.S.C. § 109(e). The current amounts are \$336,900 for unsecured debt, and \$1,010,650 for secured debt, and are subject to adjustment every three years. 11 U.S.C. §§ 104, 109(e).

¹⁹¹ 866 F.2d 1040 (8th Cir. 1989).

¹⁹² *Id.* at 1041.

Chapter 7 liquidation, the debtors asserted that the bankruptcy court lacked subject matter jurisdiction over the case due to the Section 109(e) deficiency.¹⁹³ When the Eighth Circuit Court of Appeals affirmed the conversion to Chapter 7, it agreed with the district court that the Section 109(e) deficiency was more akin to a failure to state a claim than a jurisdictional defect.¹⁹⁴ The circuit court noted that the congressional grant of subject matter jurisdiction over bankruptcy cases comes from 28 U.S.C. §§ 1334 and 157 and, unlike the requirements for diversity jurisdiction in 28 U.S.C. § 1332, makes no reference to an amount in controversy.¹⁹⁵ This analysis applies with equal force to Section 303(b). To the extent that the *Rudd* court rejected the notion that “a case filed by an ineligible debtor is a nullity, and the court has no jurisdiction to convert the nonexistent case to another chapter,”¹⁹⁶ it would be illogical to conclude that an involuntary case filed by an insufficient number of creditors is incapable of being cured.¹⁹⁷ Instead, it more closely resembles a failure to state a claim, which does not implicate subject matter jurisdiction.¹⁹⁸

3. 11 U.S.C. § 109(h)(1): Credit Counseling Requirement

Both Section 301 (governing voluntary petitions) and Section 303 (governing involuntary petitions) state that a bankruptcy case can only be commenced by or against one who

¹⁹³ *Id.*

¹⁹⁴ *Id.* at 1041-42.

¹⁹⁵ *Id.* The *Rudd* court relied in part on a Fifth Circuit case in which the court recognized a split in authority over whether § 109 was jurisdictional. *Promenade Nat'l Bank v. Phillips*, 844 F.2d 230, 236 n.2 (5th Cir. 1988). The Fifth Circuit noted that “the courts holding that the issue is not jurisdictional generally have engaged in an analysis of the issue, while the courts holding that it is a matter of jurisdiction have not.” *Promenade*, 844 F.2d at 235 n.2. Once again, the Supreme Court’s concern about “drive-by jurisdictional rulings” appears well-founded. *Steel Co. v. Citizens for a Better Env't*, 523 U.S. 83, 91 (1998); *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 511 (2006); *supra* note 187.

¹⁹⁶ *Rudd*, 866 F.2d at 1041.

¹⁹⁷ At least one bankruptcy court within the Second Circuit agrees with the reasoning in *Pugh*. *In re Toronto*, 165 B.R. 746, 756 (Bankr. D. Conn. 1994) (“Section 109(e) relates to the eligibility of a debtor for chapter 13 relief, not the jurisdiction of the court.”); *cf.* *Cavaliere v. Sapir*, 208 B.R. 784, 786 (D. Conn. 1997) (describing § 109(e) as jurisdictional, though without any meaningful analysis); *In re Rifkin*, 124 B.R. 626, 629 (Bankr. E.D.N.Y. 1991) (same); *In re Jerome*, 112 B.R. 563, 564 (Bankr. S.D.N.Y. 1990) (same).

¹⁹⁸ *Cf. Rudd*, 866 F.2d at 1041.

“may be a debtor.”¹⁹⁹ In 2005, amendments to the Bankruptcy Code added the requirement that an individual must receive credit counseling within the 180 days preceding the bankruptcy petition in order to be a debtor.²⁰⁰ In the context of a voluntary Chapter 13 petition where the debtor did not complete his credit counseling within the allotted time, one bankruptcy court has held that it “simply lacks jurisdiction over a debtor’s case where the debtor fails to comply with [the credit counseling requirement].”²⁰¹ For debtors facing involuntary bankruptcy, this holding must have prompted shouts of joy—if they simply refused credit counseling, they could not qualify as debtors under Title 11 and the court would have to dismiss the involuntary petition for lack of subject matter jurisdiction. Not surprisingly, later courts have rejected this contention both on the basis of its sheer absurdity²⁰² and statutory construction.²⁰³ Nonetheless, it provides a good example of a nonjurisdictional²⁰⁴ provision within Title 11 that cannot be logically distinguished from the statutory requirements imposed by Section 303(b).

¹⁹⁹ 11 U.S.C. §§ 301(a), 303(a) (2006).

²⁰⁰ 11 U.S.C. § 109(h)(1) (“[A]n individual may not be a debtor under this title unless such individual has, during the 180-day period preceding the date of filing of the petition by such individual, received . . . credit counseling . . .”).

²⁰¹ *In re Giles*, 361 B.R. 212, 214 (Bankr. D. Utah 2007).

²⁰² *In re Allen*, 378 B.R. 151, 153 (Bankr. N.D. Tex. 2007) (“While the court recognizes that § 303(a) requires that a person who is the subject of an involuntary case qualify as a debtor, interpreting this provision as requiring that an involuntary debtor comply with section 109(h)(1) would lead to an absurd result.”).

²⁰³ *Id.* at 153 (finding that because “[t]he statutory language of section 109(h)(1) requires that the credit counseling occur before ‘the filing of the petition by such individual,’” the requirement only applied to voluntary cases) (emphasis in original).

²⁰⁴ That the credit counseling requirement is unrelated to subject matter jurisdiction is further evidenced by the fact that there are bases, albeit limited ones, upon which a court may waive the requirement. 11 U.S.C. § 109(h)(3)(A). The ability of the court to waive the credit counseling requirement is fundamentally inconsistent with the conclusion that this requirement is subject-matter jurisdictional. *Morrison v. Allstate Indem. Co.*, 228 F.3d 1255, 1261 (11th Cir. 2000) (once a court determines that it lacks subject matter jurisdiction, “the court’s sole remaining act is to dismiss the case for lack of jurisdiction”).

C. *Breaking Down BDC*

1. The Second Circuit's Argument "fails on its own terms"²⁰⁵

Perhaps one of the biggest weaknesses of the Second Circuit's holding in *BDC* that Section 303(b) is jurisdictional is its surprisingly superficial reasoning. As the *Trusted Net* court pointed out in its first analysis of Section 303(b), the Second Circuit failed to address the interaction of Section 303(b) with the other subsections of Section 303 and "ignore[d] the fact that subject matter jurisdiction turns only upon whether the court has the statutorily-conferred power to hear the case before it, and therefore has nothing to do with the speedy determination of claims or whether an alleged debtor—or any other party—is 'properly before the . . . court.'"²⁰⁶ To the extent that the Second Circuit specifically considered the proper characterization of Section 303(b) and explicitly held that it is jurisdictional in nature,²⁰⁷ its holding can hardly be called a "drive-by jurisdictional ruling[]." ²⁰⁸ At the same time, it also falls short of the detailed treatment given by courts that have found Section 303(b) to be nonjurisdictional.²⁰⁹ The *BDC* court provided only a summary rationale for its conclusion, stating that "[w]hether an alleged debtor is properly before the bankruptcy court in an involuntary case is a threshold determination that should be made at the earliest possible stage of the proceedings," so that creditors cannot "haul a solvent debtor with whom they have legitimate disputes into bankruptcy court and force it to defend an involuntary bankruptcy proceeding while the bankruptcy court leaves for a later merits determination whether the debtor is even properly before it."²¹⁰ While the propriety of the involuntary petition is undoubtedly something that should be

²⁰⁵ *In re Trusted Net Media Holdings, LLC*, 525 F.3d 1095, 1102 (11th Cir. 2008).

²⁰⁶ *Id.* at 1102 (quoting *In re BDC 56 LLC*, 330 F.3d 111, 118 (2d Cir. 2003)).

²⁰⁷ *BDC 56*, 330 F.3d at 118 ("We believe the more sound view is that the [§ 303(b) undisputed claim] requirement is subject matter jurisdictional, and now so hold.").

²⁰⁸ *Arbaugh v. Y & H Corp.*, 546 U.S. 500, 511 (2006) (quoting *Steel Co. v. Citizens for a Better Env't*, 523 U.S. 83, 91 (1998)).

²⁰⁹ See, e.g., *Trusted Net*, 525 F.3d at 1100-04; *In re Saunders*, 379 B.R. 847, 855-57 (Bankr. D. Minn. 2007); *In re Earl's Tire Svc., Inc.*, 6 B.R. 1019, 1021-23 (D. Del. 1980).

²¹⁰ *BDC 56*, 330 F.3d at 118.

resolved as soon as possible,²¹¹ labeling Section 303(b) as jurisdictional allows precisely the opposite to occur since a lack of subject matter jurisdiction can be raised at any time.²¹² The *BDC* court failed to address this issue.

The unworkable nature of the *BDC* holding became evident in *In re MarketXT Holdings*,²¹³ a 2006 decision by a bankruptcy court in the Second Circuit. In *MarketXT*, the court entered involuntary Chapter 11 relief against a debtor who failed to file a motion opposing the petition.²¹⁴ Approximately six months later, a nonpetitioning creditor sought to have the case against the debtor dismissed for lack of subject matter jurisdiction, alleging that the petitioning creditor's claim was subject to a bona fide dispute.²¹⁵ While maintaining that it was adhering to the Second Circuit's ruling that Section 303(b) is jurisdictional,²¹⁶ the bankruptcy court nevertheless denied the creditor's motion to dismiss, relying on Section 303(d)'s provision that only the debtor may file an answer to an involuntary petition.²¹⁷ Although the court acknowledged the circuit's position that Section 303(b) is subject-matter jurisdictional,²¹⁸ it simultaneously stripped the provision of two

²¹¹ *Id.*; FED. R. BANKR. P. 1001.

²¹² *Trusted Net*, 525 F.3d at 1102 (“*BDC*’s rationale also fails on its own terms . . . [because] if § 303(b)’s requirements are subject matter jurisdictional, an involuntary debtor could raise a § 303(b) challenge at any point in the proceedings, whereas if § 303(b) is non-jurisdictional, § 303(h) and Rule 1013 would require that the issue of the petitioning creditors’ compliance with § 303(b) be determined at the outset—as a threshold matter—or be forever waived.”). Admittedly, while the nonjurisdictional approach would put the issue of the petition’s propriety to rest if not raised within a certain period of time, it would do so at the expense of an actual determination of that issue. However, there are other provisions of the Bankruptcy Code that provide the debtor with protection against frivolous involuntary petitions, such that it is not necessary to rely on a jurisdictional reading of § 303(b) to accomplish this goal. See *infra* Part III.C.2.

²¹³ 347 B.R. 156 (Bankr. S.D.N.Y. 2006).

²¹⁴ *Id.* at 158.

²¹⁵ *Id.* at 158-59.

²¹⁶ *Id.* at 160. Interestingly, the *MarketXT* court failed to even acknowledge *United Marine, LLC v. Just for Windows, Inc.*, decided one year before *BDC*, in which the court recognized that “[t]he weight of the authority clearly supports the holding that the § 303(b) requirements are not jurisdictional and that the affirmative defense that a petition does not comply with such requirements may be waived pursuant to § 303(h).” No. 01 Civ. 5066(HB), 2002 WL 72933, at *2 (Bankr. S.D.N.Y. Jan. 17, 2002). In *United Marine*, the court found that the debtor had waived his right to challenge the petition based on an insufficient number of petitioning creditors and failure to allege unsecured claims in the required aggregate amount. *Id.* at *1.

²¹⁷ *MarketXT*, 347 B.R. at 160; see 11 U.S.C. § 303(d) (2006) (“The debtor, or a general partner in a partnership debtor that did not join in the petition, may file an answer to a petition under this section.”); see also *supra* Part III.A.2.

²¹⁸ *MarketXT*, 347 B.R. at 160.

of the most fundamental characteristics of subject matter jurisdiction—namely, that it can be raised by any party, at any time, and that when it is lacking, it cannot be conferred by the actions of the parties or the court.²¹⁹ The court’s justification was that this interpretation was necessary to preserve the function of Section 303(d).²²⁰ Interestingly, the court held that the Second Circuit’s policy of determining “whether an alleged debtor is properly before the bankruptcy court in an involuntary proceeding” as early as possible is carried out by Section 303(h),²²¹ which directs a bankruptcy court to enter relief against a debtor who fails to controvert an involuntary petition.²²² Accordingly, there should be no need to achieve this objective by calling Section 303(b) jurisdictional. In fact, treating it as such would actually undermine the goal of resolving the propriety of the petition as soon as possible.²²³

In the end, the only sensible part of the *MarketXT* opinion is the outcome.²²⁴ It is apparent and well-recognized that a creditor is not able to move for dismissal of an

²¹⁹ *Id.*; see *supra* Part III.A.1 (discussing the fundamental rules of subject matter jurisdiction).

²²⁰ *MarketXT*, 347 B.R. at 160; see 11 U.S.C. § 303(d). The bankruptcy court acknowledged that subject matter jurisdiction can be challenged by any party but stated that this was “no justification for invalidating another part of the same statute [i.e., § 303(d)].” *MarketXT*, 347 B.R. at 160. The court also stated that “[n]othing in *BDC 56 LLC* suggests that the jurisdictional aspect of § 303(b) would trump the command of § 303(h) that an order for relief be entered if the petition is not ‘timely controverted,’” *Id.* at 162, indicating that it might also reject a debtor’s motion for dismissal for a § 303(b) deficiency if it is untimely.

²²¹ *Id.* at 161-62 (internal quotation marks omitted) (quoting *In re BDC 56 LLC*, 330 F.3d 111, 118 (2d Cir. 2003)).

²²² 11 U.S.C. § 303(h) (“If the petition is not timely controverted, the court shall order relief against the debtor in an involuntary case under the chapter under which the petition was filed.”).

²²³ See *supra* note 212 and accompanying text.

²²⁴ What *MarketXT* demonstrates is a lower court bound by an unworkable precedent, that must fashion a coherent argument for its holding from contradictory authority. In its battle to make sense of *BDC*’s holding, the bankruptcy court even misconstrued a passage from a widely recognized authority on bankruptcy when it stated that “subject matter jurisdiction arises ‘in other contexts under section 303, most notably subsections (b) and (h).’” *MarketXT*, 347 B.R. at 161 (quoting 2 COLLIER ON BANKRUPTCY ¶ 303.02[6] (Alan N. Resnick & Henry J. Sommer eds., 15th ed. 2002)). In fact, that source was referring to the issue addressed in *BDC*, and went on to conclude that “[t]he better argument is that the . . . requirements of section 303(b) can be waived.” COLLIER, *supra* note 5, ¶ 303.08[3]. Moreover, the *MarketXT* court also downplayed the significance of the Supreme Court’s holding in *Arbaugh*, stating merely that “in recent decisions the Supreme Court has narrowed the effect of the term [‘jurisdiction’].” *MarketXT*, 347 B.R. at 162. Rather than acknowledge that the Second Circuit improvidently labeled § 303(b) as jurisdictional, the bankruptcy court simply declined to extend *BDC*. *Id.*

involuntary bankruptcy case.²²⁵ However, it remains unclear how the Second Circuit will respond to a case where the debtor waits to raise the Section 303(b) defense until after an order for involuntary bankruptcy relief has been entered. The mere fact that the outcome of this hypothetical is uncertain points to the flaws in the circuit's current approach.²²⁶

2. Making Sense of *BDC*

What makes the Second Circuit's conclusion in *BCD* all the more puzzling is that the determination that Section 303(b) is subject-matter jurisdictional was unnecessary given the procedural posture of the case.²²⁷ Specifically, the debtor filed a timely response and therefore could raise affirmative defenses based on the substantive requirements of Section 303(b),²²⁸ thus making the issue of Section 303(b)'s construction superfluous.²²⁹ The circuit court purported to address the issue to determine the proper standard of appellate review²³⁰ when it adopted a plenary standard of review²³¹ over the more commonly-accepted

²²⁵ See, e.g., *MarketXT*, 347 B.R. at 160 (interpreting § 303(d) to mean that “only [t]he debtor, or a general partner in a partnership debtor that did not join in the petition, may file an answer”) (emphasis added). “A ‘creditor is not authorized to contest an involuntary petition because a creditor may have an incentive to protect a preference or to gain some unfair advantage at the expense of other creditors.’” *Id.* (quoting *In re Westerleigh Dev. Corp.*, 141 B.R. 38, 40 (Bankr. S.D.N.Y. 1992)) (internal alterations omitted); *In re Taylor & Assocs., L.P.*, 191 B.R. 374, 378-79, 381 (Bankr. E.D. Tenn. 1996) (this rule “prohibit[s] creditors from contesting an involuntary petition in order to prevent creditors from protecting a preference or retaining some other unfair advantage”); *In re New Era Co.*, 115 B.R. 41, 44-45 (Bankr. S.D.N.Y. 1990) (listing cases in support of this proposition). This interpretation is also reinforced by FED. R. BANKR. P. 1011(a).

²²⁶ The two apparent options in such a case are: (1) that the court will apply the *BDC* holding that § 303(b) is jurisdictional, and allow the debtor to obtain dismissal even after entry of relief by showing that the debt is subject to a bona fide dispute; or (2) that it will distinguish *BDC* on the grounds that the debtor from *BDC* moved for dismissal before entry of relief, and deny dismissal to the untimely debtor. See, e.g., *United Marine L.L.C. v. Just for Windows, Inc.*, No. 01 Civ. 5066(HB), 2002 WL 72933 (Bankr. S.D.N.Y. Jan. 17, 2002) (where the court essentially took the latter approach). Because *MarketXT* has already demonstrated that *BDC*'s version of subject matter jurisdiction allows for exceptions, there is no principled way of predicting what further exceptions may be made on a case-by-case basis.

²²⁷ The debtor answered the involuntary petition within twenty days, as mandated by FED. R. BANKR. P. 1011(b). *Key Mech. Inc. v. BDC 56 LLC*, No. 01 Civ. 10169(RWS), 2002 WL 449856, at *1 (S.D.N.Y. Mar. 22, 2002); *supra* note 72.

²²⁸ *Key Mech.*, 2002 WL 449856, at *1.

²²⁹ See *supra* note 72.

²³⁰ *In re BDC 56 LLC*, 330 F.3d 111, 118-19 (2d Cir. 2003).

²³¹ *Id.* at 116-17, 119.

review for clear error.²³² However, if the court's objective was simply to apply a more rigorous standard of review than the clearly erroneous standard, it could have accomplished it without construing Section 303(b) as subject-matter jurisdictional.²³³ In sum, the facts and procedural posture of *BDC* simply did not require the court to hold that Section 303(b) is jurisdictional.

The Second Circuit's concern with the potential unfairness of involuntary bankruptcy cases is legitimate—treating Section 303(b) as substantive could lead to the undesirable result of allowing creditors to force debtors into bankruptcy on the basis of disputed claims²³⁴ because a Section 303(b) deficiency would be an affirmative defense that is waived if not timely raised.²³⁵ However, while the waiver of affirmative defenses for failure to raise them in a timely manner arguably leads to unfairness in any circumstance, it nonetheless is uniformly accepted by the federal courts.²³⁶ The

²³² See *id.* at 118 n.3 (collecting cases in which other circuit courts “held that the clearly erroneous standard of review applies on appeal of a bankruptcy court’s determination that a bona fide dispute exists.”).

²³³ The courts that have adopted a per se rule of reviewing a bankruptcy court’s findings regarding a bona fide dispute for clear error have justified it on the basis that such a determination “will often depend . . . upon an assessment of witnesses’ credibilities and other factual considerations.” *In re Rimell*, 946 F.2d 1363, 1365 (8th Cir. 1991). When facts are in dispute, this approach is proper. FED. R. BANKR. P. 8013. However, in cases where the facts concerning the claim are not in dispute, the question of whether there is a “bona fide dispute” under § 303(b) can be treated as a question of law and given de novo review. See, e.g., *In re Dilley*, 339 B.R. 1, 5 (B.A.P. 1st Cir. 2006) (“Although some appellate courts suggest that the existence of a bona fide dispute is a fact question and thus the clearly erroneous standard always applies, we decline to adopt a per se rule.”) (footnote omitted). Although the *Dilley* court cited *BDC* for this proposition, it did so without adopting *BDC*’s holding that § 303(b) is jurisdictional. *Id.*

²³⁴ See *supra* note 70.

²³⁵ See, e.g., *In re Trusted Net Media Holdings, LLC*, 550 F.3d 1035, 1037 (11th Cir. 2008) (en banc) (“[W]e conclude that § 303(b)’s requirements are not subject matter jurisdictional in nature, and therefore can be waived.”); *In re Mason*, 709 F.2d 1313, 1318 (9th Cir. 1983) (finding that the debtor “waived his right to present [a § 303(b)] defense by failing to raise it in an answer to the petition”).

²³⁶ 5 CHARLES ALLEN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1278 (3d ed. 2008) [hereinafter FEDERAL PRACTICE AND PROCEDURE] (“It is a frequently stated proposition of virtually universal acceptance by the federal courts that a failure to plead an affirmative defense as required by Federal Rule 8(c) results in the waiver of that defense and its exclusion from the case”); *id.* at § 1278 n.1 (collecting cases). The purpose of this rule is to provide the plaintiff with “fair notice of the defense that is being advanced.” *Rogers v. McDorman*, 521 F.3d 381, 385 (5th Cir. 2008) (internal quotation marks and citation omitted). “The concern is that a defendant should not be permitted to lie behind a log and ambush a plaintiff with an unexpected defense.” *Id.* (internal quotation marks and citation omitted). “A bankruptcy pleading also is subject to the Rule 8 requirements as to the pleading of . . . affirmative defenses” FEDERAL PRACTICE AND PROCEDURE § 1229.

apparent desire of the Second Circuit to protect debtors from unscrupulous creditors that would use involuntary bankruptcy as a way to force disputed payments is sensible, but distorting well-established legal principles is a poor way to achieve it.

In addition, the Bankruptcy Code also provides other avenues that do not depend upon Section 303(b) being classified as subject-matter jurisdictional for a debtor to object to an involuntary petition that threatens to work an injustice. For example, the debtor can request that the bankruptcy court abstain from hearing the case on the basis that the parties would be better served by non-bankruptcy proceedings.²³⁷ Alternatively, a debtor that has failed to timely object to an involuntary petition can move to vacate the order for relief by showing that it has “a meritorious defense and that arguably one of the four conditions for relief applies—mistake, inadvertence, surprise, or excusable neglect.”²³⁸ In its evaluation

²³⁷ 11 U.S.C. § 305(a) (“The court, after notice and a hearing, may dismiss a case under this title, or may suspend all proceedings in a case under this title, at any time if . . . the interests of creditors and the debtor would be better served by such dismissal or suspension . . .”). The power to abstain under § 305(a) applies to both voluntary and involuntary bankruptcy cases. *See, e.g., In re Mountain Dairies, Inc.*, 372 B.R. 623, 634-35 (Bankr. S.D.N.Y. 2007) (“Even if [the creditor] were an eligible petitioner under 11 U.S.C. § 303, this Court would be compelled to abstain pursuant to 11 U.S.C. § 305 because this is essentially a two-party dispute for which the parties have adequate remedies in state court. The bankruptcy court is not a collection agency.” (internal footnote omitted)); *In re Aerovias Nacionales de Colombia S.A.*, 303 B.R. 1, 9 n.11 (Bankr. S.D.N.Y. 2003) (“The legislative history of § 305(a)(1) indicates that Congress had in mind a debtor undertaking a voluntary out-of-court restructuring and an involuntary case then being ‘commenced by a few recalcitrant creditors to provide a basis for future threats to extract full payment. The less expensive out-of-court workout may better serve the interests of the case.’”) (quoting H.R. REP. No. 95-595, 95th Cong., 1st Sess. 325 (1977)); *In re Spade*, 258 B.R. 221, 225-31, 37 (Bankr. D. Colo. 2001) (discussing the showing necessary for a bankruptcy court to abstain from hearing an involuntary case, collecting cases, and concluding that “a court may consider any factors it considers to be relevant to the determination of whether a dismissal of the case or a suspension of all proceedings would better serve the interests of the creditors and the debtor,” including the debtor’s “legitimate interest in avoiding the stigma that attaches to those forced into bankruptcy”); *In re ABQ-MCP Joint Venture*, 153 B.R. 338, 342 (Bankr. D.N.M. 1993) (“A court may properly abstain from hearing an involuntary bankruptcy case which is essentially a two-party dispute, where the creditor has adequate state law remedies, and the debtor has no significant assets for the bankruptcy court to administer.”); *In re G-N Partners*, 48 B.R. 459, 461 (Bankr. D. Minn. 1985) (“While there may be other situations in which dismissal under § 305(a) is appropriate, the one most clearly applicable is that in which an out of court ‘work-out’ has been accomplished or is soon to be accomplished and a few recalcitrant creditors have filed an involuntary bankruptcy petition.”).

²³⁸ *In re Hutter Assocs., Inc.*, 138 B.R. 512, 516 (W.D. Va. 1992) (emphasis and alterations removed); *see also In re High Voltage Eng’g Corp.*, 360 B.R. 369, 381-82 (Bankr. D. Mass. 2007) (discussing Rule 60(b) requirements in the context of a bankruptcy case). “[T]here is a strong policy of determining cases on their merits and we therefore view defaults with disfavor.” *In re Worldwide Web Sys., Inc.*, 328 F.3d 1291, 1295 (11th Cir. 2003) (referring to an involuntary debtor’s Rule 60(b) motion to

of a motion to vacate an order for relief, the court may consider the existence of meritorious affirmative defenses based on a Section 303(b) deficiency.²³⁹ Finally, a debtor can object to an improper claim using the same defenses that would be available in a typical non-bankruptcy action to collect.²⁴⁰ Consequently, the apparent likelihood that the bankruptcy court will disallow a legitimately disputed claim should counteract the incentive of creditors to file involuntary petitions to coerce the payment of disputed debts. Therefore, the Bankruptcy Code already incorporates safeguards to ensure that debtors are not improperly subjected to involuntary bankruptcy proceedings.²⁴¹

CONCLUSION

While involuntary petitions may be a small percentage of total bankruptcy cases,²⁴² the proper application of the law is of the utmost importance to those against whom an involuntary

vacate a default judgment in a bankruptcy case). As the court in *Hutter* recognized, “Bankruptcy Rules 7055 (‘Default’) and 9024 (‘Relief from Judgment or Order’) make Fed.R.Civ.P. 60(b) applicable to [an involuntary] case.” 138 B.R. at 516; *see also In re Paczesny*, 282 B.R. 646, 647 (Bankr. N.D. Ill. 2002) (involuntary relief was entered after debtor failed to timely answer, but court subsequently granted debtor’s motion to vacate the order for relief); *In re Morris*, 115 B.R. 752, 754-55 (Bankr. E.D.N.Y. 1990) (vacating order for relief pursuant to FED. R. CIV. P. 60(a), where the order was prematurely entered due to administrative error).

²³⁹ *See Jet Star Enters. Ltd. v. CS Aviation Servs.*, No. 01 Civ. 6590(DAB), 2004 WL 350733, at *8 (S.D.N.Y. Feb. 25, 2004) (explaining that “the Second Circuit requires consideration of [affirmative] defenses when ruling on motions to set aside default judgments”) (citing *Enron Oil Corp. v. Diakuhara*, 10 F.3d 90, 96 (2d Cir. 1993)).

²⁴⁰ 11 U.S.C. § 502(b) (“[I]f [an] objection to a claim is made, the court, after notice and a hearing . . . shall allow such claim . . . except to the extent that . . . such claim is unenforceable against the debtor and property of the debtor, under any agreement or applicable law for a reason other than because such claim is contingent or unmatured . . .”); *Travelers Cas. & Sur. Co. of Am. v. Pac. Gas & Elec. Co.*, 549 U.S. 443, 450 (2007) (“[It is a] settled principle that ‘[c]reditors’ entitlements in bankruptcy arise in the first instance from the underlying substantive law creating the debtor’s obligation, subject to any qualifying or contrary provisions of the Bankruptcy Code. . . .’ That principle requires bankruptcy courts to consult state law in determining the validity of most claims.”) (quoting *Raleigh v. Ill. Dep’t of Revenue*, 530 U.S. 15, 20 (2000)); *In re Shaffner*, 320 B.R. 870, 876 (Bankr. W.D. Mich. 2005) (“Section 502(b)(1) permits the objecting party to challenge the validity of the claim for any of the myriad reasons that would arise either under the agreement itself (e.g., the amount owed is \$1,000, not \$2,000) or under applicable non-bankruptcy laws (e.g., lack of consideration or the statute of frauds).”).

²⁴¹ Additionally, § 303(i) provides that costs and attorney’s fees may be awarded to a debtor who succeeds in having the involuntary petition dismissed. 11 U.S.C. § 303(i)(1). If the petition was filed in bad faith, this provision also allows the court to award actual and punitive damages. 11 U.S.C. § 303(i)(2).

²⁴² *See supra* note 5.

petition is commenced. The need for uniform application of the bankruptcy laws²⁴³ suggests that the Second Circuit's position in *BDC* deserves close scrutiny. Although the Bankruptcy Code's jurisdictional structure and statutory language provide sufficient arguments against the Second Circuit's interpretation of Section 303(b) as subject-matter jurisdiction,²⁴⁴ the Supreme Court's recent decision in *Arbaugh*²⁴⁵ provides the proverbial nail in the coffin and illustrates the importance of proper statutory construction on a scale much larger than just that of the Bankruptcy Code. Therefore, it is time for the Second Circuit to confront the unworkable precedent that it created in *BDC* and explicitly hold that Section 303(b) of the Bankruptcy Code does not pertain to subject matter jurisdiction.

Rachel Green[†]

²⁴³ See, e.g., *In re Frushour*, 433 F.3d 393, 400 (4th Cir. 2005) ("Uniformity among the circuits is also important in the bankruptcy context."); *Gonzales v. Parks*, 830 F.2d 1033, 1035 (9th Cir. 1987) (noting the importance of "the uniformity of federal bankruptcy law, a uniformity required by the Constitution") (citing U.S. CONST. art. I, § 8, cl. 4); Robert K. Rasmussen, *A Study of the Costs and Benefits of Textualism: The Supreme Court's Bankruptcy Cases*, 71 WASH. U. L.Q. 535, 573 (1993) ("The unavoidable conclusion is that the Court is much more concerned with ensuring uniformity in the implementation of federal bankruptcy law than with the content of such law.")

²⁴⁴ See *supra* Part III.

²⁴⁵ See *supra* Part III.A.3.

[†] J.D. Candidate, Brooklyn Law School, 2010; B.S., Cornell University, 2001. I would like to thank my parents, my brother, and Nate for their constant support and encouragement. Thanks also to everyone on *Brooklyn Law Review* for their hard work and their friendship.

The Sovereign Debt Dilemma

When it becomes necessary for a state to declare itself bankrupt, in the same manner as when it becomes necessary for an individual to do so, a fair, open, and avowed bankruptcy is always the measure which is both least dishonourable to the debtor, and least hurtful to the creditor.¹

—Adam Smith

INTRODUCTION

Over the past two decades, the securitization of sovereign lending and the emergence of the secondary debt market have transformed the contours of the global financial system.² Although public debt remains one of the most effective tools to implement domestic economic policy,³ fundamental changes in the design and structure of sovereign financing have stymied the efficient restructuring of state obligations.⁴ In particular, the late-1980s shift from syndicated bank lending⁵ to securitized bond financing⁶ has resulted in a vast

¹ 2 ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS 468 (Edwin Cannan ed., The Univ. of Chi. Press 1976) (1776).

² Michael Wolfgang Waibel, Sovereign Debt Restructuring 6 (Winter 2003) (unpublished manuscript, available at http://forschungsnewsletter.univie.ac.at/fileadmin/user_upload/int_beziehungen/Internetpubl/waibel.pdf).

³ For example, public debt can fund human capital development and physical infrastructure projects, mitigate the effects of temporary economic downturns, and redistribute “resources from future generations to the current one.” EDUARDO BORENSZTEIN ET AL., AMERICAN DEVELOPMENT BANK, LIVING WITH DEBT: HOW TO LIMIT THE RISKS OF SOVEREIGN FINANCE 3-4 (2006).

⁴ A. Mechele Dickerson, *A Politically Viable Approach to Sovereign Debt Restructuring*, 53 EMORY L.J. 997, 1012 (2004).

⁵ Under the syndicated lending practices of the 1970s, relatively small groups of commercial banks would extend credit to a sovereign “on identical terms . . . pursuant to a single loan agreement.” Lee C. Buchheit & Ralph Reisner, *The Effect of the Sovereign Debt Restructuring Process on Inter-Creditor Relationships*, 1988 U. ILL. L. REV. 493, 500 (1988). From the perspective of the debtor, syndicated lending facilitates the acquisition of funds, which would otherwise be unattainable from an individual financing source. Likewise, for both lenders and borrowers, syndicate loans promote efficiency by aggregating initial negotiations and subsequent loan administration into a single, collaborative endeavor. *Id.*

⁶ In response to the fallout from the Latin American debt crisis of the early-1980s, United States Treasury Secretary Nicholas Brady sought “to ‘securitize’ sovereign loans by converting loan obligations into bonds.” Philip J. Power, *Sovereign Debt: The Rise of the Secondary Market and Its Implications for Future Restructurings*,

diversification of sovereign creditors and a substantial increase in the collective action problems among them.⁷ This change, combined with the lack of reliable enforcement regimes and restructuring institutions, has led to a global sovereign debt dilemma.⁸

In conjunction with the return of securitized bond financing,⁹ the global debt crisis of the 1980s also spurred the emergence of a secondary market in sovereign debt.¹⁰ To policy makers and sovereign debtors alike, this new market presented a host of challenges that were not present under previous financing schemes.¹¹ Accordingly, when subsequent financial crises necessitated the renegotiation of sovereign obligations, the syndicate loan restructuring model¹² no longer functioned in

64 FORDHAM L. REV. 2701, 2720 (1996). Under the so-called “Brady Plan,” syndicate bank loans were pooled together and exchanged for debt-securities guaranteed by United States Treasury Bills. Jessica W. Miller, Comment, *Solving the Latin American Sovereign Debt Crisis*, 22 U. PA. J. INT’L ECON. L. 677, 685-86 (2001). After repackaging, the securities were sold in the public markets and the sale proceeds used to satisfy outstanding sovereign loan obligations. Power, *supra*, at 2720. As a result of securitization, individual bondholders replaced commercial bank syndicates. *Id.* at 2719. Although sovereign lending has evolved to include other types of bond instruments, the securitization of debt remains the principle means of sovereign financing today. Christopher C. Wheeler & Amir Attaran, *Declawing the Vulture Funds: Rehabilitation of a Comity Defense in Sovereign Debt Litigation*, 39 STAN. J. INT’L L. 253, 261 (2003).

⁷ William W. Bratton & G. Mitu Gulati, *Sovereign Debt Reform and the Best Interest of Creditors*, 57 VAND. L. REV. 1, 20-22 (2004).

⁸ Dickerson, *supra* note 4, at 1012-13.

⁹ Following the independence movement of the early nineteenth century, newly sovereign nations in Latin America began to procure external financing through government bond issues in European capital markets. BORENSZTEIN ET AL., *supra* note 3, at 63. After enjoying almost a century of heavy capital inflows, World War I brought most sovereign financing to a halt. *Id.* at 63, 75-76. With the onset of World War II and the resulting European capital controls, the United States replaced Britain as the center of global capital. *Id.* at 76. By the time the credit markets thawed in the 1970s, New York had emerged as the dominant capital market and syndicated lending had replaced bond financing as the primary mode of sovereign borrowing. *Id.* at 74, 79.

¹⁰ See *infra* Part II.

¹¹ See *infra* Part II.B.

¹² Between 1982 and 1983, over fifteen countries declared that they would fall into arrears or suspend payments on approximately \$90 billion in foreign syndicate loans. Power, *supra* note 6, at 2708 n.27. Due to the relatively small number of lenders and the interconnected nature of the affected notes, principles of “shared sacrifice” dominated the syndicate loan restructuring atmosphere. *Id.* at 2710. Accordingly, bank advisory committees were formed to promote “equity among banks, . . . and . . . [make] it harder for individual banks to hold out for special treatment.” Charles Lipson, *Bankers’ Dilemmas: Private Cooperation in Rescheduling in Sovereign Debts*, WORLD POLITICS, Vol. 38, No. 1 200, 212 (Oct. 1985). In accordance with these equity principles, lenders were asked to extend gap financing to sovereign debtors equal to their pro rata share of credit exposure. Jill E. Fisch & Caroline M. Gentile, *Vultures or Vanguard?: The Role of Litigation in Sovereign Debt Restructuring*, 53 EMORY L. J. 1043, 1058 (2004). As a temporary stopgap measure, bridge financing permitted a

the fluid and dynamic secondary market.¹³ Initially, the international debate centered on whether public institutions or private actors should lead the call to restructuring reform.¹⁴ However, with the death of the Sovereign Debt Restructuring Mechanism,¹⁵ independent, contractual remedies continue to govern sovereign debt restructuring and will do so for at least the foreseeable future.¹⁶

In 2002, soon after Argentina declared a \$132 billion public debt default,¹⁷ significant contractual reforms began to permeate throughout the sovereign financing market.¹⁸ At first, Mexico took center stage when it announced the first large-scale sovereign bond issuance in New York to incorporate collective action clauses (the “CACs”).¹⁹ By providing for the supermajority modification of certain repayment matters, CACs sought to curb the inefficiencies posed by the unanimous

debtor to make interest payments while creditors worked to reschedule the principal due on the loan. Power, *supra* note 6, at 2709-10. Likewise, due to the discretionary enforcement nature of the International Lending Supervision Act of 1983, which required lenders to accumulate additional reserves, federal banking regulators were not above making “friendly” calls” to incentivize uncooperative lenders to participate in the restructuring process. *Id.* at 2713. In addition, cross-default clauses in the syndicate loan agreements discouraged maverick litigation by requiring all legal proceeds to be shared pro rata with fellow lenders. Anne Krueger, First Deputy Managing Director, International Monetary Fund, Speech at the Economics Society Dinner, The Evolution of Emerging Market Capital Flows: Why We Need to Look Again at Sovereign Debt Restructuring (Jan. 21, 2002), available at <http://imf.org/external/np/speeches/2002/012102.htm>.

¹³ See *infra* Part II.B.

¹⁴ See generally LUCIO SIMPSON, UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT, G-24 DISCUSSION PAPER SERIES, THE ROLE OF THE IMF IN DEBT RESTRUCTURINGS: LENDING INTO ARREARS, MORAL HAZARD, AND SUSTAINABILITY CONCERNS, 1-9 (2006).

¹⁵ Under the auspices of the International Monetary Fund, the SDRM was based on Chapter 11 of the United States Bankruptcy Code and sought to ensure the “orderly . . . and rapid restructuring of . . . debt, while protecting asset values and creditors’ rights.” Sean Hagan, *Designing a Legal Framework to Restructure Sovereign Debt*, 36 GEO. J. INT’L L. 299, 337-40 (2005); ANNE O. KRUEGER, INT’L MONETARY FUND, A NEW APPROACH TO SOVEREIGN DEBT RESTRUCTURING 4, 21 (2002).

¹⁶ See *infra* Part I.A.

¹⁷ Clifford Krauss, *Argentine Leader Declares Default on Billions in Debt*, N.Y. TIMES, Dec. 24, 2001, at A1.

¹⁸ See *infra* Part III.

¹⁹ United Mexican States, Prospectus, at 7 (Dec. 4, 2002). Prior to Mexico’s 2003 issuance, unanimous action clauses (the “UACs”) governed the vast majority of sovereign bonds issued pursuant to New York law. Under a UAC, the modification of reserved matters cannot be effectuated without unanimous bondholder consent. As a result, small factions of holdout creditors can derail the restructuring process. Sergio J. Galvis & Angel L. Saad, *Collective Action Clauses: Recent Progress and Challenges Ahead*, 35 GEO. J. INT’L L. 713, 714-15 (2004). Because CACs impair the ability of a holdout creditor to derail debt rescheduling, they are generally regarded as a more efficient means to facilitate sovereign debt restructuring.

action clauses, which had previously dominated the market.²⁰ Several months later, Uruguay followed with a similar debt issuance that incorporated CACs along with aggregation principles and a pseudo-trustee structure.²¹ In the event of a debt restructuring, aggregation enables a supermajority of bondholders to cram down the modification of certain reserved matters across multiple series of bonds.²² Likewise, the weak trustee structure underlying the notes provided bondholders with a centralized figure that could both initiate collective legal actions as well as distribute any resulting legal award.²³

Although these contractual reforms pushed sovereign financing forward, none provided a comprehensive solution to the creditor holdouts that pose the sovereign debt dilemma.²⁴ Under collective action theory, rational, self-interested individuals will choose personal gain over the pursuit of collective objectives.²⁵ As a result, some form of coercion is required to obtain the optimal aggregate outcome.²⁶ In the case of a sovereign debt default, the potential recoveries from holdout-litigation motivate creditors to abstain from the voluntary restructuring process.²⁷ Without an indenture trustee to strip bondholders of their right to pursue individual legal remedies,²⁸ the Mexican and the Uruguayan reforms have failed to fully embrace effective contractual coercion.²⁹ Given this inability to efficiently coerce creditor cooperation, the holdout problem will persist, and the sovereign debt dilemma will remain.

Part I of this Note begins with a brief examination of the primary differences between lending in the public and

²⁰ Fisch & Gentile, *supra* note 12, at 1093.

²¹ República Oriental del Uruguay, Trust Indenture Filed April 10, 2004, 13-15, 35-36 [hereinafter República Oriental, Indenture]; *see also* Galvis & Saad, *supra* note 19, at 717.

²² Galvis & Saad, *supra* note 19, at 722.

²³ República Oriental, Indenture, *supra* note 21, at 14; Galvis & Saad, *supra* note 19, at 723-24.

²⁴ *See infra* Part IV.

²⁵ Indeed, "even if all of the individuals in a large group are rational and self-interested, and would gain if, as a group, they acted to achieve their common interest or objective, they will still not voluntarily act to achieve that common or group interest." MANCUR OLSON, JR, *THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS* 2 (1965).

²⁶ *Id.*

²⁷ Wheeler & Attaran, *supra* note 6, at 259-60.

²⁸ Fisch & Gentile, *supra* note 12, at 1105; *see also* Galvis & Saad, *supra* note 19, at 723.

²⁹ *See infra* Part IV.

private spheres, as well as a cursory review of the current state of sovereign debt. Part II explores the collective action challenges that resulted from the rise of the secondary debt market and the inability of public institutions to effectively resolve this problem. Part III contends that the Mexican and the Uruguayan models fail to adequately combat the complexities of the holdout problem. To better address this issue, as well as provide for greater efficiency in debt restructuring, Part IV advocates for the inclusion of an indenture trustee clause in subsequent sovereign financing contracts.

I. BACKGROUND OF THE SOVEREIGN DEBT CRISIS

A. *Differences Between Sovereign and Private Lending*

To fully appreciate the role of holdout creditors and the resulting sovereign debt dilemma, it is first necessary to understand the fundamental differences between public and private borrowing.³⁰ In the context of private lending, creditors have recourse to legal regimes that will enforce the payment obligations of debtors.³¹ Similarly, bankruptcy institutions protect distressed borrowers from financial dismemberment³² and ensure “equal treatment” among similarly situated creditors.³³ As a result, private financing schemes provide both

³⁰ Bratton & Gulati, *supra* note 7, at 10-12.

³¹ *Id.* at 11. In the sphere of private lending, the maxim *pacta sunt servanda* continues to apply. Henry T. C. Hu & Jay Lawrence Westbrook, *Abolition of the Corporate Duty to Creditors*, 107 COLUM. L. REV. 1321, 1389-90 (2007). Accordingly, if a debtor fails to comport with his or her promise to pay, courts will enforce this obligation in accordance with debtor-creditor and contract law. *Id.*; *see, e.g.*, Gerdes v. Kennamer, 155 S.W.3d. 541, 546 (Tex. App. 2004) (holding that a court may order a judgment debtor to turnover property “subject to the debtor’s control” even though the property is located outside of the United States) (quoting TEX. CIV. PRAC. & REM. CODE ANN. § 31.002(b)(1) (Vernon 1997)).

³² For example, the automatic stay provision of the United States Bankruptcy Code “provides the debtor with relief from the pressure and harassment of creditors seeking to collect on their claims” as well as “breathing space . . . to focus on its rehabilitation or reorganization.” 3 COLLIER ON BANKRUPTCY § 362.03 (2005); 11 U.S.C.A. § 362 (2006).

³³ Rory Macmillan, *Towards a Sovereign Debt Work-Out System*, 16 NW. J. INT’L L. & BUS. 57, 74 (1995). The equitable distribution of assets among similarly situated creditors is a core principle of the United States Bankruptcy Code. 5 COLLIER ON BANKRUPTCY § 541.01 (2007). Accordingly, U.S. bankruptcy courts “should aim to treat similarly situated creditors similarly.” *Till v. SCS Credit Corp.*, 541 U.S. 465, 477 (2004).

debtors and creditors with access to legal authorities that will enforce the reasonable expectations of the lending agreement.³⁴

In the world of sovereign financing, however, things are different,³⁵ since creditors lack recourse to reliable enforcement institutions when the borrower fails to pay.³⁶ In the United States, prior to the passage of the Foreign Sovereign Immunities Act (the “FSIA”), sovereign debtors enjoyed an unqualified immunity in both state and federal courts.³⁷ With a judiciary that recognized absolute sovereign immunity, lenders relied solely on the President to compel payment from recalcitrant sovereign debtors.³⁸ Today, though the United States Supreme Court has held that debt obligations are a “commercial activity” no longer subject to sovereign immunity,³⁹

³⁴ Mitu Gulati & George Triantis, *Contracts Without Law: Sovereign Versus Corporate Debt*, 75 U. CIN. L. REV. 977, 986 (2007). “The assurance of protection from the consequences of debtor default is a fundamental necessity in the commercial world, whose order depends upon the predictability of the debtor-creditor relationship and the realization of reasonable expectations.” E. Hunter Taylor, Jr., *Recent Developments in Commercial Law*, 11 RUTGERS-CAM. L.J. 527, 657 (1980).

³⁵ Bratton & Gulati, *supra* note 7, at 11.

³⁶ *Id.* In addition to the lack of enforcement mechanisms, secured lending is usually not an option when contracting with a sovereign. Patrick Bolton & David A. Skeel, Jr., *Inside the Black Box: How Should a Sovereign Bankruptcy Framework Be Structured?*, 53 EMORY L. J. 763, 793 (2004).

³⁷ Bradford R. Clark, *Domesticating Sole Executive Agreements*, 93 VA. L. REV. 1573, 1618 (2007). When the United States Supreme Court first examined sovereign immunity in *The Schooner Exchange v. McFaddon*, the Court found “that the sovereign power of the nation is alone competent to avenge wrongs committed by a sovereign, that the questions to which such wrongs give birth are rather questions of policy than of law, [and] that they are for diplomatic, rather than legal discussion.” *The Schooner Exchange v. McFaddon*, 11 U.S. (7 Cranch) 116, 146 (1812).

³⁸ Clark, *supra* note 37, at 1618.

[E]arly Presidents embraced the role of chief negotiator by espousing and settling claims of U.S. citizens against foreign nations barred by foreign sovereign immunity. Presidents would decide, in their discretion, whether and how to espouse such claims. Even if the President agreed to espouse a claim, he retained wide-ranging discretion in disposing of it. He could “compromise it, seek to enforce it, or waive it entirely.” . . . In the end, whatever compensation the President secured for the claimant was almost certainly greater than any amount the claimant could recover on his or her own, since foreign sovereign immunity foreclosed access to U.S. courts.

Id. at 1627-29. Outside the United States, one of the most infamous attempts at sovereign debt collection occurred in 1902, when the British and German navies fired on the Venezuelan coast and threatened to invade unless the debts of their subjects were paid in full. Likewise, it was not until 1907 that Luis Drago, the Argentine politician, first espoused the doctrine that sovereign debt cannot justify an armed conflict or occupation of a debtor state. See Lee C. Buchheit, *The Role of the Official Sector in Sovereign Debt Workouts*, 6 CHI. J INT’L L. 333, 336-37 (2005).

³⁹ *Republic of Argentina v. Weltover, Inc.*, 504 U.S. 607, 620 (1992).

lenders continue to face daunting debt collection challenges.⁴⁰ For example, because FSIA does not permit a creditor to seize sovereign assets located outside of the U.S. border,⁴¹ sovereign debtors have transferred assets out of the United States on the eve of declaring default.⁴² Once these monies have exited the country, the lender often remains without an effective means to collect on the sovereign debt.⁴³

When compared to commercial borrowing, sovereign lending also carries a heightened expectation of breach.⁴⁴ While economic or political factors may give rise to a sovereign's default, the absence of realistic enforcement procedures provides an incentive for nation-states to ignore debt obligations even when they are able to pay.⁴⁵ Rather than face the political, economic, or social consequences of conservative fiscal policies, sovereigns may choose instead to default opportunistically.⁴⁶ As a result, a common assumption underlying the sovereign financing process is that the borrower will inevitably fail to pay.⁴⁷ Consequently, a primary challenge for sovereign lenders is to devise a contractual mechanism that will realize the reasonable expectations of the lender-borrower relationship when the debtor inevitably defaults.⁴⁸

In addition to the lack of effective enforcement mechanisms, there is similarly no global institution to address

⁴⁰ Bratton & Gulati, *supra* note 7, at 11. For instance, two problems that continue to plague sovereign debt satisfaction are: (1) the difficulty in identifying sovereign property that is subject to execution, and (2) the inability to liquidate a sovereign debtor. *Id.*

⁴¹ Jonathan C. Lippert, *Vulture Funds: The Reason Why Congolese Debt May Force a Revision of the Foreign Sovereign Immunities Act*, 21 N.Y. INT'L L. REV. 1, 14-15 (2008). Indeed, the scope of FSIA extends only to "property in the United States." 28 U.S.C. § 1601(a) (2008).

⁴² In fear of creditor enforcement actions, Argentina removed assets from the United States and deposited them in the Bank for International Settlements before declaring a default in 2001. Bratton & Gulati, *supra* note 7, at 35.

⁴³ In the majority of cases, the sovereign's courts cannot seize the sovereign's assets. Hal S. Scott, *A Bankruptcy Procedure for Sovereign Debtors?*, 37 INT'L LAW 103, 116-17 (2003).

⁴⁴ Robert B. Ahdieh, *Between Mandate and Market: Contract Transition in the Shadow of the International Order*, 53 EMORY L.J. 692, 694 (2004). In addition, the transaction costs of dealing in sovereign debt are higher than the costs of similar corporate transactions. Gulati & Triantis, *supra* note 34, at 986.

⁴⁵ Fisch & Gentile, *supra* note 12, at 1048-49.

⁴⁶ *Id.* Professors Fisch and Gentile argue that holdout litigation serves an important role in frustrating the desirability of an opportunistic default. *Id.* at 1047. Although this may well be the case, it remains to be seen whether such benefits are outweighed by the restructuring disruptions that such creditors pose.

⁴⁷ *Id.* at 1044.

⁴⁸ *Id.* at 1090.

the problem of sovereign debt restructuring.⁴⁹ Whereas legal tribunals can allocate the financial rights of debtors and creditors in bankruptcy, sovereigns are not subject to domestic insolvency proceedings.⁵⁰ Although both academics and multinational institutions have put forth proposals for the creation of a global sovereign insolvency regime,⁵¹ these efforts have failed to garner sufficient support for their implementation.⁵² Most recently, Anne Krueger of the International Monetary Fund (the “IMF”) called for the formation of a Sovereign Debt Restructuring Mechanism (the “SDRM”).⁵³ However, due to pushback from both debtor and creditor states, the SDRM was placed indefinitely on hold.⁵⁴ With the rise of contractual approaches to sovereign debt restructuring, the current prospects for a global sovereign insolvency regime appear to be nil.⁵⁵ As a result, lenders and borrowers are left to develop their own contractual devices to effectuate the efficient restructuring of sovereign obligations.

B. *History of Modern Sovereign Financing*

The roots of the holdout problem in sovereign debt restructuring can be traced to the years spanning the early 1970s to the early 1980s, when lending to sovereign debtors experienced exponential growth.⁵⁶ During this period, syndicate loans⁵⁷ from commercial banks in the United States and

⁴⁹ Galvis & Saad, *supra* note 19, at 714. Although creditors and debtors can currently enter into informal agreements under the supervision of the IMF through Paris Club (sovereign creditors and sovereign debtors) and London Club (sovereign debtors and private creditors) negotiations, this system is noncompulsory and has been criticized for its inefficiency. Dickerson, *supra* note 4, at 1008-12.

⁵⁰ See generally, Caroline Atkinson, *Forget Sovereign Bankruptcy Plans* (2002), available at <http://www.cfr.org/publication.html?id=4584>.

⁵¹ KRUEGER, *supra* note 15, at 4.

⁵² Dickerson, *supra* note 4, at 998.

⁵³ KRUEGER, *supra* note 15, at 21.

⁵⁴ Adam Brenneman, Comment, *Gone Broke: Sovereign Debt, Personal Bankruptcy, and a Comprehensive Contractual Solution*, 154 U. PA. L. REV. 649, 679 (2006).

⁵⁵ *Id.*

⁵⁶ For example, in the ten year period between 1973 and 1983, foreign debt in Latin America increased by more than 700%. Miller, *supra* note 6, at 680 (quoting Roy MacMillan, *The Next Sovereign Debt Crisis*, 31 STAN. J. INT'L L. 395, 311 n.31 (1995) (citing PEDRO-PABLO KUCZYNSKI, LATIN AMERICAN DEBT 14 (1988))).

⁵⁷ “A syndicated loan is one that is provided by a group of lenders and is structured, arranged, and administered by one or several commercial or investment banks known as arrangers.” STANDARD & POOR’S, A GUIDE TO THE LOAN MARKET 7 (2009), available at <https://www.lcdcomps.com/d/pdf/LoanMarketguide.pdf>.

Western Europe functioned as the dominant source of financing for sovereigns in the developing world.⁵⁸ After the 1979 energy crisis,⁵⁹ however, the Federal Reserve Board (the “Fed”) increased interest rates to combat growing domestic inflation, and as a result capital flew from developing countries back into the United States.⁶⁰ In response to the Fed’s higher discount rate, lenders in the United States hiked prime rates on outstanding sovereign loans.⁶¹ To the sovereigns, this had the detrimental effect of increasing both the nominal value of interest payments as well as the real rate of interest on their debt.⁶² Consequently, on August 22, 1982, Mexico became the first nation of the 1980s financial crisis to announce that it would be unable to service its outstanding loan obligations.⁶³ Less than one year later, fifteen additional countries declared

⁵⁸ Fisch & Gentile, *supra* note 12, at 1054; *see also* Power, *supra* note 6, at 2707. During this period, U.S. financial institutions were awash in deposits from oil-exporting nations, Fisch & Gentile, *supra* note 12, at 1054, while an economic downturn and rising inflation at home reduced the domestic demand for credit. Power, *supra* note 6, at 2707. Given the surplus of petrodollar deposits and the rising price of raw material exports from developing nations, commercial banks viewed sovereigns as a justifiable credit risk. *Id.* Indeed, lenders believed “sovereign borrowers were immune from bankruptcy risk and would not suspend debt servicing.” Alberto Gonzalo Santos, *Beyond Baker and Brady: Deeper Debt Reduction for Latin American Sovereign Debtors*, 66 N.Y.U. L. REV. 66, 74 (1991). As a result, financial institutions would routinely ignore sound lending practices such as profitability analysis and investment requirements. *Id.* at 73-74. To sovereign debtors, rising inflation in the United States counteracted high interest rates, *id.* at 72, and also rendered the real rate of interest negative for a few years, increasing the desirability of borrowing in U.S. dollars. *Id.* at 72 n.41. Encouraged by the liberal lending practices of U.S. banks combined with highly favorable financing costs, many countries pursued unsustainable development through excessive foreign borrowing at the expense of conservative fiscal policies. *Id.* at 74-75.

⁵⁹ The overthrow of the Shah of Iran resulted in an energy crisis that doubled the price of oil within a year. Jon H. Sylvester, *Impracticability, Mutual Mistake, and Related Contractual Bases for Equitably Adjusting the External Debt of Sub-Saharan Africa*, 13 NW. J. INT’L L. & BUS. 258, 264 (1992). Although some debtor nations are petroleum producers (e.g., Venezuela), the vast majority are not. *Id.* at 264 n.30. Accordingly, to compensate for the increased cost of petroleum products, sovereign debtors borrowed more heavily from commercial banks. Power, *supra* note 6, at 2707-08. At the same time, however, global recession precipitated a reduction in gross returns on the commodity exports that nations used to service their debt. *Id.* at 2708.

⁶⁰ Santos, *supra* note 58, at 74-75.

⁶¹ Edward Cowan, *Bank Lending Rate Set at Record 14% by Federal Reserve*, N.Y. TIMES, May 5, 1981, at A1.

⁶² Power, *supra* note 6, at 2708.

⁶³ Lee C. Buchheit, *A Quarter of a Century of Sovereign Debt Management: An Overview*, 35 GEO. J. INT’L L. 637, 637 (2004). Although earlier in 1982 Argentina suspended payment on \$37 billion in foreign debt after its defeat in the Falkland Islands War, “it was the Mexican default that shook the financial world.” RICHARD JOLLY ET AL., UN CONTRIBUTIONS TO DEVELOPMENT THINKING AND PRACTICE 142 (2004).

that they too would fall into arrears or suspend payments on approximately \$90 billion in foreign debt.⁶⁴

At the time of the crisis, many commercial banks had extended loans to sovereign debtors in amounts that greatly exceeded their capacity to lend.⁶⁵ To avoid having to declare significant balance sheet losses, commercial banks jointly extended bridge loans to sovereign debtors, which permitted them to make interest payments while creditors worked to reschedule the principal due on the loans.⁶⁶ Although creditors with larger exposure to the debt crisis were more willing to provide funds to engage in gap financing measures,⁶⁷ peer and regulatory pressures ensured cooperation even among the smallest and most recalcitrant lenders.⁶⁸ In addition to austerity programs,⁶⁹ the IMF also instituted policies conditioning new loans on the ability of a nation to obtain

⁶⁴ Power, *supra* note 6, at 2709 n.28; *see also* Steven M. Cohen, Note, *Give Me Equity or Give Me Debt: Avoiding a Latin American Debt Revolution*, 10 U. PA. J. INT'L BUS. L. 89, 97 (1988).

⁶⁵ Fisch & Gentile, *supra* note 12, at 1057. For example, in the United States, nine of the nation's largest financial institutions had loaned more than 250% of their aggregate capital to sovereign nations. *Id.* Under United States banking regulations, lenders had to declare a loan as non-performing if interest on the note was over 90-days past due. *Id.* If the sovereign debtors defaulted on their loans, lenders would have almost certainly faced bankruptcy. *Id.*

⁶⁶ Power, *supra* note 6, at 2709-10. Under this approach, if banks were continuing to receive interest payments in a timely fashion, they could continue to carry the sovereign notes as assets on their balance sheets and avoid bankruptcy. *Id.* at 2710; Fisch & Gentile, *supra* note 12, at 1057.

⁶⁷ Creditors with heavy exposure to the crisis were more willing to provide gap financing for two principal reasons. First, like other creditors, bridge loans would ensure that they could maintain sovereign loans as an asset on their balance sheets. Since these creditors were more heavily exposed to the crises in the region, their prospects for bankruptcy were more acute than those of minor participants. Similarly, these large lenders wanted to maintain good working relationships with the sovereigns. In many cases, the banks looked forward to developing new relationships with local businesses and opening up retail banks in the sovereign nations. *See* Fisch & Gentile, *supra* note 12, at 1058-60.

⁶⁸ Free-riding creditors were a potential problem if larger banks provided the entire financing necessary to avoid default. *Id.* at 1060. Under this scenario, sovereign debtors would have sufficient funds to make interest payments on all their outstanding notes. Consequently, less-exposed creditors would receive the benefit of timely interest payments without having to incur the costs and additional exposure required by providing gap financing. To secure full compliance, members of bank advisory committees were assigned to oversee smaller banks within their geographical region. *Id.* at 1060-61. Because smaller banks sought to grow and develop their working relationships with other financial institutions, larger lenders would threaten international and domestic market isolation if the smaller banks failed to participate in the restructuring of sovereign debt. *Id.* at 1061.

⁶⁹ Such "programs usually involve[d] cutting public spending, devaluing the national currency to stimulate exports and reducing imports." Burton Bollag, *U.N. Critical of I.M.F. Austerity Plan*, N.Y. TIMES, Sept. 6, 1989, at D7.

additional financing from all of its current lenders.⁷⁰ Because of these collective pressures, between 1982 and 1984 commercial banks successfully restructured over forty loan agreements with more than thirty different countries.⁷¹ While the comparatively homogenous views of syndicate bank lenders reduced creditor coordination problems and facilitated the efficient rescheduling of sovereign debt, the subsequent rise of bond financing in the mid-1980s presented new collective action challenges that threatened to hinder the successful restructuring of sovereign obligations.⁷²

II. THE EMERGENCE OF A SECONDARY MARKET IN SOVEREIGN DEBT

A. *Beginnings of a Secondary Market: Inter-Bank Swaps and Brady Bonds*

Although the extension of bridge loans by bank advisory committees and multilateral institutions⁷³ helped to temporarily stave off losses from debtor nations,⁷⁴ several years of cyclical restructuring fatigued creditors, and as a result many institutions opted out of the process.⁷⁵ As the crisis continued to worsen, a secondary market in sovereign debt began to emerge.⁷⁶ Initially, this market consisted entirely of inter-bank swaps,⁷⁷ but as the sovereign debt crisis

⁷⁰ Fisch & Gentile, *supra* note 12, at 1061.

⁷¹ *Id.* at 1063.

⁷² Bratton & Gulati, *supra* note 7, at 20-21.

⁷³ In 1985, at the World Bank Meeting in Seoul, South Korea, United States Secretary of the Treasury James A. Baker III proposed a plan whereby multinational institutions such as the IMF and World Bank would extend an additional \$9 billion in loans to debtor states. Under the terms of the plan, borrowing nations would adopt austerity measures in exchange for the funds. To some observers, the differences between the Baker Plan and the private restructuring organized by bank advisory committees were minimal. Santos, *supra* note 58, at 76-77.

⁷⁴ *Id.*

⁷⁵ Development Committee, Joint Ministerial Committee of the Boards of Governors of the World Bank and the International Monetary Fund, A Strategy for Restoration of Growth in Middle-Income Countries That Face Debt-Servicing Difficulties 12-13 (1986), available at http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2001/02/15/000178830_98101901582392/Rendered/PDF/multi_page.pdf.

⁷⁶ Sylvester, *supra* note 59, at 272.

⁷⁷ Power, *supra* note 6, at 2715.

deteriorated, banks began to trade their foreign loan assets for cash.⁷⁸

After several years of accumulating cash from sovereign loan exchanges, many banks had attained a level of loan-loss reserves that could sustain substantial write-off losses from sovereign notes.⁷⁹ Soon thereafter, it became clear to lenders that the principal on sovereign loans would not be repaid at “any time in the foreseeable future.”⁸⁰ To reduce the debt burden on commercial banks, United States Treasury Secretary Nicholas Brady announced a plan to “securitize” sovereign debts by converting loan obligations into bonds.⁸¹ Under the Brady Plan, syndicated bank loans were pooled together and exchanged for Brady Bonds guaranteed by United States Treasury Bills.⁸² After repackaging, the bonds were sold in the public markets and the proceeds used to satisfy the sovereign’s outstanding debt.⁸³ Importantly, this securitization

⁷⁸ Even though it was highly unlikely that the sovereign debts would be ever be fully repaid, the secondary market became quite popular with some investors. *Id.* at 2716. At first, this market was principally composed of large corporate investors seeking debt-for-equity swaps.

In a debt-for-equity swap, an investor approaches a large debtor nation and expresses an interest in investing in an industry or specific business. The investor proposes to buy outstanding debt from a specific creditor or on the open market for a fraction of the face amount of the outstanding loan. The investor then sells the outstanding loan to the debtor nation for the face amount or for a discounted amount of local currency The investor then uses the sale proceeds to buy an equity stake in the local business, and makes further capital investment.

Sylvester, *supra* note 59, at 272. But as lenders increasingly tried to exit from the unraveling sovereign debt market, they rapidly reduced the price of their sovereign loan assets. Rory Macmillan, *supra* note 56, at 328. Given the availability of fire sale prices, investors with no interest in equity swaps began to purchase the heavily discounted notes. Power, *supra* note 6, at 2718. Even if the debtor only paid back a fraction of the loan’s face value, an investor could realize a potentially large profit. Macmillan, *supra* note 33, at 328. Similarly, because interest continued to accrue on the face value of the notes, interest payments alone could yield “an above-market rate of return.” Power, *supra* note 6, at 2719.

⁷⁹ Power, *supra* note 6, at 2719.

⁸⁰ Miller, *supra* note 6 at 685. By 1989, “the pretense of keeping . . . [the] loans on the books at face value could not longer be maintained.” Macmillan, *supra* note 56, at 313.

⁸¹ Power, *supra* note 6, at 2720. Under the so-called “Brady Plan,” commercial banks agreed to partially forgive sovereign debt obligations “in exchange for both a commitment on the part of the debtors to adopt specified reforms designed to achieve sustainable growth . . . and greater assurances of the collectability of the debt.” Fisch & Gentile, *supra* note 12, at 1067.

⁸² Miller, *supra* note 6 at 685.

⁸³ Power, *supra* note 6, at 2720. The securitization of sovereign lending was quite popular with the market. Accordingly, within five years of initiating the Brady Plan, “more than half of the affected debt stock had been traded in the hands of non-

process replaced debts owed to commercial banks with obligations to a group of individual bondholders.⁸⁴ As a result of the Brady Plan, sovereign financing “shifted” from the banking business to the securities markets⁸⁵ and, although the techniques have changed,⁸⁶ the securitization of debt remains the principle means of sovereign lending today.⁸⁷

B. *The Emergence of the Holdout Problem*

Unlike the homogenized bank syndicates of the 1970s and 1980s, post-Brady bondholders are diverse.⁸⁸ Whereas “[b]ank lenders are repeat players, constrained to cooperate with one another,”⁸⁹ groups of bondholders constantly change as the securities are bought and sold in the market.⁹⁰ Similarly, the vast majority of sovereign bondholders lack any relationship with the debtor, because they became creditors through secondary trading.⁹¹ In the absence of a rapport with either the sovereigns or with each other, bondholders do not feel the same pressures to “compromise their . . . claims” or share sacrifice.⁹² Instead of investing with a common purpose, the liquid secondary market aggregates investors⁹³ with vastly divergent short-term and long-term goals.⁹⁴ Given the relative anonymity among them,⁹⁵ there is little collective pressure to cooperate.⁹⁶ Consequently, sovereign bondholders pose a collective action problem whereby holdout creditors can derail a

bank investors.” Wheeler & Attaran, *supra* note 6, at 261 (quoting Lee C. Buchheit, *Sovereign Debtors and Their Bondholders*, in UNITAR TRAINING PROGRAMMES ON FOREIGN ECONOMIC RELATIONS: SOVEREIGN DEBTORS AND THEIR BONDHOLDERS 7).

⁸⁴ Power, *supra* note 6, at 2719. For the syndicated bank lenders, securitization enabled them to escape from the sovereign debt market. Fisch & Gentile, *supra* note 12, at 1067.

⁸⁵ Wheeler & Attaran, *supra* note 6, at 261.

⁸⁶ In the past ten years, sovereign lending has moved from Brady Bonds to other types of bond instruments. See Miller, *supra* note 6, at 687.

⁸⁷ Wheeler & Attaran, *supra* note 6, at 261.

⁸⁸ *Id.*

⁸⁹ Bratton & Gulati, *supra* note 7, at 20.

⁹⁰ Fisch & Gentile, *supra* note 12, at 1071.

⁹¹ Wheeler & Attaran, *supra* note 6, at 261.

⁹² Dickerson, *supra* note 4, at 1013.

⁹³ Fisch & Gentile, *supra* note 12, at 1071.

⁹⁴ *Id.*

⁹⁵ Ahdieh, *supra* note 44, at 704.

⁹⁶ Wheeler & Attaran, *supra* note 6, at 261.

potentially successful restructuring.⁹⁷ It is this tyranny of the minority that poses the sovereign debt dilemma.

1. The Unanimous Action Requirement and the Vulture Fund Holdouts

Currently, the United States dominates the market for sovereign bond issuances, and New York law governs the majority of U.S.-issued sovereign bonds.⁹⁸ Until Mexico's sovereign bond issuance in 2003, the vast majority of these bonds incorporated unanimous action clauses (the "UACs").⁹⁹ Under a UAC, any alteration to a bond's repayment terms cannot be effectuated without the unanimous consent of all bondholders.¹⁰⁰ As a result, small factions of minority creditors can derail a widely approved restructuring by withholding their support.¹⁰¹

Along with the disruptive power of minority bondholders, the creation of a secondary market in sovereign debt also brought about the rise of "[f]unds specializing in distressed assets."¹⁰² Generally, these "vulture funds" purchase deeply discounted sovereign debt on the secondary market¹⁰³ and later attempt to collect on their claim in full.¹⁰⁴ Although

⁹⁷ Dickerson, *supra* note 4, at 1013.

⁹⁸ Wheeler & Attaran, *supra* note 6, at 259. Historically, most sovereign financing activity took place in European capital markets. However, with the onset of World War I in 1914 and the subsequent global depression, sovereign lending shifted west. As the dominant capital market in the United States, New York emerged as the new leader in sovereign finance. By the time the credit markets thawed in the 1970s, New York had already established itself as the center of the sovereign financing establishment, a position it maintains to this day. BORENSZTEIN ET AL., *supra* note 3, at 74-76.

⁹⁹ Dickerson, *supra* note 4, at 1013-14. In part, the use of UACs in sovereign bonds can be traced to the United States' implementation of the Trust Indenture Act of 1939. Pursuant to the Act, corporate bonds issued in the United States were required to incorporate UACs. Although the Act did not apply to sovereign bonds, commentators have noted that the inclusion of UACs in sovereign financing contracts may simply be the result of "drafting momentum." Buchheit & Gulati, *Sovereign Bonds and the Collective Will*, 51 EMORY L.J. 1317, 1329-30 (2002).

¹⁰⁰ Dickerson, *supra* note 4, at 1013-14.

¹⁰¹ Brenneman, *supra* note 54, at 680.

¹⁰² Wheeler & Attaran, *supra* note 6, at 254.

¹⁰³ *Id.* As a business model, the vulture fund structure can reap significant rewards. In one case, Elliott Associates, a New York-based fund, earned over 494% on a single investment in Peruvian debt. *See id.* at 258.

¹⁰⁴ *Id.* at 262. While champerty laws prevent a third party from purchasing a secondary debt with the sole intention of immediately litigating the claim to obtain full recovery, sovereigns have resoundingly failed in their attempt to combat vulture funds through champerty statutes. *See James Thuo Gathii, The Sanctity of Sovereign Loan Contracts and Its Origins in Enforcement Litigation*, 38 GEO. WASH. INT'L L. REV. 251,

there is usually no reasonable expectation that the debt will be fully repaid,¹⁰⁵ vulture funds “refuse to participate” in the restructuring process¹⁰⁶ because they are immune to the “peer or regulatory” pressures that permeate syndicated bank lending.¹⁰⁷ As a result, these funds circumvent traditional sovereign debt collection procedures and utilize litigation to obtain the full face value of their claims.¹⁰⁸

For a bond issued with UACs, the vulture fund litigation strategy poses substantial problems for the restructuring process.¹⁰⁹ Since an amendment to repayment terms cannot take effect without all outstanding bondholders agreeing to the alteration, the sovereign debtor has incentives to make side payments to any recalcitrant creditors.¹¹⁰ In doing so, the sovereign debtor inadvertently encourages future holdouts.¹¹¹ Not only does a holdout receive the benefit of a side payment, it may also continue to pursue legal remedies to recover on the full face value of its claim.¹¹² If such litigation proves successful, it depletes the total funds available to satisfy the claims of other similarly-situated creditors.¹¹³ Thus, instead of promoting an orderly distribution of assets, the ability of a vulture fund to derail the restructuring process encourages the financial butchering of a sovereign’s foreign exchange reserves.¹¹⁴ “[A] single default” can activate cross-default clauses in other debt instruments and quickly flood the sovereign in an unexpected “avalanche of redeemed debt.”¹¹⁵ Even if litigation proves to be unsuccessful, the unanimity requirements of a UAC provision allow a single holdout to

311-12 (2006); *see also* Elliott Assocs. v. Banco de la Nacion, 194 F.3d 363, 381 (2d Cir. 1999) (holding that the New York champerty statute “is not violated when . . . the accused party’s ‘primary goal’ is . . . [the] satisfaction of a valid debt and its intent is only to sue absent full performance.”).

¹⁰⁵ *See generally* Fisch & Gentile, *supra* note 12, at 1044.

¹⁰⁶ Wheeler & Attaran, *supra* note 6, at 254, 263.

¹⁰⁷ *Id.* at 262.

¹⁰⁸ Fisch & Gentile, *supra* note 12, at 1045. The litigation by some vulture funds has become increasingly aggressive. In the case of the Republic of Congo, vulture funds have attempted to collect on claims by attaching assets held by United States corporations doing business with the nation. *See generally*, Lippert, *supra* note 41.

¹⁰⁹ *See* Dickerson, *supra* note 4, at 1013-15; Fisch & Gentile, *supra* note 12, 1045-46; Wheeler & Attaran, *supra* note 6, at 262-63.

¹¹⁰ Wheeler & Attaran, *supra* note 6, at 259-60.

¹¹¹ *Id.*

¹¹² *Id.*

¹¹³ *Id.*

¹¹⁴ *Id.* at 260-61

¹¹⁵ *Id.* at 260.

bring the entire restructuring process to a halt during the pendency of the suit.¹¹⁶ Although these holdouts may provide valuable benefits to the sovereign financing process,¹¹⁷ they can also thwart a potentially successful restructuring¹¹⁸ and impose heavy burdens on the citizenry of the debtor nation.¹¹⁹ Consequently, holdout bondholders can obstruct the efficient restructuring of sovereign obligations and therefore create the sovereign debt dilemma.¹²⁰

2. Inability of Public Institutions to Solve the Sovereign Debt Crisis

In 2002, to combat the efficiency costs of the holdout problem, Anne Krueger of the International Monetary Fund called for the creation of a Sovereign Debt Restructuring Mechanism (the “SDRM”) under the auspices of the IMF.¹²¹ Based on Chapter 11 of the United States Bankruptcy Code,¹²² the SDRM sought to ensure the “orderly . . . and rapid restructuring of . . . debt while protecting asset values and creditors’ rights.”¹²³ However, the plan ran into problems as soon as it was announced. On the one hand, debtor-states criticized the SDRM for its infringement on national sovereignty and its potential to increase the cost of credit.¹²⁴ On the other hand, lenders argued that a uniform means to restructure sovereign debt would reduce the number of potential investors.¹²⁵ Most importantly, however, the United States disapproved of any global regime to effect sovereign debt

¹¹⁶ Dickerson, *supra* note 4, at 1013-14.

¹¹⁷ According to Professors Fisch and Gentile, “[h]oldout creditors . . . serve as a check on opportunistic defaults and onerous restructuring terms.” Moreover, the enforcement of debt obligations by the judiciary “enhances the operation of the sovereign debt market by lowering the cost of financing to sovereign debtors and increasing the value of the obligation to creditors.” Fisch & Gentile, *supra* note 12, at 1112.

¹¹⁸ Wheeler & Attaran, *supra* note 6, at 254.

¹¹⁹ *Id.*

¹²⁰ *Id.* at 262 (quoting, G. Mitu Gulati & Kenneth N. Klee, *Sovereign Piracy*, 56 BUS. LAW 635, 637-38 (2001)).

¹²¹ KRUEGER, *supra* note 15, at 1, 21.

¹²² *Id.* at 1, 4.

¹²³ *Id.* at 1, 4. The SDRM was modeled closely on Chapter 11 bankruptcy proceedings in the United States. *See id.* at 21.

¹²⁴ Brenneman, *supra* note 54, at 677-78.

¹²⁵ Arturo C. Porzecanski, *A Critique of Sovereign Bankruptcy Initiatives: The IMF and G7 Should Curb Financial Assistance to Countries in Trouble*, BUS. ECON., Jan. 2003, at 39, 44.

restructuring.¹²⁶ Accordingly, in April 2003, United States Treasury Secretary John W. Snow stated that it was “neither necessary nor feasible to continue working on the SDRM.”¹²⁷ Given the resistance of the United States and the investment community to any “statutory bankruptcy-like process,”¹²⁸ the SDRM proposal was placed on hold.¹²⁹ Today, any prospect for the establishment of a formal nation-state restructuring regime appears to be dead.¹³⁰

In the debate leading up to the demise of the SDRM, Treasury Secretary John W. Snow noted that “a contractual approach . . . would help promote a more orderly restructuring process . . . [because] [t]he source of . . . [the] problem . . . lies in the relationships and agreements . . . [between] debtors and their creditors.”¹³¹ Given the prevalence of UACs prior to 2003 and the resulting holdout problem, the IMF,¹³² the United States,¹³³ and the Group of 10 (the “G-10”),¹³⁴ advocated for a full transition from unanimous action clauses to collective action clauses in sovereign financing contracts. Through the use of CACs, it was believed that the collective action problem could be mitigated, since a supermajority vote could bind a minority of holdout creditors.¹³⁵

¹²⁶ John W. Snow, U.S. Sec’y of Treas., Statement at the Meeting of the International Monetary and Financial Committee (Apr. 12, 2003), *available at* <http://www.imf.org/external/spring/2003/imfc/state/eng/usa.htm>.

¹²⁷ *Id.*

¹²⁸ Galvis & Saad, *supra* note 19, at 715. Since adoption of the SDRM would require an amendment to the IMF charter, the proposal would have required the affirmative vote of U.S. representatives to the IMF. *See* Dickerson, *supra* note 4, at 1017.

¹²⁹ Ahdieh, *supra* note 44, at 708.

¹³⁰ Breneman, *supra* note 54, at 679.

¹³¹ Snow, *supra* note 126.

¹³² International Monetary and Financial Committee, International Monetary Fund, Communiqué, Dubai (Sept. 21, 2003) *available at* <http://www.imf.org/external/np/cm/2003/092103a.htm>.

¹³³ Ahdieh, *supra* note 44, at 708.

¹³⁴ WORKING GROUP ON CONTRACTUAL CLAUSES, GROUP OF TEN, REPORT OF THE G-10 WORKING GROUP ON CONTRACTUAL CLAUSES 3-6 (2002) [hereinafter WORKING GROUP ON CONTRACTUAL CLAUSES], *available at* <http://www.bis.org/publ/gten08.pdf>. The “Group of 10” “refers to the group of countries that have agreed to participate in the [IMF’s] General Arrangements to Borrow, a supplementary borrowing arrangement that can be invoked if the IMF’s resources are estimated to be below member’s [sic] needs.” INTERNATIONAL MONETARY FUND, FACT SHEET, A GUIDE TO COMMITTEES, GROUPS, AND CLUBS, 4 (2009), *available at* <http://www.imf.org/external/np/exr/facts/pdf/groups.pdf>. The members of the G-10 are: Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland, the United Kingdom, and the United States. *Id.*

¹³⁵ BARRY EICHENGREEN ET AL., INTERNATIONAL MONETARY FUND, CRISIS RESOLUTION: NEXT STEPS, 12-15 (2003).

III. THE MEXICAN AND URUGUAYAN MODELS

A. *The Mexican Model: Rise of the Collective Action Clause*¹³⁶

In February 2003, Mexico became the first major issuer to incorporate collective action clauses (the “CACs”) into sovereign bonds governed by New York law.¹³⁷ Although other large capital markets had included CACs in sovereign bonds for quite some time, the New York markets had been hesitant to incorporate them.¹³⁸ Unlike unanimous action clauses, CACs enable a sovereign to amend certain reserved matters¹³⁹ on an outstanding bond by mere supermajority vote.¹⁴⁰ Both academics and multinational

¹³⁶ Although other nations had previously incorporated collective action clauses into their sovereign bond indentures, Mexico’s debt offering in 2003 was by far the largest and most visible. *See generally* Mark Gugiatti & Anthony Richards, *The Use of Collective Action Clauses in New York Law Bonds of Sovereign Borrowers* 6 (2004) (unpublished manuscript, available at <http://www.law.georgetown.edu/international/documents/gugiatti.pdf>).

¹³⁷ Galvis & Saad, *supra* note 19, at 715; *see also* United Mexican States, Prospectus, at 7 (Dec. 4, 2002).

¹³⁸ These markets include London, Brussels, Luxemburg, and Tokyo. Hagan, *supra* note 15 at 317-18; *see also* Elmar B. Koch, Essay, *Collective Action Clauses: The Way Forward*, 35 GEO. J. INT’L L. 665, 667 (2004). In various forms, collective action clauses have been the norm under English law since the late 19th Century. Andrew G. Haldane et al. *Optimal Collective Action Clause Thresholds* 9 (2004) (unpublished manuscript, available at <http://www.bankofengland.co.uk/publications/workingpapers/wp249.pdf>). However, in the United States, the Trustee Indenture Act of 1939 prohibits the use of CACs in corporate bonds. Trust Indenture Act of 1939 § 316(b), 15 U.S.C. § 77ppp (2004); *see also* Mark J. Roe, *The Voting Prohibition in Bond Workouts*, 97 YALE L.J. 232, 250 (1987). Due in large part to market practice, the prohibition on CACs in the corporate context migrated to sovereign bonds. Bratton & Gulati, *supra* note 7, 55. In 2002, the G-10 Working Group on Contractual Clauses issued a report calling for the inclusion of CACs in future sovereign bond agreements. WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 134, at 3-4. In particular, the Working Group noted that the inclusion of CACs would diverge from market practice in both New York and Germany. *Id.*

¹³⁹ In Mexico’s 2003 issuance, reserved matters included: “payment dates, payment amounts, interest rates, . . . payment currency . . . governing law, specified events of default, *pari passu* ranking, and submission to the jurisdiction of New York courts.” Galvis & Saad, *supra* note 19, at 715-16.

¹⁴⁰ *Id.* at 715. Within the realm of collective clauses, there is much diversity. Although the general approval threshold was set by Mexico at 75%, some countries, such as Brazil, have required up to 85% approval. Likewise, though the majority of collective action clauses measure the voting base as the percentage of all outstanding bondholders, other nations have provided that the voting base will only consist of those holders who are present at a bondholder meeting. Similarly, other issues arise when the issuing nation or a state-owned entity is a holder of its own bonds. To combat the potential of undue influence in the approval process, most indentures have incorporated disenfranchisement clauses that prevent the state or entity from voting

institutions alike view CACs as “the most critical component” of curbing disruptive holdout litigation.¹⁴¹ Since a supermajority of bondholders can impose new repayment terms on recalcitrant holdout creditors,¹⁴² CACs are an effective restraint on the “tyranny of the minority” problem.¹⁴³ To address the new risk of the majority abusing its bargaining power at the expense of minority bondholders,¹⁴⁴ heightened approval thresholds may be utilized.¹⁴⁵ Not surprisingly, CACs have been widely regarded as a necessary but potentially insufficient means to achieve the efficient restructuring of sovereign debt.¹⁴⁶

Pursuant to Mexico’s 2003 bond issuance, three-fourths of bondholders can ratify an amendment to certain reserved matters, such as repayment terms.¹⁴⁷ To curb investor concerns that the English quorum approach¹⁴⁸ would interfere with majority bondholder rights, the Mexican issuance provided for an approval threshold based on the total principal remaining on all outstanding bonds.¹⁴⁹ In addition to CACs, Mexico also incorporated a disenfranchisement clause.¹⁵⁰ As one of the

on matters that require majority approval. *Id.* at 719-22; *see also* WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 131, at 1-6; Brenneman, *supra* note 54, at 681.

¹⁴¹ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 134, at 3.

¹⁴² Brenneman, *supra* note 54, at 681; *see also supra* Part II.

¹⁴³ Buchheit & Gulati, *supra* note 99, at 1325 (quoting FRANCIS B. PALMER, COMPANY PRECEDENTS 271 (2d ed. 1881)).

¹⁴⁴ Fisch & Gentile, *supra* note 12, at 1094-95.

¹⁴⁵ The approval threshold represents the percentage of bondholders that must accept an amendment to the bond’s repayment terms. WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 134, at 4. In its contractual reform recommendations, the G-10 Working Group suggested that a 75% threshold would provide optimal benefits. *Id.* On the one hand, a higher threshold would increase the probability of holdout litigation. *Id.* On the other hand, too low a threshold may enable majority abuse of minority bondholders. Fisch & Gentile, *supra* note 12, at 1094-95. Initially, investors in the United States were hesitant to accept this change out of a concern that the threshold represented the percentage of holders actually present at a bondholders’ meeting. WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 4. To address this concern, the G-10 recommended that the threshold percentage be based on the total principal remaining on all outstanding bonds. *Id.*

¹⁴⁶ *See generally*, WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 3-7 (noting several recommendations to thwart holdouts in sovereign debt restructuring).

¹⁴⁷ United Mexican States, Prospectus, at 7 (Dec. 4, 2002); Galvis & Saad, *supra* note 19, at 715.

¹⁴⁸ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 4. Under the English quorum approach, the approval threshold is based on the percentage of bonds that are represented at the bondholders’ meeting, *not* the total number of bonds outstanding. Galvis & Saad, *supra* note 19, at 719.

¹⁴⁹ United Mexican States, Prospectus, at 7 (Dec. 4, 2002).

¹⁵⁰ Galvis & Saad, *supra* note 19, at 715.

recommendations made by the G-10,¹⁵¹ disenfranchisement clauses ensure that “[b]onds owned or controlled directly or indirectly, by the Issuer or by any public sector instrumentality of the Issuer . . . be disregarded and deemed not to be [o]utstanding.”¹⁵² To curb concerns over potential vote manipulation by the sovereign debtor,¹⁵³ such provisions limit the ability of a sovereign to distort the outcome of a proposed debt restructuring by having bondholders vote against their interests and in favor of the sovereign’s dictates.¹⁵⁴ Although Mexico limited the scope of its disenfranchisement clause,¹⁵⁵ the 2003 issuance did prohibit bonds “owned directly or indirectly by the [Mexican] federal government” from being counted in any subsequent vote.¹⁵⁶ Within a year of Mexico’s drastic contractual reforms, both CACs and disenfranchisement clauses became standard market practice in New York.¹⁵⁷

¹⁵¹ In 2002, the G-10 formed a Working Group on Contractual Clauses “to consider how sovereign debt contracts could be modified in order to make the resolution of debt crises more orderly.” WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 1. To that end, in September of 2002, the Working Group issued a Report with recommendations of contractual provisions to include in future sovereign financing agreements. *Id.* For the Working Group, the objectives to be achieved were:

- (i) to foster early dialogue, coordination, and communication among creditors and a sovereign caught up in a sovereign debt problem;
- (ii) to ensure that there are effective means for creditors and debtors to re-contract, without a minority of debt-holders obstructing the process; and
- (iii) to ensure that disruptive legal action by individual creditors does not hamper a workout that is underway, while protecting the interest of the creditor group.

Id.

¹⁵² *Id.* at 7.

¹⁵³ Galvis & Saad, *supra* note 19, at 720.

¹⁵⁴ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 4.

¹⁵⁵ The wording of the Mexican disenfranchisement clause is somewhat narrower than that suggested by the G-10. Galvis & Saad, *supra* note 19, at 720. Under the G-10’s wording, bonds “owned or controlled” by the sovereign would be prohibited. WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 17. Because Mexico limited its provision to bonds “owned” by the federal government, this might be viewed as more favorable to the sovereign debtor. Galvis & Saad, *supra* note 19, at 720. Although most other sovereigns have followed Mexico’s lead, Uruguay adopted the G-10’s recommendation word-for-word. *Id.*

¹⁵⁶ *Id.*

¹⁵⁷ INTERNATIONAL MONETARY FUND, PROGRESS REPORT TO THE INTERNATIONAL MONETARY AND FINANCIAL COMMITTEE ON CRISIS RESOLUTION 5 (Apr. 20, 2004), available at <http://www.imf.org/external/np/pdr/cr/2004/eng/042004.pdf>. In less than a year after Mexico made its initial offering using CACs, over 11 countries, including Chile, Indonesia, Israel, Italy, Peru, Poland, Turkey and Venezuela, incorporated CACs into their bonds governed by New York law. *Id.* at 3; see also Koch, *supra* note 138, at 673. Indeed, although “there were no sovereign bonds with CACs on

However, while the Mexican reforms were necessary, they were insufficient to achieve effective creditor cooperation in the absence of other coercive legal remedies.¹⁵⁸

B. *The Uruguayan Model*

In March 2003, Uruguay became the second country to issue sovereign bonds incorporating CACs.¹⁵⁹ Like the Mexican model, Uruguay provided for both a 75% approval threshold on reserved matters¹⁶⁰ as well as an issuer disenfranchisement provision.¹⁶¹ In addition to the incorporation of reforms adopted from the Mexican model,¹⁶² the Uruguayan issuance also included aggregation principles¹⁶³ and a weak-trustee structure.¹⁶⁴ When compared to the Mexican reforms, the Uruguayan additions appear to provide a superior means to tackle several of the unresolved collective action problems.¹⁶⁵ However, though the Uruguayan issuance appears to better constrain the power of holdout creditors, it too fails to fully address the collective action crisis of sovereign debt restructuring.¹⁶⁶

the New York market in 2002, in 2003 nearly 50% . . . of all new sovereign bonds under New York law included CACs." *Id.*

¹⁵⁸ Robert B. Gray, Chairman, Int'l. Primary Mkt. Ass'n., Remarks at UNCTAD Fourth Inter-Regional Debt Management Conference (Nov. 11, 2003), available at <http://www.efmagroup.net/getdoc/7514dd4b-4c34-4bc0-a266-f77648b5638a/111103-RBG-UNCTAD-Speech-PDF.aspx>.

¹⁵⁹ República Oriental, Indenture, *supra* note 21, at 13-15; *see also* Galvis & Saad, *supra* note 19, at 717.

¹⁶⁰ República Oriental, Indenture, *supra* note 21, at 36. Under the Uruguayan issuance, reserved matters are very similar to those included under the Mexican model. *See supra* note 139 and accompanying text. In Uruguay's 2003 issue, reserved matters include: payment dates, principal amounts, interest rates, currency, percentage of votes required for taking any action, *pari passu* rankings, governing law, and submission to New York courts' jurisdiction. República Oriental, Indenture, *supra* note 21, at 38.

¹⁶¹ Unlike the Mexican issuance, the Uruguayan disenfranchisement clause mirrored the G-10 recommendations exactly. Galvis & Saad, *supra* note 19, at 720; *see also* República Oriental, Indenture, *supra* note 21, at 25.

¹⁶² Galvis & Saad, *supra* note 19, at 717.

¹⁶³ República Oriental, Indenture, *supra* note 21, at 36; *see also* Galvis & Saad, *supra* note 19, at 722-23.

¹⁶⁴ República Oriental, Indenture, *supra* note 21, at 15; *see also* Galvis & Saad, *supra* note 19, at 724.

¹⁶⁵ *See infra* Part III.B.1-2.

¹⁶⁶ *See infra* Part IV.

1. Aggregation

Under the Mexican model, voting provisions and approval thresholds apply individually to each bond series, and as a result, hinder the efficient restructuring of sovereign debt.¹⁶⁷ In the absence of aggregation, an issuer must seek approval of a restructuring plan from the requisite percentage of holders in each individual bond series.¹⁶⁸ Consequently, collective action problems arise both among bondholders within the same class, as well as among the various series of bonds.¹⁶⁹ As the number of series increases, or when different modification provisions govern several different series of bonds, this process becomes progressively complex.¹⁷⁰ Similarly, the repeated renegotiation of identical terms across multiple bond series can prove to be incredibly inefficient to the sovereign debt restructuring process.¹⁷¹ Without aggregation, a group of rogue bondholders within a single series can hold up a potentially successful restructuring.¹⁷² In an effort to ameliorate these holdout creditors and move the restructuring process forward, a sovereign may “purchase” the consent of dissenting creditors through side payments, and inadvertently create a run on the sovereign debtor’s assets.¹⁷³ Moreover, even

¹⁶⁷ See Galvis & Saad, *supra* note 19, at 722-23. A single bond issuance may incorporate multiple series of bonds. For example, after the debt crisis of 2001, Argentina had to restructure 152 different bonds, issued in 14 different countries, denominated in seven different currencies, and subject to eight different governing laws. Dr. Guillermo Nielsen, Argentine Republic Sec’y of Finance, Speech at Dubai on Argentina’s Restructuring Guidelines (Sept. 22, 2003), available at http://www.argentinedebtinfo.gov.ar/documentos/discurso_gn_dubai_con_diap_english.pdf.

¹⁶⁸ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 3.

¹⁶⁹ Galvis & Saad, *supra* note 19, at 722.

¹⁷⁰ David A. Skeel, Jr., Review Essay, *Can Majority Voting Provisions Do It All?*, 52 EMORY L. J. 417, 422-23 (2003). For example, an issuer could experience significant problems if one series of bonds is governed by CACs and another series has incorporated UACs.

¹⁷¹ Barry Eichengreen & Ashoka Mody, *Is Aggregation a Problem for Sovereign Debt Restructuring?* 1 (Jan. 2003) (unpublished manuscript, available at, <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=801485>); see also INTERNATIONAL MONETARY FUND, LEGAL DEPARTMENT, THE RESTRUCTURING OF SOVEREIGN DEBT—ASSESSING THE BENEFITS, RISKS, AND FEASIBILITY OF AGGREGATING CLAIMS 8 (2003) (available at <http://www.imf.org/external/np/pdr/sdrm/2003/090303.pdf>).

¹⁷² Fisch & Gentile, *supra* note 12, at 1094. The success of such a holdout strategy will ultimately depend on whether the bonds incorporate UACs or CACs. LEGAL DEPARTMENT, INTERNATIONAL MONETARY FUND, THE RESTRUCTURING OF SOVEREIGN DEBT—ASSESSING THE BENEFITS, RISKS, AND FEASIBILITY OF AGGREGATING CLAIMS 8 (2003).

¹⁷³ Eichengreen & Mody, *supra* note 171, at 3. Under the United States Bankruptcy Code, private debtors and creditors can avoid this outcome because of the effect of the automatic stay (which halts attempts by creditors to collect on their debt

assuming that a change in repayment terms could be effectuated across multiple series of bonds, the size of a single holdout's stake may be large enough to make the entire restructuring meaningless.¹⁷⁴

To address these efficiency issues and conform to the contractual recommendations of the G-10,¹⁷⁵ Uruguay became the first sovereign to incorporate aggregation principles that provide for the cram down modification of reserved matters across multiple series of bonds.¹⁷⁶ Under this provision, an amendment to repayment terms can be imposed against multiple bond series.¹⁷⁷ Specifically, cram down can occur if the proponents of the modification obtain the support of “[h]olders of not less than 85% in aggregate principal amount of the Outstanding Debt Securities of all Series affected by that Modification (taken in aggregate) . . . and [h]olders of not less than 66-2/3% in aggregate principal amount of the Outstanding Debt Securities of that Series (taken individually).”¹⁷⁸ When combined with Uruguay's 75% approval threshold CAC,¹⁷⁹ aggregation allows the issuer to impose repayment term amendments on a one-third-minority holdout.¹⁸⁰

Most importantly, the incorporation of an aggregation clause encourages the type of collaboration and shared sacrifice that was commonplace during the era of syndicated bank lending.¹⁸¹ Because of cram down, aggregation permits the

during the pendency of the case) and avoidable preference provisions (which void transactions that were made on the eve of filing the bankruptcy petition). 11 U.S.C. §§ 362, 547 (2005).

¹⁷⁴ For example, if the holdout was the cause of the sovereign's financial crisis.

¹⁷⁵ Although the G-10 Working Group did “not [focus] on the technicalities of [aggregation provisions] in any detail,” their 2002 report did note that such clauses have “a great deal of potential” and “[merit] further exploration.” WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 138, at 6.

¹⁷⁶ Alinna Arora & Rodrigo Olivares Caminal, *Rethinking the Sovereign Debt Restructuring Approach*, 9 L. & BUS. REV. AM. 629, 663-64 (2003).

¹⁷⁷ Galvis & Saad, *supra* note 19, at 722.; *see also* WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6-7.

¹⁷⁸ República Oriental, Indenture, *supra* note 21, at, 36; *see also* Galvis & Saad, *supra* note 19, at 722.

¹⁷⁹ Galvis & Saad, *supra* note 19, at 723.

¹⁸⁰ By providing for aggregation, the Uruguayan model “effectively reduces” the approval threshold “from 75% to two-thirds.” Galvis & Saad, *supra* note 19, at 723. For example, under CAC with a 75% approval threshold, a minority faction of one-third of outstanding bondholders in a single series can block any amendment to reserved matters for that series. With aggregation, however such holdouts have less power. If the proponents of a reserved matter modification can obtain the approval of 85% of the aggregate principle of all outstanding series, the amendment can be crammed down on a single series with no more than one-third holdouts.

¹⁸¹ *See supra* Part I.B.

issuer to focus on areas of collective agreement across multiple bond series.¹⁸² Similarly, the threat of cram down encourages the holders of different bonds to work together to arrive at a settlement that is jointly advantageous.¹⁸³ With the presence of a highly liquid secondary market, moreover, recalcitrant bondholders remain free to avoid what they may deem as inequitable concessions by selling their bonds in the open market.¹⁸⁴ As a result, the Uruguay model promotes and fosters collaboration among creditors while providing an avenue for those who wish to opt out of the process.¹⁸⁵

2. Fiscal Agency and Trust Structures

While collective action clauses make the restructuring of sovereign debt somewhat easier, they only solve a portion of the holdout problem.¹⁸⁶ Under both UACs and CACs, sovereign bonds issued pursuant to New York law generally incorporate a fiscal agency structure.¹⁸⁷ Under this approach, each bondholder retains an individual right to seek legal remedies against the sovereign debtor in the event of default.¹⁸⁸ Although direct legal actions were at one time quite rare,¹⁸⁹ litigation to collect against sovereign debtors is increasing.¹⁹⁰ In the absence of “sharing clauses,”¹⁹¹ litigating creditors under both the Mexican and the Uruguayan models do not have to divide legal awards

¹⁸² Galvis & Saad, *supra* note 19, at 722.

¹⁸³ Fisch & Gentile, *supra* note 12, at 1090-95.

¹⁸⁴ *Id.*

¹⁸⁵ *Id.*

¹⁸⁶ Fisch & Gentile, *supra* note 12, at 1093-95; *see also* Skeel, *supra* note 170, at 423-24.

¹⁸⁷ Galvis & Saad, *supra* note 19, at 723; *see also* WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 138, at 6. Although sovereign bonds issued in England have incorporated trust deeds for quite some time, sovereign bonds in the United States typically utilize a fiscal agent. WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6; Fisch & Gentile, *supra* note 12, at 1102.

¹⁸⁸ INTERNATIONAL MONETARY FUND, PROGRESS REPORT TO THE INTERNATIONAL MONETARY AND FINANCIAL COMMITTEE ON CRISIS RESOLUTION 10 n.16 (Apr. 20, 2004) [hereinafter IMF, PROGRESS REPORT], *available at* <http://www.imf.org/external/np/pdr/cr/2004/eng/042004.pdf>.

¹⁸⁹ Bratton & Gulati, *supra* note 7, at 34.

¹⁹⁰ IMF, PROGRESS REPORT, *supra* note 188, at 13.

¹⁹¹ For example, in the event that a court awards a litigating bondholder a “disproportionate” judgment, a sharing clause may require that bondholder to turn over any overpayment to the fiscal agent for a pro rata distribution to other bondholders. Although such clauses were common during the era of syndicated bank lending, they are rarely found in sovereign bond financing. Lee C. Buchheit, *Changing Bond Documentation: The Sharing Clause*, 17 INT’L FIN. L. REV. 17, 17-18 (1998).

pro rata with fellow bondholders.¹⁹² Because sovereign debt restructuring qualifies as a “zero sum game,”¹⁹³ litigation becomes “infectious”¹⁹⁴ as creditors race to seize a defaulting sovereign’s assets.¹⁹⁵ Accordingly, though the introduction of CACs and aggregation principles begin to address the holdout problem, civil suits by dissenting bondholders continue to reduce both the net pool of assets available to other creditors as well as the potential for a successful restructuring of the sovereign’s outstanding debt.¹⁹⁶

Under the Mexican model, the fiscal agent is a relatively weak entity that controls merely the distribution of payments and simple forms of interaction between the issuer and the bondholders.¹⁹⁷ As an agent of the sovereign debtor, the fiscal agent does not represent the interests of the bondholder class.¹⁹⁸ Pursuant to most Fiscal Agency Agreements,¹⁹⁹ the fiscal agent “acts solely . . . for the issuer and does not have any fiduciary relationship to the bondholders.”²⁰⁰ As a result, the fiscal agent has very limited bondholder duties.²⁰¹ In most cases, these obligations are confined to: giving notice of specified events, assembling a bondholder meeting if petitioned by the requisite percentage of bondholders, and appointing a chairperson at the bondholder meeting.²⁰² Given that the fiscal agent has no power to file suit against the sovereign debtor,²⁰³ and that the creditors retain an individual right to litigate on their claims,²⁰⁴ the fiscal agency structure is ineffective in preventing disruptive litigation on the part of holdout creditors.²⁰⁵

¹⁹² Brenneman, *supra* note 54, at 680.

¹⁹³ Buchheit, *supra* note 191, at 18. In other words, the sovereign’s assets that are available to satisfy bondholder debt are limited. Therefore, as one creditor collects on its claim, another creditor is left with fewer assets to satisfy its claim.

¹⁹⁴ *Id.*

¹⁹⁵ *Id.*

¹⁹⁶ Fisch & Gentile, *supra* note 12, at 1093-95 .

¹⁹⁷ Macmillan, *supra* note 33, at 65-66.

¹⁹⁸ Fisch & Gentile, *supra* note 12, at 1102 .

¹⁹⁹ The Fiscal Agency Agreement is the controlling document that governs the sovereign debtor and fiscal agent relationship. Macmillan, *supra* note 56, at 341.

²⁰⁰ *Id.* at 341-42.

²⁰¹ *Id.* at 341.

²⁰² *Id.*

²⁰³ See also Working Group on Contractual Clauses, Group of Ten, Report of the G-10 Working Group on Contractual Clauses (2002), available at <http://www.bis.org/publ/gten08.pdf>.

²⁰⁴ Fisch & Gentile, *supra* note 12, at 1102.

²⁰⁵ *Id.* at 1103.

To avoid some of the collective action and efficiency problems inherent in multiple civil suits,²⁰⁶ the Uruguayan model incorporates a weakened trustee structure instead of the fiscal agency model.²⁰⁷ Although the trustee structure does not preclude individual bondholders from filing suit to recover outstanding amounts payable,²⁰⁸ the trustee does have the power to initiate legal action on behalf of the bondholder class.²⁰⁹ Accordingly, in the event of a default, the trustee is an “identifiable leader” to coordinate collective bondholder action.²¹⁰ Similarly, when engaged in litigation, the trustee acts for the benefit of the entire bondholder class and distributes any resulting award pro rata.²¹¹ In accordance with G-10 recommendations,²¹² the trustee is also responsible for gathering and distributing financial information concerning the sovereign debtor in the event of a debt restructuring.²¹³ Yet, while the Uruguayan trustee structure plays a more prominent role in addressing the holdout problem, the model fails to

²⁰⁶ See *supra* Part II.B.

²⁰⁷ República Oriental, Indenture, *supra* note 21, at 17; see also Part IV. Uruguay was the first nation to utilize an Indenture Trustee in a New York sovereign financing agreement. Galvis & Saad, *supra* note 19, at 724. In addition, by incorporating a weak trustee structure, the success of Uruguay’s issuance also demonstrated a market willingness to move away from the traditional fiscal agency model.

²⁰⁸ In particular, if the Republic fails to make payments when due, individual bondholders can sue to recover. República Oriental, Indenture, *supra* note 21, at 15; see also Arora & Caminal, *supra* note 176, at 663; Galvis & Saad, *supra* note 19, at 724.

²⁰⁹ The trustee can initiate such action on the request of 25% of outstanding bondholders. República Oriental, Indenture, *supra* note 21, at 15. Pursuant to the indenture, “the Trustee, in its own name . . . shall be entitled and empowered to institute any action or proceedings at law or in equity for the collection of . . . sums . . . due and unpaid.” *Id.* at 13. Some academics suggest that the primary benefits of a trustee could be achieved without shifting from the fiscal agent structure. By “concentrat[ing] ‘the right to sue’ in a single representative of bondholders” the same benefits could be obtained. Galvis & Saad, *supra* note 19, at 723-25.

²¹⁰ Macmillan, *supra* note 56, at 341.

²¹¹ Galvis & Saad, *supra* note 19, at 722-24.

²¹² WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 3.

²¹³ Such information includes:

- (i) a description of the economic or financial circumstances that . . . explain the request for the proposed Modification;
- (ii) if the Republic . . . [has] entered into a standby, extended funds, or similar program with the International Monetary Fund . . . ; and
- (iii) a description of the Republic’s proposed treatment of its other major creditor groups

República Oriental, Indenture, *supra* note 21, at 37; see also, Galvis & Saad, *supra* note 19, at 722; Anna Gelpern, *How Collective Action is Changing Sovereign Debt*, 22 INT’L FINANCIAL L. REV., 19 (2003).

prevent rogue litigation,²¹⁴ and as a result, collective action problems remain.

IV. THE SUPER TRUSTEE SOLUTION

In 2002, when the G-10 reported on contractual solutions to the sovereign debt crisis, it recommended both the inclusion of CACs²¹⁵ and the incorporation of a “super” trustee structure.²¹⁶ Under the “super” trustee model, bondholders generally do not have the right to bring legal actions in their individual capacity.²¹⁷ Rather, the authority to file suit against the sovereign debtor usually lies solely with an indenture trustee.²¹⁸ As a result, litigation can only be brought on the trustee’s own initiative or upon the direction of a specified percentage of outstanding bondholders.²¹⁹ Similar to the Uruguay model, if any resulting legal action proves successful, the trustee, as representative of the entire bondholder class, must share any award pro rata.²²⁰

For the better part of the last century, the Trust Indenture Act of 1939 (the “Act”) has mandated a trust indenture structure for public corporate bonds.²²¹ Under the Act, the trustee is an agent of the bondholders and owes to them a duty of good faith.²²² Although the trustee’s duties are limited outside of the default scenario,²²³ the trustee does ensure compliance with the terms of the indenture even when the debtor is paying as required.²²⁴ If a debtor defaults,

²¹⁴ Galvis & Saad, *supra* note 19, at 723-24. Although the Uruguayan Trustee curbs holdout litigation on accelerated amounts (those payments not yet due), it fails to effectively control individual legal actions for past amounts due. *Id.* at 724 n.23.

²¹⁵ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 3.

²¹⁶ Other recommendations included: revised provisions for calling bondholder meetings; majority enforcement of acceleration clauses; provisions requiring appropriate information to be disseminated to bondholders; and disenfranchisement provisions from the issuer of the bonds. *See id.* at 2-7.

²¹⁷ IMF, PROGRESS REPORT, *supra* note 157, at 10 n.16.

²¹⁸ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6-7.

²¹⁹ IMF, PROGRESS REPORT, *supra* note 157, at 10 n.16.

²²⁰ WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6-7.

²²¹ Yakov Amihud, Kenneth Garbade, & Marcel Kahan, *A New Governance Structure For Corporate Bonds*, 51 STAN. L. REV. 447, 485 (1999). In addition, the Trust Indenture Act of 1939 explicitly exempts governments, both domestic and foreign, from its requirements. 15 U.S.C.A. § 77ddd(a)(6) (1998); *see also* Macmillan, *supra* note 56, at 339-41.

²²² Macmillan, *supra* note 586, at 339-41.

²²³ Fisch & Gentile, *supra* note 12, at 1105.

²²⁴ Macmillan, *supra* note 586, at 339-40.

however, the trustee's duties become much more complex.²²⁵ In accordance with the Act, the debtor's default triggers the trustee's fiduciary duties to the bondholder class.²²⁶ In addition, the trustee is the only entity that is able to accelerate principal amounts due on outstanding bonds.²²⁷ Unlike a fiscal agent, the trustee acts as a fiduciary for bondholders, and thus only the trustee may file suit against the debtor.²²⁸ Unless the trustee fails to comport with its fiduciary obligations, bondholders are limited in the types of lawsuits they can bring.²²⁹ Consequently, the Act both limits the ability of holdouts to pursue obstructive litigation tactics and provides bondholders with a centralized fiduciary to enforce the payment obligations of recalcitrant corporate debtors.²³⁰

Although the Uruguayan model provides for an indenture trustee with some control over the sovereign debt restructuring process, it fails to solve the holdout dilemma, because rogue creditors can continue to file adversary actions.²³¹ By incorporating CACs and aggregation principles but failing to preclude suits by individual bondholders,²³² the model fails to live up to its full potential.²³³ Under a "super" trustee approach akin to that required by the Trust Indenture Act, the holdout problem could be greatly curbed.²³⁴ Whereas CACs and aggregation clauses in the Uruguay model prevent holdout creditors from halting the restructuring process itself, the problem of a race to the sovereign debtor's assets remains.²³⁵ By entrusting an individual or entity with exclusive power to file suit for default, the "super" trustee structure prevents vulture funds and rogue creditors from disrupting the restructuring process with threats of costly and cumbersome litigation.²³⁶ Similarly, the pool of assets to be distributed among equally

²²⁵ *Id.* at 339-41.

²²⁶ *Id.*

²²⁷ Fisch & Gentile, *supra* note 12, at 1104.

²²⁸ *Id.* at 1105. Although the Act provides that public, corporate bondholders have an absolute right to sue for past amounts due (as opposed to accelerated amounts), this provision could be removed from sovereign bond trustee indentures. See Galvis & Saad, *supra* note 19, at 724 n.23.

²²⁹ Fisch & Gentile, *supra* note 12, at 1105.

²³⁰ *Id.*

²³¹ See Galvis & Saad, *supra* note 19, at 723.

²³² See *id.* at 723-25.

²³³ See Fisch & Gentile, *supra* note 12, at 1094.

²³⁴ See *id.*

²³⁵ See *id.*

²³⁶ See WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6-7.

situated bondholders is not raided, but distributed pro rata.²³⁷ In addition, the case law and legal theories that have been applied to the trustee structure in the corporate context for almost three-fourths of a century could easily be transplanted to sovereign debt.²³⁸ Furthermore, fiduciary duties should curb fears that a trustee will not be aggressive in defending bondholders' interests.²³⁹ Accordingly, the application of a "super" trustee structure should be the next step in solving the holdout crisis at the core of the sovereign debt dilemma.²⁴⁰

CONCLUSION

Although sovereign financing has undergone significant contractual reforms over the past decade,²⁴¹ these efforts have generally failed to adequately address the inefficiencies created by holdout strategies. In the absence of a global sovereign debt restructuring regime, both creditors and debtors will need to continue to rely on contractual methods to effectuate sovereign debt restructuring.²⁴² Notwithstanding the laudable improvements made by Mexico and Uruguay,²⁴³ additional refinements are necessary. In particular, the sovereign financing market should move towards the incorporation of a super trustee indenture. With a trustee to coerce creditor cooperation and ensure equitable treatment among bondholders, the super trustee fills in the gaps left by the early reforms. Therefore, the super trustee is necessary to curb the holdout problem and finally extinguish the sovereign debt dilemma.

James M. Hays II

²³⁷ See *id.*

²³⁸ See Macmillan, *supra* note 33, at 65.

²³⁹ Fisch & Gentile, *supra* note 12, at 1107.

²⁴⁰ See WORKING GROUP ON CONTRACTUAL CLAUSES, *supra* note 1384, at 6-7.

²⁴¹ See *supra* Part III.

²⁴² See *supra* Part II.B.2.

²⁴³ See *supra* Part III.

[†] J.D. Candidate, Brooklyn Law School, 2010; B.A., Southern Methodist University, 2003. For their unwavering support and continued encouragement, I extend my sincerest thanks and appreciation to my family. In particular, I would like to acknowledge my parents for instilling in me a regard for intellectual curiosity and a respect for perseverance. Likewise, I am deeply grateful to Molly E. Madden for her steadfast belief in me and her invaluable role in both my personal and professional life. Finally, I would like to thank the members of the *Brooklyn Law Review* for their diligent work in editing this Note.

Unscrambling the Organic Eggs

THE GROWING DIVERGENCE BETWEEN THE DOJ AND THE FTC IN MERGER REVIEW AFTER *WHOLE FOODS*

INTRODUCTION

When the merger between Whole Foods Market Inc. (“Whole Foods”) and Wild Oats Markets, Inc. (“Wild Oats”) was first announced in 2007, few people suspected that this \$565 million merger would set off such a massive (organic) food fight, forcing the Federal Trade Commission (“FTC”), the Department of Justice (“DOJ”), and the judiciary to re-evaluate their roles in the merger process.¹ On March 6, 2009, Whole Foods and the FTC announced a settlement.² The tortuous legal battle between the FTC and Whole Foods might be over, but the most important result of the battle is not the settlement between the FTC and Whole Foods. Rather, it is the D.C. Circuit Court of Appeals’ controversial decision in *FTC v. Whole Foods Market, Inc.*, which articulated a preliminary injunction standard making it much easier for the FTC to block future mergers, compared to its antitrust enforcement counterpart at the DOJ.³ In a system of shared responsibility for enforcement of the federal antitrust laws, such an outcome is unacceptable. Merging parties should expect the same treatment and burden in the merger review process. The substantive outcome of a proposed merger should not depend on the arbitrary allocation of the merger to either the FTC or the DOJ for review. However, after *Whole Foods*, the outcome of a proposed transaction might very well depend on which antitrust enforcement agency is reviewing it.

¹ See Andrew Martin, *Whole Foods Makes Offer for a Smaller Rival*, N.Y. TIMES, Feb. 22, 2007, at C1. The deal was a tender offer for all of Wild Oats stock at a price of \$18.50 per share—a 23% premium over the average share price in January of 2007. *Id.*

² Press Release, Fed. Trade Comm’n, FTC Consent Order Settles Charges that Whole Foods’ Acquisition of Rival Wild Oats was Anticompetitive (March 6, 2009), available at <http://www.ftc.gov/opa/2009/03/wholefoods.shtm>.

³ 548 F.3d 1028 (D.C. Cir. 2008), *amended and reissued*, 548 F.3d 1028 (D.C. Cir. 2008).

In *Whole Foods*, the FTC sought a preliminary injunction to stop the proposed merger of two organic supermarkets, Whole Foods and Wild Oats.⁴ The FTC argued that Whole Foods and Wild Oats were the two largest competitors in the “premium, natural, and organic supermarkets” or “PNOS” market and a merger of the two companies would harm consumers by reducing competition in a number of geographic markets.⁵ The D.C. District Court rejected the FTC’s argument that PNOS were a distinct market and concluded that a merger of Whole Foods and Wild Oats would not substantially lessen competition in the broad market of all supermarkets.⁶ Nearly a full year after the closing of the merger and the integration of the two firms,⁷ a panel of D.C. Circuit judges (Judge Brown, Judge Tatel, and Judge Kavanaugh) issued three separate opinions reversing the denial of the preliminary injunction.⁸ In *Whole Foods*, the D.C. Circuit explicitly articulated a standard that significantly reduces the FTC’s burden of proof in its request for preliminary injunctions. This lowered preliminary injunction standard, and the ability of the FTC to commence administrative proceedings, create a disturbing perception that the outcome of a challenged merger depends on which agency is reviewing the merger, rather than on the antitrust merits of the case.

The effect of the *Whole Foods* decision has already been felt. Less than a week after Whole Foods and the FTC settled, a \$1.4 billion merger between CCC Holdings, Inc., (“CCC Holdings”) and Mitchell International, Inc., (“Mitchell International”) was abandoned after a judge relied on *Whole*

⁴ Preliminary injunctions are used by the antitrust enforcement agencies to “preserve the status quo by preventing the consummation of a merger.” AMERICAN BAR ASSOCIATION, SECTION OF ANTITRUST LAW, MERGER AND ACQUISITIONS: UNDERSTANDING THE ANTITRUST ISSUES 546 (3d. ed. 2008) [hereinafter UNDERSTANDING THE ANTITRUST ISSUES]. Normally, a plaintiff seeking injunctive relief must prove: (1) irreparable harm if the injunction is not granted, (2) this injury outweighs any harm to the defendant by the injunction, (3) the plaintiff has a substantial likelihood of success on the merits, and (4) the injunction is in the public interest. *Id.* at 564-65. For a more in depth discussion for each of the different factors, see *id.* at 570-95.

⁵ *Whole Foods*, 548 F.3d at 1032.

⁶ *FTC v. Whole Foods Mkt, Inc.*, 502 F. Supp. 2d 1, 49-50 (D.D.C. 2007), rev’d., 533 F.3d 869 (D.C. Cir. 2008).

⁷ See Alicia Wallace, *Boulder’s Whole Foods-Wild Oats: One Year Later*, BOULDER DAILY CAMERA, Aug. 25, 2008, at D1. There were significant changes in personnel, suppliers, distribution systems, and leasing agreements. *Id.*

⁸ *FTC v. Whole Foods Mkt, Inc.*, 533 F.3d 869 (D.C. Cir. 2008), *reprinted as amended*, 548 F.3d 1028 (D.C. Cir. 2008).

Foods to issue a preliminary injunction that stopped the proposed transaction.⁹ In November 2008, the three-judge panel had amended and reissued its original opinions in *Whole Foods*, so that Judge Tatel no longer concurred in Judge Brown's opinion, but rather only with the judgment of the court.¹⁰ As a result of Judge Tatel's revision, Judge Brown's opinion was no longer the majority opinion of the court and there were questions about the precedential value of the decision.¹¹ However, *FTC v. CCC Holdings, Inc.* made it clear that *Whole Foods* and its articulation of the preliminary injunction standard for the FTC was now binding precedent.¹²

In a system where the DOJ and the FTC have shared responsibility for enforcement of the federal antitrust laws, merging parties should expect comparable treatment and burden, as well as a comparable outcome, regardless of whether the FTC or the DOJ is reviewing their merger.¹³ Antitrust enforcement has an enormous impact on the economy, so consistency, predictability, and fairness are crucial in the merger review process. However, the settlement between the FTC and *Whole Foods* after a prolonged and expensive fight, and the termination of the proposed merger between CCC Holdings and Mitchell International, provide disturbing illustrations that the choice of enforcement agency for merger review clearly does influence the outcome of a transaction. In light of *Whole Foods*, the best outcome for parties to a proposed merger would be for the DOJ to clear the proposed transaction.

This Note will address the growing divergence in merger enforcement between the FTC and the DOJ. It argues

⁹ See *FTC v. CCC Holdings, Inc.*, 605 F. Supp. 2d 26, 30 (D.D.C. 2009). On March 9, 2009, Judge Collyer granted the FTC's request for a preliminary injunction. On March 11, 2009, the parties announced the termination of the merger. See Press Release, CCC Information Services Inc., CCC-Mitchell mutually agree to terminate merger (March 11, 2009), available at http://ccc.cccis.com/filebin/pdf/CCCMITCHELL_Nonpursuit.pdf.

¹⁰ *Whole Foods*, 548 F.3d at 1028.

¹¹ See *id.* at 1061 n.8 (Kavanaugh, J., dissenting). According to Judge Kavanaugh, "this confused decision will invite years of uncertainty and litigation over what the holding of this case is—a separate but important problem with the Court's approach." *Id.*

¹² See *FTC v. CCC Holdings, Inc.*, 605 F. Supp. 2d at 36. According to Judge Collyer, "precedents irrefutably teach that in this context 'likelihood of success on the merits' has a less substantial meaning than in other preliminary injunction cases. *Heinz* not only emphasized this point but *Whole Foods* makes clear that *Heinz* remains good law." *Id.* at 36 n.11.

¹³ See ANTITRUST MODERNIZATION COMM'N, REPORT AND RECOMMENDATIONS 129 (2007), available at http://govinfo.library.unt.edu/amc/report_recommendation/amc_final_report.pdf [hereinafter ANTITRUST MODERNIZATION REPORT].

that the FTC's lower preliminary injunction standard and its ability to commence administrative litigation gives the FTC a significant advantage over the DOJ in challenging a merger and extracting a settlement, a result that is unacceptable in a dual enforcement system. Specifically, the Note argues that after *Whole Foods*, the ultimate decision as to whether a merger may proceed depends on which agency is reviewing the transaction, which can lead to both expensive litigation and disruptive post-closing divestitures. Part I examines the relevant antitrust statutes, the enforcement agencies involved, and how the merger review process works. Part II reviews the history of the FTC's challenge to the merger between Whole Foods and Wild Oats. It begins with a discussion of the merger, followed by a discussion of the district court and the D.C. Circuit Court of Appeals' decisions. It concludes with a review of the settlement between the FTC and Whole Foods. Part III discusses how the divergence in preliminary injunction standards applicable to the DOJ and the FTC, and the ability of the FTC to pursue administrative trials, produce inconsistent results between the FTC and the DOJ in merger enforcement. This Part argues that due to these divergences, the choice of which antitrust enforcement agency is to review a proposed merger is outcome-determinative.¹⁴ Finally, Part IV suggests two approaches to harmonizing the divergences: a judicial solution and a legislative solution. It argues that in light of *CCC Holdings, Inc.*, a judicial solution is unlikely, so the most politically promising solution to stem the growing divergence between the DOJ and the FTC enforcement standards is for Congress to amend the Federal Trade Commission Act to specify that the same preliminary injunction standard applies to both enforcement agencies.

¹⁴ See American Bar Association, Public Comments Submitted to AMC Regarding Government Enforcement Institutions: Differential Merger Enforcement Standards, at 9 (Oct. 28, 2005), available at http://govinfo.library.unt.edu/amc/public_studies_fr28902/enforcement_pdf/051028_ABA_Fed_Enforc_Inst_Differential_S tandards.pdf [hereinafter ABA Comments re Differential Standards].

I. THE PROCESS OF MERGER REVIEW BY THE DOJ AND THE FTC

The DOJ's Antitrust Division and the FTC have shared responsibility for enforcement of the federal antitrust laws.¹⁵ To avoid duplication of effort, the agencies consult with one another and a proposed transaction is "cleared" to one agency or the other for review in a process known as the "clearance process."¹⁶ This Part examines how the dual enforcement system functions. It begins with a discussion of the relevant federal antitrust statutes. It then discusses the clearance process between the FTC and the DOJ, the decision to challenge a proposed transaction in court or in an administrative trial (for the FTC), and the potential remedies available to the enforcement agencies, such as divestitures, for a merger with anticompetitive concerns.

A. *Overview of the Applicable Antitrust Statutes*

At the federal level, the framework for the merger review process is contained in a few relevant statutes, namely Sections 1 and 2 of the Sherman Act, Section 7 of the Clayton Act, and Section 5 of the FTC Act.¹⁷ Modern antitrust law really began with the enactment of the Sherman Act in 1890.¹⁸ Section 1 of the Sherman Act prohibits contracts, combinations, or conspiracies in restraint of trade.¹⁹ Section 2 prohibits monopolies and attempts at monopolies.²⁰ The Sherman Act only prohibits restraints of trade that are unreasonable.²¹ To build upon the protection afforded in the Sherman Act,

¹⁵ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 129; Fed. Trade Comm'n, Bureau of Competition, Guide to the Antitrust Laws, the Enforcers, <http://www.ftc.gov/bc/antitrust/enforcers.shtm>.

¹⁶ Fed. Trade Comm'n, Bureau of Competition, Guide to the Antitrust Laws, Mergers: Premerger Notification and the Merger Review Process, http://www.ftc.gov/bc/antitrust/premerger_notification.shtm.

¹⁷ Of course, every state and the District of Columbia have its own statutes regulating the competitive effects of mergers and acquisitions. See Stephen Calkins, *Perspectives on State and Federal Antitrust Enforcement*, 53 DUKE L.J. 673, 678 (2003). However, this note will only focus on the federal statutes.

¹⁸ See Scott A. Sher, *Closed But Not Forgotten: Government Review of Consummated Mergers Under Section 7 of the Clayton Act*, 45 SANTA CLARA L. REV. 41, 45 (2004) (tracking the development of modern antitrust law and analyzing numerous cases involving closed merger challenges by the FTC and DOJ).

¹⁹ 15 U.S.C. § 1 (2006).

²⁰ 15 U.S.C. § 2 (2006).

²¹ Fed. Trade Comm'n, Bureau of Competition, Guide to the Antitrust Laws, the Antitrust Laws, http://www.ftc.gov/bc/antitrust/antitrust_laws.shtm.

Congress passed the Clayton Act in 1914,²² amended it in 1950 with the Celler-Kefauver amendments to close some loopholes,²³ and amended it again in 1976 with the Hart-Scott-Rodino Antitrust Improvement Act of 1976 (“HSR Act”).²⁴

1. The Clayton Act

Today, the principal federal antitrust statute is the Clayton Act, specifically Section 7, which prohibits mergers or acquisitions “in any line of commerce or in any activity affecting commerce in any section of the country, [when] the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly.”²⁵ Even though the Clayton Act was much more expansive than the Sherman Act, the lack of a requirement for pre-closing notifications meant that the government could challenge an anticompetitive transaction only after it closed.²⁶ By then, it was often too late to enforce the Clayton Act.²⁷ Aware of the substantial costs and time involved in such post-consummation challenges, Congress enacted the HSR Act with the goal of “giving the government antitrust agencies a fair and reasonable opportunity to detect and investigate large mergers of questionable legality before they are consummated.”²⁸ The HSR Act and the establishment of the premerger notification program would give the antitrust enforcement agencies such an opportunity.

2. The Hart-Scott-Rodino Act

The HSR Act is often credited with establishing the modern merger review process by giving the DOJ and the FTC

²² Clayton Act, 15 U.S.C. § 18 (2006).

²³ Celler-Kefauver Amendments, Pub. L. No. 81-899, 64 Stat. 1125 (Codified as amended at 15 U.S.C. § 18 (2006)). After the Celler-Kefauver amendments, Section 7 of the Clayton Act covers both asset acquisitions and stock acquisitions. *See* *Brown Shoe Co. v. United States*, 370 U.S. 294, 316 (1962). In *Brown Shoe*, the Supreme Court reviewed the legislative history and purpose of the amendments. *See id.* at 315-23.

²⁴ Hart-Scott-Rodino Antitrust Improvement Act of 1976, Pub. L. No. 94-435, 90 Stat. 1383 (Current version at 15 U.S.C. § 18a (2006)).

²⁵ 15 U.S.C. § 18 (2006).

²⁶ *See* *Sher*, *supra* note 18, at 52-53.

²⁷ Litigation often took years and even if the government won, there was often no remedy because the firms were already well integrated. *Id.* at 52-54.

²⁸ H.R. REP. No. 94-1373, at 5 (1976), *as reprinted in* 1976 U.S.C.C.A.N. 2637, 2637.

the ability to block mergers before consummation.²⁹ Before the passage of the HSR Act, it was very difficult to challenge a merger successfully.³⁰ Without advance notice of the transaction, mergers were typically challenged after they were already consummated.³¹ The government also had very little time to prepare, and carried the burden of proof for obtaining a preliminary injunction.³² Since these challenges often took years to litigate, it was very difficult for courts to come up with an appropriate remedy to restore competition—“that is, to unscramble the eggs”—because it was very difficult to recreate the acquired entity as an independent “competitively viable firm.”³³ So even when the government was successful in its challenge, it was often a hollow victory and too late to gain any “meaningful relief.”³⁴

The Hart-Scott-Rodino Act changed all of this by requiring the parties to notify the FTC and the DOJ about mergers and acquisitions of certain sizes before they occur, and to give the antitrust agencies time to review such transactions before consummating the proposed transaction.³⁵ Under the HSR Act, the parties to certain proposed transactions must notify both the FTC and the DOJ by submitting a “Notification and Report Form” with some information about the parties and

²⁹ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 47, 151; Sher, *supra* note 18, at 52-54.

³⁰ See H.R. REP. No. 94-1373, at 8 (1976), as reprinted in 1976 U.S.C.C.A.N. 2637, 2640; see also Sher, *supra* note 18, at 52-54 (discussing the difficulties of post-consummation challenges).

³¹ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 47; Sher, *supra* note 18, at 52-53.

³² See H.R. REP. No. 94-1373, at 8 (1976), as reprinted in 1976 U.S.C.C.A.N. 2637, 2640 (“[W]ithout advance notice of an impending merger, data relevant to its legality, and at least several weeks to prepare a case, the government often has no meaningful chance to carry its burden of proof, and win a preliminary injunction against a merger that appears to violate section 7. The weight of this burden cannot be overemphasized.”).

³³ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 47. “Unscrambling the eggs” is a term used to express the difficulty of a divestiture remedy when a merger is already closed and the assets of the combined firms are integrated. H.R. REP. No. 94-1373, at 4-5 (1976), as reprinted in 1976 U.S.C.C.A.N. 2637, 2640-41 (After closing, “the acquired firm’s assets, technology, marketing systems, and trademarks are replaced, transferred, sold off, or combined with those of the acquiring firm. Similarly, its personnel and management are shifted, restrained, or simply discharged. In these ways, the acquiring and acquired firms are, in effect, irreversibly ‘scrambled’ together.”).

³⁴ H.R. REP. No. 94-1373, at 8 (1976), as reprinted in 1976 U.S.C.C.A.N. 2637, 2640. See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 151.

³⁵ See 15 U.S.C. § 18a(a) (2006); see also Fed. Trade Comm’n, Bureau of Competition, Guide to the Antitrust Laws, Mergers: Premerger Notification and the Merger Review Process, http://www.ftc.gov/bc/antitrust/premerger_notification.shtm.

the proposed transaction.³⁶ The HSR Act does not require a premerger filing for all mergers or acquisitions.³⁷ The filing thresholds are updated annually, but generally, the parties must be of a certain size and the deal must be of a certain value.³⁸ Under the HSR Act, advance notice must be provided to both the DOJ and the FTC even though only one agency will review the proposed merger.³⁹ The HSR Act also enables the agencies to obtain documents and other necessary information from the parties and third parties to assess whether to challenge the proposed transaction.⁴⁰ Congress's solution to the time constraint problem was to establish a thirty-day waiting period.⁴¹ During this time, the parties are prohibited from closing their deal unless the waiting period is granted early termination by the FTC or the DOJ.⁴² As a result of the HSR Act, challenges to consummated deal are relatively rare because the agencies are able to challenge mergers before they are consummated.⁴³ Nevertheless, over the last decade, the FTC has been much more aggressive in challenging closed deals where the anticompetitive concerns were not apparent during the merger review process.⁴⁴

³⁶ Fed. Trade Comm'n, Bureau of Competition, *Hart-Scott-Rodino Premerger Notification Program Introductory Guide I, What is the Premerger Notification Program? An Overview*, at 6, (2008) available at <http://www.ftc.gov/bc/hsr/introguides/guide1.pdf> [hereinafter *Guide I*]. Copies of the form and instructions are available at <http://www.ftc.gov/bc/hsr/hsrforms.htm>.

³⁷ In addition to the size of the parties or of the deal, there are a limited number of exceptions to the HSR Act. See 15 U.S.C. § 18a(c) (2006) (exempting certain transactions from HSR Act's requirements).

³⁸ See Fed. Trade Comm'n, Bureau of Competition, *Hart-Scott-Rodino Premerger Introductory Guide II, To File or Not to File—When you must file a premerger notification report form*, available at <http://www.ftc.gov/bc/hsr/introguides/guide2.pdf>. Guide II describes the criteria used to determine whether a transaction is subject to the requirements of the HSR Act.

³⁹ *Guide I*, *supra* note 36, at 11.

⁴⁰ *Id.* at 12.

⁴¹ See *id.* at 9. In the case of an all cash tender offer or an acquisition in bankruptcy, there is a fifteen-day waiting period. *Id.*

⁴² *Id.* at 9.

⁴³ Sher, *supra* note 18, at 41.

⁴⁴ See *id.* at 42 (describing how since 2001, the FTC has challenged consummated mergers involving MSC Software, Chicago Bridge, Airgas, and Aspen Technology, as well as seriously investigated dozens more); see also ABA Comments re Differential Standards, *supra* note 14, at app. a 2-7 (list of mergers and acquisitions the FTC has challenged post-closing); D. Bruce Hoffman & M. Sean Royall, *Administrative Litigation at the FTC: Past, Present, and Future*, 71 ANTITRUST L.J. 319, 319-20 (2003) ("The FTC today is aggressively continuing to use the administrative litigation process in the manner envisioned by the agency's creators . . . the FTC's administrative litigation process has become the forum in which many of our day's most complex and interesting antitrust issues are being litigated.").

3. The Federal Trade Commission Act

The FTC is an administrative agency created by Congress in 1914 under the Federal Trade Commission Act (“FTC Act”).⁴⁵ Only the FTC can bring cases under the FTC Act.⁴⁶ The FTC was formed to police “unfair methods of competition”⁴⁷ and “unfair or deceptive acts or practices in or affecting commerce.”⁴⁸ Congress created the FTC to supplement the DOJ’s enforcement of the antitrust laws, and to help develop and clarify antitrust policy by giving the FTC adjudicative power under Section 5(b) of the FTC Act.⁴⁹ As a result, the FTC can challenge a transaction in federal courts as well as through an internal administrative proceeding (known as a Part III proceeding) before an administrative law judge.⁵⁰ Whether it wins or loses at the federal court level, the FTC can still challenge a transaction through administrative litigation. This allows the FTC to initiate an administrative proceeding to challenge a transaction pre-consummation or post-consummation.⁵¹

B. Overview of the Merger Review Process

Although the DOJ’s Antitrust Division and the FTC have shared responsibility for enforcement of the federal antitrust laws, in practice, only one agency is responsible for

⁴⁵ See 15 U.S.C. § 41 (2006). Fed. Trade Comm’n, Bureau of Competition, Guide to the Antitrust Laws, The Antitrust Laws, http://www.ftc.gov/bc/antitrust/antitrust_laws.shtm.

⁴⁶ See *id.* Unlike the DOJ, the FTC does not have criminal enforcement authority.

⁴⁷ The term “unfair methods of competition” is generally thought to mean the same as the prohibitions of the Sherman and Clayton Acts. ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 129.

⁴⁸ 15 U.S.C. § 45(a)(2) (2006) (“The Commission is hereby empowered and directed to prevent persons, partnerships, or corporations . . . from using unfair methods of competition in or affecting commerce and unfair or deceptive acts or practices in or affecting commerce.”).

⁴⁹ Hoffman & Royall, *supra* note 44, at 319-20. See American Bar Association, Section of Antitrust Law, Public Comments Submitted to AMC Regarding Dual Federal Merger Enforcement, at 2 (Oct. 28, 2005) (“[T]he FTC was designed to function as an expert body in antitrust law, capable of assessing and adjudicating the competitive effects of complex transactions”) [hereinafter ABA Comments re Dual Enforcement]; Fed. Trade Comm’n, Guide to the Federal Trade Commission, <http://www.ftc.gov/bcp/edu/pubs/consumer/general/gen03.shtm> (“Congress created the FTC as a source of expertise and information on the economy.”).

⁵⁰ The FTC may seek a preliminary permanent injunction in federal court under 15 U.S.C. § 53(b) (2006). It can commence an administrative proceeding under 15 U.S.C. § 45(b)-(c) (2006).

⁵¹ See generally Hoffman & Royall, *supra* note 44.

investigating a particular merger. As the law enforcement agency of the executive branch, the DOJ is entrusted with the power to bring criminal antitrust cases or civil actions seeking an injunction and to take steps to remedy past violations.⁵² As an administrative agency, the FTC is allowed to seek a preliminary or permanent injunction in federal court or commence an internal administrative proceeding.⁵³ Despite criticism that this dual enforcement arrangement was unnecessarily duplicative, it has worked relatively well with few conflicts between the two agencies.⁵⁴ One of the reasons for the lack of clashes is that over the years, the two agencies have developed a “clearance process” where the FTC and the DOJ will consult with each other, and the matter is “cleared” to one agency for review.⁵⁵

1. The Clearance Process

To avoid duplication of enforcement efforts, the DOJ and the FTC will consult with each other to decide which agency will conduct a formal investigation of a particular transaction. During the waiting period, the FTC and the DOJ will assign the filing to a specific division or section within the agency having expertise over the industry of the proposed transaction.⁵⁶ Initially, both agencies will perform a preliminary review of the proposed transaction.⁵⁷ If the assigned division or section within one agency determines that a formal investigation is necessary, that agency will seek clearance from the other agency to conduct an investigation.⁵⁸ Since only one agency will be conducting the investigation of the proposed

⁵² Dep’t of Justice, Antitrust Div., Mission of the Antitrust Division, <http://www.justice.gov/atr/about/mission.htm>.

⁵³ See *supra* note 50.

⁵⁴ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 129. According to the Antitrust Modernization Report, “[c]ritics contend that having two agencies enforce the federal antitrust laws entails unnecessary duplication and can result in inconsistent antitrust policies, additional burdens on businesses, or other obstacles to efficient and fair federal antitrust enforcement.” *Id.*

⁵⁵ *Id.* at 132-33; *Guide I*, *supra* note 36, at 11.

⁵⁶ AMERICAN BAR ASSOCIATION, SECTION OF ANTITRUST LAW, THE MERGER REVIEW PROCESS: A STEP-BY-STEP GUIDE TO FEDERAL MERGER REVIEW 26 (3d. ed. 2006) [hereinafter MERGER REVIEW GUIDE].

⁵⁷ *Guide I*, *supra* note 36, at 11.

⁵⁸ MERGER REVIEW GUIDE, *supra* note 56, at 26-27, 134-36. At the DOJ, the Office of Operations will ask the Premerger Office at the FTC for clearance to investigate. At the FTC, the Premerger Office will notify the Office of Operations at the DOJ to coordinate which agency will conduct the investigation. *Id.* at 27.

transaction, neither agency will contact the parties or third parties until it has been decided which agency will be responsible for investigating the proposed transaction.⁵⁹ This minimizes the potential for confusion and duplication of efforts if both agencies contacted the parties at different times for the same matter.⁶⁰

This clearance process determines which agency will conduct the investigation; this is usually the agency with the most relevant staff expertise and experience in the industry potentially affected by the proposed merger.⁶¹ For example, the FTC is responsible for industries where consumer spending is high, such as health care, pharmaceuticals, food, energy, computer technology, and internet services.⁶² If there are disputes over which agency has more expertise in a given area, the matter is passed to increasingly senior staff until it is resolved, potentially all the way up to the Chairman of the FTC and the Assistant Attorney General for Antitrust at the DOJ.⁶³ As a result of the clearance process, only one agency takes control of the investigation.

⁵⁹ *Guide I*, *supra* note 36, at 11; MERGER REVIEW GUIDE, *supra* note 56, at 136. However, any interested person, including the parties to the proposed transaction, is free to present information to either or both agencies at any time. *Guide I*, *supra* note 36, at 11.

⁶⁰ See *Guide I*, *supra* note 36, at 11.

⁶¹ See MERGER REVIEW GUIDE, *supra* note 56, at 27. In 2002, the DOJ and the FTC reached an accord to explicitly allocate certain industries to each agency. See Dep't of Justice, Antitrust Div. & Federal Trade Comm'n, Memorandum of Agreement Between the Federal Trade Commission and the Antitrust Division of the United States Department of Justice Concerning Clearance Procedures for Investigations (Mar. 5, 2002), available at <http://www.justice.gov/atr/public/10170.pdf>. However, the accord was short-lived and ended after objections from Senator Ernest F. Hollings. As a result, the two agencies have continued to decide based on staff expertise and experience. See Lauren Kearney Peay, *The Cautionary Tale of the Failed 2002 FTC/DOJ Merger Clearance Accord*, 60 VAND. L. REV. 1307, 1308-10 (2007) (discussing the failure of the 2002 accord and potential approaches to improving the interaction between the FTC, the DOJ, and Congress).

⁶² Fed. Trade Comm'n, Bureau of Competition, An FTC Guide to the Antitrust Laws: The Enforcers, <http://www.ftc.gov/bc/antitrust/enforcers.shtm>. However, if one agency decides not to initiate an investigation, even in an industry where it has quite an amount of expertise in, the other agency is free to start an investigation. See MERGER REVIEW GUIDE, *supra* note 56, at 135.

⁶³ See ABA Comments re Dual Enforcement, *supra* note 49, at 11 & n.16; MERGER REVIEW GUIDE, *supra* note 56, at 135. Although not common, these disputes between the agencies may cause significant delays in the merger review process. *Id.* at 136.

2. Further Investigation Required: Second Request

Once clearance is granted, the investigating agency will notify the merging parties that an investigation has been opened. The investigating agency can now obtain information from various sources, including the merging parties.⁶⁴ After the initial investigation, the agency can decide to do three things: it can grant early termination of the waiting period, allow the waiting period to expire, or it can issue a Request for Additional Information (a “second request”).⁶⁵ The second request is commonly used to allow the staff more time to investigate,⁶⁶ and will often require the parties to provide more information about the transaction and its potential anticompetitive effects.⁶⁷ After the parties have substantially complied with the second request for information, there is an additional thirty-day waiting period, after which the agency must decide whether to approve the merger, seek a preliminary injunction in federal court to stop the merger, or seek a voluntary agreement not to close the deal until further investigation can be completed.⁶⁸ During this time, the parties can meet with review officials to argue that their transaction should not be challenged.⁶⁹ The investigating agency can also grant an early termination of the waiting period, or allow it to expire if they decline to pursue a challenge.⁷⁰ Either way, the parties are free to close their transaction at that point.

3. Agency Action: Approve or Litigate

At the DOJ, the staff’s recommendation is first reviewed by the appropriate section chiefs and increasingly senior officials before it goes to the ultimate decision maker, the

⁶⁴ See *Guide I*, *supra* note 36, at 12.

⁶⁵ See MERGER REVIEW GUIDE, *supra* note 56, at 138.

⁶⁶ See *id.* at 27-29. The extended waiting period is normally thirty days from the date of substantial compliance by both merging parties. It is ten days for “a cash tender offer or certain bankruptcy filings.” *Guide I*, *supra* note 36, at 13.

⁶⁷ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 151; see also *Guide I*, *supra* note 36, at 12.

⁶⁸ MERGER REVIEW GUIDE, *supra* note 56, at 29; see also 15 U.S.C. § 18a(e)(2) (2006). Parties are often willing to extend the time period voluntarily because it gives them more time to prepare and to meet with reviewing officials to persuade them not to challenge the transaction. MERGER REVIEW GUIDE, *supra* note 56, at 247-50.

⁶⁹ MERGER REVIEW GUIDE, *supra* note 56, at 29.

⁷⁰ 15 U.S.C. § 18a(b)(2) (2006); see also *Guide I*, *supra* note 36, at 10.

Assistant Attorney General for Antitrust.⁷¹ At the FTC, the staff recommendation is forwarded to the appropriate deputy directors and directors before it goes to the final decision makers, the five commissioners.⁷² If the merger is approved, or if thirty days has passed since the parties substantially complied with the second request, the parties are free to consummate their transaction.⁷³ If the investigating agency determines that a merger may substantially lessen competition, the agency can try to reach a settlement, or it can seek a preliminary and permanent injunction in the appropriate district court to enjoin the consummation of the merger.⁷⁴

In practice, the agencies will usually try to negotiate with the merging parties to reach a settlement, either through a consent decree where the parties agree to a divestiture of certain assets to ease concerns about the merger's anticompetitive effects, or through a less common "fix-it-first" restructuring of their transaction.⁷⁵ Depending on the circumstances, parties can either abandon the transaction,⁷⁶ or agree to settle as a way to avoid costly and time-intensive litigation that could delay the closing of the transaction and the ensuing efficiencies of the merger.⁷⁷

C. *Proceeding to Litigation: Challenging a Proposed Transaction*

1. Seeking Injunctive Relief in Federal Court

If the merging parties and the enforcement agency fail to negotiate a settlement, the FTC and the DOJ are authorized

⁷¹ MERGER REVIEW GUIDE, *supra* note 56, at 29. For a more in depth discussion of the process and the decision makers that are involved, see *id.* at 232-36, 242-47.

⁷² *Id.* at 29-30. Parties are given the opportunity to present their case at each step of the approval process. This includes meeting with each commissioner separately. A majority vote of the commissioners is necessary for any action. For a more in depth discussion of the process and the decision makers that are involved, see *id.* at 232-42.

⁷³ *Guide I*, *supra* note 36, at 13.

⁷⁴ MERGER REVIEW GUIDE, *supra* note 56, at 30-31, 254.

⁷⁵ See *id.* at 252-53; DEP'T OF JUSTICE, ANTITRUST DIV., ANTITRUST DIVISION POLICY GUIDE TO MERGER REMEDIES 1 (2004) [hereinafter DOJ MERGER REMEDIES GUIDE], available at <http://www.usdoj.gov/atr/public/guidelines/205108.pdf>; see also Lawrence M. Frankel, *The Flawed Institutional Design of U.S. Merger Review: Stacking the Deck Against Enforcement*, 2008 UTAH L. REV. 159, 181 (2008).

⁷⁶ See MERGER REVIEW GUIDE, *supra* note 56, at 256-57.

⁷⁷ See *id.* at 255.

to seek injunctive relief in federal court to enjoin a transaction that they believe raises competitive concerns.⁷⁸ The DOJ and the FTC have different approaches when seeking injunctive relief. Unlike the FTC, the DOJ does not have another avenue for permanent relief other than the federal court process. As a result, the DOJ often asks for both a preliminary injunction and a permanent injunction from the district courts.⁷⁹ If the DOJ's request is denied, the parties can usually consummate their merger without further concerns of antitrust litigation.⁸⁰ In contrast, the FTC only seeks a preliminary injunction; if the FTC loses, the parties still have to worry about the FTC potentially pursuing costly and lengthy administrative litigation.⁸¹

Due to the need to close a proposed transaction quickly (to enjoy the efficiencies that come from a merger and to avoid the costs of litigation), preliminary injunctions are particularly important to both the parties and the enforcement agencies.⁸² If the district court denies the injunction, the agencies normally treat the denial as final and will not take any further action.⁸³ As a result, the parties can close the merger relatively quickly.⁸⁴ However, if a court grants the injunction, the parties will most likely abandon the transaction because very few firms can withstand the time, costs, and uncertainty involved in an appeal or an administrative trial.⁸⁵

⁷⁸ The DOJ is authorized to seek an injunction under 15 U.S.C. § 25 (2006). The FTC is authorized to seek an injunction under 15 U.S.C. § 53(b) (2006).

⁷⁹ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 138.

⁸⁰ The ABA Section of Antitrust Law has not been able to find any example of the DOJ seeking a permanent injunction after failing to obtain a preliminary injunction. ABA Comments re Differential Standards, *supra* note 14, at 5.

⁸¹ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 139.

⁸² See MERGER REVIEW GUIDE, *supra* note 56, at 255 ("In many cases, the preliminary injunction motion will determine the outcome of the case.").

⁸³ UNDERSTANDING THE ANTITRUST ISSUES, *supra* note 4, at 546 ("An unsuccessful effort to obtain a preliminary injunction can be the plaintiff's final battle to block a merger. . .").

⁸⁴ See MERGER REVIEW GUIDE, *supra* note 56, at 255; UNDERSTANDING THE ANTITRUST ISSUES, *supra* note 4, at 546-47. This is assuming the Court of Appeals fails to grant a stay pending appeal by the enforcement agency or the FTC decides not to pursue an administrative proceeding. See *infra* Part I.C.2.

⁸⁵ UNDERSTANDING THE ANTITRUST ISSUES, *supra* note 4, at 547. ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 139.

2. The FTC Pursuing Administrative Litigation

If the DOJ fails to obtain an injunction, it will abandon any further litigation.⁸⁶ However, the FTC has pursued administrative proceedings after losing at the preliminary injunction stage.⁸⁷ The decision by the FTC to pursue an administrative proceeding is made on a case-by-case basis with the standard being whether the “pursuit of administrative litigation after the denial of a preliminary injunction motion would serve the public interest.”⁸⁸ Some of the criteria the FTC uses in its decision include “the district court’s factual findings and conclusions of law; any new evidence developed during the preliminary injunction proceeding; whether the transaction raises important issues of fact, law, or merger injunction policy that need resolution in administrative litigation; the costs and benefits of further proceedings; and any additional relevant factor.”⁸⁹ An administrative proceeding takes place before an FTC administrative law judge, with review by the five commissioners.⁹⁰ The decision can then be appealed to a federal appellate court.⁹¹

3. Relief: Structural Remedies and Conduct Remedies

Merger concerns can be resolved through negotiation, resulting in a settlement, or through litigation in court. The FTC has stated that its remedial objective is to “prevent the anticompetitive effects likely to result from a merger that the [FTC] has determined is unlawful.”⁹² Similarly, according to the DOJ, “[a]lthough the remedy should always be sufficient to redress the antitrust violation, the purpose of a remedy is not to enhance premerger competition but to restore it.”⁹³ Coming up with an appropriate remedy can be extremely difficult. In

⁸⁶ MERGER REVIEW GUIDE, *supra* note 56, at 255.

⁸⁷ *Id.*

⁸⁸ Press Release, Fed. Trade Comm’n, FTC Closes Its Investigation of Arch Coal’s Acquisition of Triton Coal Company’s North Rochelle Mine (June 13, 2005) (internal quotation marks omitted), *available at* <http://www.ftc.gov/opa/2005/06/archcoal.htm>.

⁸⁹ *Id.*

⁹⁰ Hoffman & Royall, *supra* note 44, at 322.

⁹¹ *Id.*

⁹² Federal Trade Comm’n, Negotiating Merger Remedies, Statement of the Bureau of Competition of the Federal Trade Commission, at 4 (April 2, 2003), *available at* <http://www.ftc.gov/bc/bestpractices/bestpractices030401.pdf>.

⁹³ DOJ MERGER REMEDIES GUIDE, *supra* note 75, at 4.

fashioning a remedy, the speed, certainty, cost, efficacy, and ease of monitoring are all important factors that need to be taken into consideration.⁹⁴

Merger remedies usually take two basic forms.⁹⁵ The first form, a structural remedy, usually involves a divestiture or the sale of assets by the merged firms.⁹⁶ The second form, a conduct remedy, is usually an injunctive provision that regulates or changes the business conduct of the merged firm.⁹⁷ Structural remedies are preferred because they require very little ongoing monitoring by the enforcement agency.⁹⁸ In contrast, conduct remedies are not preferred because of the monitoring costs involved, and the fact that consumers would ultimately be harmed if the restrained firm fails to survive in a competitive market.⁹⁹ A remedy can be a combination of structural and injunctive remedies.

Divestiture is the primary post-consummation remedy for a Section 7 violation (or Section 5 of the FTC Act) because the logical solution to excessive concentration is divestiture of the assets that caused the antitrust problems.¹⁰⁰ Since the goal of divestiture is to restore competition, the agencies try to ensure that the divestiture remedy contains enough assets for the purchaser to function as a long-term viable competitor, with the hope of replacing the competition prior to the merger.¹⁰¹

II. *FTC v. WHOLE FOODS MARKET, INC.*

The tortuous legal battle between the FTC and Whole Foods might be over, but the D.C. Circuit's controversial decision in *Whole Foods* continues to give the FTC a significant advantage over the DOJ in challenging a merger and extracting a settlement. This Part examines the D.C. Circuit

⁹⁴ *Id.* at 7-8.

⁹⁵ *Id.* at 7.

⁹⁶ *Id.*

⁹⁷ *Id.*

⁹⁸ *Id.* "Structural remedies are preferred to conduct remedies in merger cases because they are relatively clean and certain, and generally avoid costly government entanglement in the market." *Id.* This is in contrast to conduct remedies, which are often "more difficult to craft, more cumbersome and costly to administer, and easier than a structural remedy to circumvent." *Id.* at 8.

⁹⁹ *Id.* at 8-9.

¹⁰⁰ See UNDERSTANDING THE ANTITRUST ISSUES, *supra* note 4, at 603; see also DOJ MERGER REMEDIES GUIDE, *supra* note 75, at 7-8.

¹⁰¹ DOJ MERGER REMEDIES GUIDE, *supra* note 75, at 9-11.

Court's fractured opinion. It begins with a discussion of the merger between Whole Foods and Wild Oats, and the FTC's decision to challenge the merger. It then reviews the district court and the D.C. Circuit's opinions, and concludes with a review of the settlement between the FTC and Whole Foods.

A. *Background: The Merger Between Whole Foods and Wild Oats*

At the time of the merger, Whole Foods and Wild Oats, respectively, were the largest and second largest nationwide operators of organic supermarkets in the United States.¹⁰² Whole Foods operated approximately 194 stores,¹⁰³ and Wild Oats operated approximately 110 stores.¹⁰⁴ Whole Foods is a Texas corporation that opened its first store in 1980.¹⁰⁵ Wild Oats is a Delaware corporation that opened its first store in 1987.¹⁰⁶ Both chains have expanded over the years by opening new stores and acquiring other premium natural and organic supermarkets.¹⁰⁷ Both Whole Foods and Wild Oats tried to differentiate themselves from other supermarkets by focusing on natural and organic products, as well as a commitment to quality and service.¹⁰⁸ In February 2007, Whole Foods announced its intent to purchase Wild Oats for an estimated \$565 million.¹⁰⁹ The market reacted positively after the announcement as investors and analysts generally applauded the merger as necessary in the face of intense competition from

¹⁰² Martin, *supra* note 1.

¹⁰³ *Id.*

¹⁰⁴ *Id.*

¹⁰⁵ Whole Foods Market Home Page, About Whole Foods Market, <http://www.wholefoodsmarket.com/company/index.php> (last visited Feb. 27, 2010).

¹⁰⁶ Whole Foods Market Home Page, Company History, <http://www.wholefoodsmarket.com/company/history.php#18> (last visited Feb. 27, 2010).

¹⁰⁷ *FTC v. Whole Foods Mkt., Inc.*, 502 F. Supp. 2d 1, 10-11 (D.D.C. 2007), *rev'd*, 533 F.3d 869 (D.C. Cir. 2008). Whole Foods acknowledges on its website that “[m]uch of the growth of [the] company has been accomplished through mergers and acquisitions. The story of Whole Foods is incomplete without honoring these notable companies in their own right.” Whole Foods Market Home Page, Company History, <http://www.wholefoodsmarket.com/company/history.php#18> (last visited Feb. 27, 2010).

¹⁰⁸ Proof Brief for Appellant FTC at 5-6, *FTC v. Whole Foods Mkt., Inc.*, 548 F.3d 1028 (D.C. Cir. 2008) (No. 07-5276).

¹⁰⁹ Martin, *supra* note 1. The deal was a tender offer for all of Wild Oats stock at a price of \$18.50 per share—a 23% premium over the average share price in January of 2007. Proof Brief for Appellant FTC, *supra* note 108, at 6.

larger rivals like Wal-Mart, Safeway, Kroger, and Trader Joe's.¹¹⁰

In February 2007, Whole Foods filed the Premerger Notification and Report Forms with the FTC and DOJ as required by the HSR Act.¹¹¹ The merger caught the FTC's attention.¹¹² After going through the clearance process, the FTC was chosen as the investigating agency due to its traditional expertise in the supermarkets industry.¹¹³ After reviewing the documents from the second request, the FTC authorized its staff to seek a preliminary injunction under Section 7 of the Clayton Act and Section 5 of the FTC Act.¹¹⁴ In June 2007, the FTC filed a complaint in the District of Columbia seeking a preliminary injunction to enjoin the merger.¹¹⁵ All five commissioners voted in favor of bringing the case.¹¹⁶

In its complaint, the FTC argued that a merger of the two biggest chains in the premium natural and organic supermarkets, or PNOS, market would "substantially lessen competition and thereby cause significant harm to consumers" by increasing prices and reducing quality and services.¹¹⁷ In defining the relevant markets, the FTC found that premium natural and organic supermarkets are different from

¹¹⁰ Martin, *supra* note 1. One commentator described the merger as "consistent with how Whole Foods has created value for shareholders for much of its history. . . ." *Id.* The chief executive of Wild Oats, Gregory Mays, said he considered the merger a "perfect marriage" and a "natural fit" because of the intense competition from much larger rivals who were eager to move into this lucrative and growing market. *Id.* In fact, Mr. Mays stated that since "the two stores were the leaders in the natural and organic marketplace . . . it [was] a 'perfect marriage' because the combined company could focus on larger rivals." *Id.*

¹¹¹ See *supra* Part I.A.2.

¹¹² See Proof Brief for Appellant FTC, *supra* note 108, at 6.

¹¹³ See FED. TRADE COMM'N, AN FTC GUIDE TO THE ENFORCERS: THE FEDERAL GOV'T, STATES AND PRIVATE PARTIES (2008), http://www.ftc.gov/bc/antitrust/factsheets/FactSheet_FedEnforcers.pdf.

¹¹⁴ Press Release, Fed. Trade Comm'n, FTC Seeks to Block Whole Foods Market's Acquisition of Wild Oats Markets (June 5, 2007), <http://www.ftc.gov/opa/2007/06/wholefoods.shtm>.

¹¹⁵ *Id.*; Complaint for Temporary Restraining Order and Preliminary Injunction Pursuant to Section 13(b) of the Federal Trade Commission Act at 2, 5-6, FTC v. Whole Foods Mkt. Inc., 502 F. Supp. 2d 1 (D.D.C. 2007) (No. 07-cv-01021) [hereinafter FTC Complaint].

¹¹⁶ Press Release, Fed. Trade Comm'n, *supra* note 114.

¹¹⁷ FTC Complaint, *supra* note 115, at 1. See Andrew Martin, *F.T.C. to Sue in Bid to Halt Food Merger*, N.Y. TIMES, Jun. 6, 2007, at C1 (Jeffrey Schmidt, the director of the FTC's Bureau of Competition, said in a statement that "Whole Foods and Wild Oats are each other's closest competitors in premium natural and organic supermarkets, and are engaged in intense head-to-head competition in markets across the country. If Whole Foods is allowed to devour Wild Oats, it will mean higher prices, reduced quality and fewer choices for consumers.").

conventional retail supermarkets because PNOS offer a unique upscale shopping experience for their customer that is characterized by a large selection of organic foods and excellent customer service.¹¹⁸ The FTC alleged that the customer base for PNOS is different from that of traditional supermarkets because PNOS customers seek an experience where the shopping environment can matter as much as the price.¹¹⁹ The FTC also alleged that Whole Foods and Wild Oats were each other's closest competitors in twenty one geographic markets and that the merger would create monopolies in eighteen cities.¹²⁰

Needless to say, the parties involved as well as analysts who follow the companies and industry were surprised by the FTC's decision.¹²¹ Analysts and reporters were quick to point out that the combined entity would only operate about 300 supermarkets.¹²² By comparison, Wal-Mart, the largest supermarket chain in the U.S., owns about 3,000 stores that sell groceries, and Kroger, the second-largest supermarket chain in the U.S., owns about 2,500 grocery stores.¹²³ Due to the need to close the merger quickly, the lawsuit at the district court level was litigated on a very fast track so as to allow the losing side sufficient time to appeal the decision before the consummation of the proposed deal, which was scheduled for August 31, 2007.¹²⁴

¹¹⁸ FTC Complaint, *supra* note 115, at 10.

¹¹⁹ *See id.* at 8-9.

¹²⁰ *Id.* at 11-12.

¹²¹ *See* Martin, *supra* note 117. One "somewhat bemused" research analyst remarked that the FTC's decision was "somewhat at odds' with the recent blurring of lines between stores like Whole Foods and Trader Joe's and more conventional chains like Publix and Wegmans" and the fact that "74 percent of natural and organic foods were now sold through mass-market channels like conventional supermarkets." *Id.* Whole Foods' Chief Executive John Mackey said in a statement that "[t]he FTC has failed to recognize the robust competition in the supermarket industry, which has grown more intense as competitors increase their offerings of natural, organic and fresh products; renovate their stores; and open stores with new banners and formats resembling Whole Foods Market." *Id.*

¹²² David Kesmodel & John R. Wilke, *Why Whole Foods Deal Is in Peril—Pending FTC Challenge To Wild Oats Deal Argues Firms Are in Narrow Arena*, WALL ST. J., Jun. 6, 2007, at A3. Whole Foods and Wild Oats together only accounted for 15% of the \$46 billion natural-foods market. *Id.*

¹²³ *Id.*

¹²⁴ *FTC v. Whole Foods Mkt., Inc.*, 502 F. Supp. 2d 1, 4 (D.D.C. 2007), *rev'd*, 533 F.3d 869 (D.C. Cir. 2008). The merger was consummated on August 28, 2007.

B. *The District Court's Opinion in FTC v. Whole Foods Market, Inc.*

In a thorough opinion, Judge Friedman denied the FTC's request for a preliminary injunction because he concluded the FTC had not demonstrated a likelihood of success on the merits—that is, that the effects of the merger “may substantially lessen competition [or] tend to create a monopoly” in a properly defined relevant product market.¹²⁵ As with most antitrust cases, the product market definition was key.¹²⁶ After going over the arguments on both sides, Judge Friedman found that the “economic evidence, market research studies, and evidence concerning the realities on the ground . . . all lead to the conclusion that the relevant product market in this case is not [PNOS] as argued by the FTC but . . . at least all supermarkets.”¹²⁷ Judge Friedman also noted that so-called conventional supermarkets like Wal-Mart, Kroger, and Safeway were all carrying natural and organic foods.¹²⁸ In fact, market research indicated that a majority of natural and organic goods are now being sold in conventional supermarkets as they move aggressively into the sale of organic foods.¹²⁹ With such stiff competition from more conventional supermarkets, Judge Friedman believed that post-merger, customers would still have plenty of competing options to choose from.¹³⁰ As a result, Judge Friedman concluded that there was no substantial likelihood that the FTC would be able to prove its asserted product market, or that the Whole Foods-Wild Oats merger would “substantially lessen competition or tend to create a monopoly.”¹³¹

Following the district court's decision, the FTC filed an emergency motion for an injunction pending the outcome of the appeal, which was unanimously denied by a three-judge panel of the D.C. Circuit Court.¹³² At that point, four federal judges had looked at the case and all concluded that the FTC had failed to meet the preliminary injunction standard. Shortly

¹²⁵ *Id.* (quoting 15 U.S.C. §§ 18, 53(b) (2006)).

¹²⁶ *FTC v. Staples, Inc.*, 970 F. Supp. 1066, 1073 (D.D.C. 1997).

¹²⁷ *Whole Foods*, 502 F. Supp. 2d at 34.

¹²⁸ *Id.* at 26-27.

¹²⁹ *Id.* at 27.

¹³⁰ *Id.* at 36.

¹³¹ *Id.* at 49-50.

¹³² *FTC v. Whole Foods Mkt., Inc.*, 548 F.3d 1028, 1033 (D.C. Cir. 2008), *amended and reissued*, 592 F. Supp. 2d 107 (D.D.C. 2009).

thereafter, the parties closed the merger.¹³³ After closing, Whole Foods started integrating Wild Oats by converting certain stores and selling other stores under the Wild Oats family.¹³⁴

C. *The D.C. Circuit's Opinions in FTC v. Whole Foods Market, Inc.*

In July 2008, almost a full year after the merger was consummated, a panel of the D.C. Circuit reversed and remanded the district court's decision to deny the FTC's motion for a preliminary injunction in a splintered decision with no majority opinion.¹³⁵ The decision was amended and reissued in November 2008.¹³⁶ Judge Brown, Judge Tatel, and Judge Kavanaugh each wrote a separate opinion in the July and November rulings.¹³⁷ In the July ruling, Judge Brown wrote the opinion for the court. Judge Tatel wrote a concurring opinion, and Judge Kavanaugh wrote a dissenting opinion.¹³⁸ According to Judge Brown, the district court committed legal error by rejecting the FTC's market definition so that it failed to give adequate weight to the FTC's evidence.¹³⁹ The majority thus held that the FTC had raised enough questions about the merits of its case against the merger.¹⁴⁰ Following the decision, Whole Foods petitioned the D.C. Circuit for a rehearing en banc. The petition for rehearing en banc was denied on November 21, 2008.¹⁴¹ However, on that same day, the three-judge panel amended and reissued its original opinions.¹⁴² The most significant difference between the July and November rulings was that Judge Tatel no longer concurred in Judge Brown's opinion but only in the judgment of the court.¹⁴³ Judge Brown and Judge Tatel continued to agree, however, that the

¹³³ Proof Brief for Appellant FTC, *supra* note 108, at 4 n.3.

¹³⁴ *Whole Foods*, 548 F.3d at 1033 ("Whole Foods has already closed some Wild Oats stores and sold others. In addition, Whole Foods has sold two complete lines of stores, Sun Harvest and Harvey's, as well as some unspecified distribution facilities.").

¹³⁵ *FTC v. Whole Foods Mkt., Inc.*, 533 F.3d 869, 869 (D.C. Cir. 2008).

¹³⁶ *Whole Foods*, 548 F.3d at 1028.

¹³⁷ *Id.*

¹³⁸ *See id.*

¹³⁹ *Id.* at 873.

¹⁴⁰ *Id.* at 882.

¹⁴¹ *Whole Foods*, 548 F.3d at 1028.

¹⁴² *Id.*

¹⁴³ *Id.*

district court's decision should be reversed and remanded for further proceedings.¹⁴⁴

Judge Brown believed that the district court used the correct standard for granting a preliminary injunction, but incorrectly applied the standard in its analysis of the product market.¹⁴⁵ According to Judge Brown, in deciding whether to grant an injunction, "a district court must balance the likelihood of the FTC's success against the equities, under a sliding scale."¹⁴⁶ However, this balancing test will often weigh in favor of the FTC because "the public interest in effective enforcement of the antitrust laws' was Congress's specific 'public equity consideration' in enacting the provision."¹⁴⁷ Thus, the FTC will usually be able to obtain a preliminary injunction by "rais[ing] questions going to the merits so serious, substantial, difficult[,] and doubtful as to make them fair ground for thorough investigation."¹⁴⁸

According to Judge Brown, the district court did not appropriately apply the standard because it incorrectly found that the FTC failed to present evidence of a likelihood of success and therefore never weighed the equities.¹⁴⁹ The district court erred when it assumed that "marginal customers,"¹⁵⁰ and not "core customers,"¹⁵¹ must be the focus of an antitrust analysis.¹⁵² Instead, Judge Brown stated that core consumers should be given consideration as a separate submarket in certain cases, such as when there is a distinct service or a specialized or "unique environment."¹⁵³ Judge Brown believed that the FTC's evidence demonstrated that there was a distinct PNOS submarket of core customers who shop exclusively at Whole Foods or Wild Oats for their unique environment.¹⁵⁴ As a result, the district court had underestimated the FTC's

¹⁴⁴ *Id.*

¹⁴⁵ *Id.* at 1034-36.

¹⁴⁶ *Id.* at 1035.

¹⁴⁷ *Id.* (quoting *FTC v. H.J. Heinz Co.*, 246 F.3d 708, 726 (D.C. Cir. 2001)).

¹⁴⁸ *Id.* (quoting *Heinz*, 246 F.3d at 714-15).

¹⁴⁹ *Id.* at 1035-36.

¹⁵⁰ A marginal consumer is someone who would switch to a competitor if his primary choice imposed a small but significant and nontransitory price increase (typically 5%). See *FTC v. Whole Foods Mkt., Inc.*, 502 F. Supp. 2d 1, 17 (D.D.C. 2007), *rev'd*, 533 F.3d 869 (D.C. Cir. 2008).

¹⁵¹ Core customers are the customers who refuse to switch despite a price increase. *Id.* at 16-17.

¹⁵² *Whole Foods*, 548 F.3d at 1037.

¹⁵³ *Id.* at 1037-39.

¹⁵⁴ *Id.* at 1039-40.

likelihood of success on the merits.¹⁵⁵ Since the district court did not reach the equities in its decision, Judge Brown and Judge Tatel both agreed to remand the case back to the district court to determine whether policy considerations weighed against the injunction.¹⁵⁶

After the D.C. Circuit's decision, the FTC made it clear that it wanted to commence an in-house administrative trial on the merger, scheduled to begin in April of 2009.¹⁵⁷ Thus, the FTC had two cases on parallel tracks: one in the district court (*Whole Foods*), and the other in an internal administrative proceeding. On January 29, 2009, the FTC announced that it would temporarily halt its review of the merger so that the FTC and Whole Foods could engage in settlement talks.¹⁵⁸ On March 6, 2009, almost 21 months after the FTC first sued Whole Foods in federal court to stop the deal, Whole Foods and the FTC announced a settlement.¹⁵⁹

D. *The Settlement Between the FTC and Whole Foods*

The consent agreement between Whole Foods and the FTC required Whole Foods to divest thirty-two Wild Oats stores and assets related to those stores in seventeen separate geographic markets.¹⁶⁰ However, out of the thirty-two stores, only thirteen stores were operating at the time of the agreement.¹⁶¹ Whole Foods had closed the other nineteen stores, but still retained control over them.¹⁶² Whole Foods was also required to divest Wild Oats intellectual property, including the rights to the Wild Oats brand.¹⁶³ The FTC believed that “[e]ven months after the acquisition, the Wild Oats brand name

¹⁵⁵ *Id.* at 1041.

¹⁵⁶ *Id.*

¹⁵⁷ See Brent Kendall, *FTC is Planning Hearings on Whole Foods Merger*, WALL ST. J., Aug. 12, 2008, at B8.

¹⁵⁸ Commission Order Withdrawing Matter From Adjudication, Fed. Trade Comm'n, *In the Matter of Whole Foods Market, Inc., and Wild Oats Markets, Inc.*, No. 9324 (2009), <http://www.ftc.gov/os/adjpro/d9324/090128orderwithdrawingmatter.pdf>.

¹⁵⁹ Press Release, Fed. Trade Comm'n, *supra* note 2.

¹⁶⁰ Analysis of Agreement Containing Consent Orders To Aid Public Comment at 3, *In the Matter of Whole Foods Market, Inc., and Wild Oats Markets, Inc.*, No. 9324 (2009), <http://www.ftc.gov/os/adjpro/d9324/090306wfanal.pdf> [hereinafter *Analysis of Agreement*].

¹⁶¹ *Id.* at 3. These are referred to as live stores. *Id.* at 3 n.4.

¹⁶² *Id.* at 3. These are referred to as dark stores. *Id.* at 3 n.4.

¹⁶³ Decision and Order at 2, *In the Matter of Whole Foods Market, Inc., and Wild Oats Markets, Inc.*, No. 9324 (2009), <http://www.ftc.gov/os/adjpro/d9324/090306wfdo.pdf> [hereinafter *Decision and Order*].

retains significant brand equity that has been developed over the past 20 years.”¹⁶⁴ While the stores may be divested to more than one FTC-approved buyer, the Wild Oats intellectual property may be divested to only a single FTC-approved buyer.¹⁶⁵ The consent agreement appointed a divestiture trustee to oversee the marketing and sale of the assets.¹⁶⁶ The consent agreement also includes an order to maintain assets, which requires Whole Foods to continue to operate the stores in a way that preserves marketability and competitiveness until a FTC-approved buyer is found.¹⁶⁷ In the end, the divestitures will only “offer relief in 17 of the 29 geographic markets alleged in the amended administrative complaint.”¹⁶⁸

III. THE CHOICE OF THE ENFORCEMENT AGENCY IS OUTCOME-DETERMINATIVE AFTER *WHOLE FOODS*

Due to the statutory authority granted to the FTC in the FTC Act, there are important procedural differences between the FTC and the DOJ. First, as seen in *Whole Foods*, the FTC enjoys a lower standard for obtaining a preliminary injunction than the DOJ.¹⁶⁹ Second, the FTC’s statutory authority to commence administrative litigation, even after a denial of a request for a preliminary injunction, creates uncertainty about the proposed transaction—a risk not faced if the DOJ is challenging the merger.¹⁷⁰ The divergence between the two agencies is most troubling when the FTC decides to pursue an administrative trial post-consummation, as seen in its battle with Whole Foods. With the status of their merger in legal limbo (and already well integrated), the parties are forced to defend their merger in a long and costly administrative proceeding, a risk they do not face if the DOJ is challenging it. This Part argues that the substantive outcome of a merger challenge depends on which agency is challenging it; that is, the choice of enforcement agency is outcome-determinative.¹⁷¹ The preliminary injunction standard for the FTC as articulated in *Whole Foods* puts the debate on whether the choice of

¹⁶⁴ Analysis of Agreement, *supra* note 160, at 3.

¹⁶⁵ Decision and Order, *supra* note 163, at 4.

¹⁶⁶ *Id.* at 3. The FTC appointed The Food Partners as the divestiture trustee.

¹⁶⁷ Analysis of Agreement, *supra* note 160, at 4.

¹⁶⁸ *Id.* at 3.

¹⁶⁹ See *infra* Part III.A.

¹⁷⁰ See *supra* Part I.C.2.

¹⁷¹ See ABA Comments re Differential Standards, *supra* note 14, at 9.

enforcement agency is outcome-determinative to rest with an empathetic “yes” in the D.C. Circuit.

A. *The Divergence in Preliminary Injunction Standards for the DOJ and the FTC*

The most significant aspect of the *Whole Foods* decision is not the substance of the decision, but rather the preliminary injunction standard for the FTC. There has been an ongoing debate about whether the FTC faces a preliminary injunction burden that is lower than that of the DOJ.¹⁷² In its report, the Antitrust Modernization Commission stated that “[t]here is at least a perception, if not a reality, that the FTC and the DOJ face different standards” and the standard for the FTC is “less burdensome, or is generally perceived to be less burdensome, than the standard applicable to DOJ actions” for obtaining a preliminary injunction.¹⁷³ After Judge Brown’s opinion in *Whole Foods*, the perception that the FTC faces a lower preliminary injunction standard is no longer just perception, it is a fact.

Traditionally, a plaintiff seeking injunctive relief must prove: (1) irreparable harm if the injunction is not granted, (2) the injury outweighs any harm to the defendant by the

¹⁷² See Report of the Section of Antitrust Law of the American Bar Association to the Antitrust Modernization Commission at 10 (2004), <http://govinfo.library.unt.edu/amc/comments/abaantitrustsec.pdf> (“DOJ has to meet the regular district court standards when seeking preliminary injunctive relief . . . subjecting itself to a full hearing on the merits and a higher standard of proof. In contrast, the FTC typically seeks only preliminary injunctive relief from the district court and does so under a standard that, as written, appears to be less demanding than that facing other litigants (including the DOJ), reserving trial on the merits for agency adjudication. Most transactions are abandoned if an injunction under any standard is granted. Thus, some lawyers believe that the apparently lower burden for the FTC could lead to different outcomes.”). *But see* Hearing on Federal Civil Remedies for Antitrust Offenses: Statement of Commissioner Thomas B. Leary Before the Antitrust Modernization Commission at 5 (2005), <http://www.ftc.gov/speeches/leary/051201civilremedies.pdf> (“The ABA submission points out, first, that some decisions seem to apply a more lenient standard when the FTC applies for a preliminary injunction than they do when the DOJ applies. It is not possible to know whether the facially different standards have been outcome-determinative; I personally doubt that they have been in recent years, and suspect our litigators would agree.”); Observations on Federal Antitrust Enforcement Institutions: Comments of W. Blumenthal to the Antitrust Modernization Commission at 6 (2005), <http://www.ftc.gov/os/2005/11/051103gcstmnntonfedantitrustenforcement.pdf> (“Because the preliminary injunction standards applied to actions brought by the FTC and DOJ appear to be substantially identical, any differences in their application would seem more likely to be based upon the specific facts of a given matter than substantive legal standards. So far as I am aware there is no evidence that any cases or group of cases were or would have been decided differently based on which of the antitrust agencies was the plaintiff.”).

¹⁷³ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 141-42.

injunction, (3) the plaintiff has a substantial likelihood of success on the merits, and (4) the injunction is in the public interest.¹⁷⁴ Due to its status as an administrative agency, the FTC is subjected to a different preliminary injunction standard than the DOJ. The DOJ is subjected to the traditional test articulated above,¹⁷⁵ while the FTC is subjected to the standard set forth in Section 13(b) of the FTC Act.¹⁷⁶ Under Section 13(b) of the FTC Act, the FTC can obtain an injunction “[u]pon a proper showing that, weighing the equities and considering the Commission’s likelihood of ultimate success, such action would be in the public interest.”¹⁷⁷ Courts have read Congress’s intent in Section 13(b) as making injunctive relief “broadly available to the FTC”¹⁷⁸ and the appropriate test to be a “public interest” test—that is, “the court evaluates whether it is in the public interest to enjoin the proposed merger.”¹⁷⁹

At the preliminary injunction stage, the FTC is not required to prove that the proposed merger would in fact violate Section 7 of the Clayton Act.¹⁸⁰ Instead, the FTC is only required to show that it is likely to succeed in showing under Section 7 that the proposed merger “may be substantially to lessen competition” or “tend to create a monopoly.”¹⁸¹ In *Whole Foods*, Judge Brown made it clear that the FTC will usually be able to obtain a preliminary injunction blocking a merger by “rais[ing] questions going to the merits so serious, substantial,

¹⁷⁴ DEP’T OF JUSTICE, ANTITRUST DIV., ANTITRUST DIVISION MANUAL IV-15 (4th ed. 2008), available at <http://www.justice.gov/atr/public/divisionmanual/atrdivman.pdf> [hereinafter ANTITRUST DIVISION MANUAL]; UNDERSTANDING THE ANTITRUST ISSUES, *supra* note 4, at 564.

¹⁷⁵ ANTITRUST DIVISION MANUAL, *supra* note 174 (“The Federal Rules do not prescribe a standard for granting or denying a PI. Traditional equitable considerations apply.”).

¹⁷⁶ See 15 U.S.C. § 53(b) (2006); see also *United States v. Gillette Co.*, 828 F. Supp. 78, 80 (D.D.C. 1993) (comparing the preliminary injunction standards for the DOJ and FTC).

¹⁷⁷ 15 U.S.C. § 53(b) (2006); see *FTC v. H.J. Heinz Co.*, 246 F.3d 708, 714 (D.C. Cir. 2001) (“Congress intended this standard to depart from what it regarded as the then-traditional equity standard.”).

¹⁷⁸ *FTC v. Exxon Corp.*, 636 F.2d 1336, 1343 (D.C. Cir. 1980) (“Congress further demonstrated its concern that injunctive relief be broadly available to the FTC by incorporating a unique ‘public interest’ standard in 15 U.S.C. § 53(b), rather than the more stringent, traditional ‘equity’ standard for injunctive relief.”).

¹⁷⁹ *H.J. Heinz*, 246 F.3d at 713.

¹⁸⁰ *E.g., id.* at 714; *FTC v. Libbey, Inc.*, 211 F. Supp. 2d 34, 44 (D.D.C. 2002) (“Congress used the words *may* be substantially to lessen competition . . . to indicate that its concern was with probabilities, not certainties.” (quoting *Brown Shoe Co. v. United States*, 370 U.S. 294, 323 (1962))).

¹⁸¹ *H.J. Heinz*, 246 F.3d at 714; *Libbey*, 211 F. Supp. 2d at 44; *FTC v. Staples, Inc.*, 970 F. Supp. 1066, 1071 (D.D.C. 1997).

difficult[,] and doubtful as to make them fair ground for thorough investigation.”¹⁸² Despite “at best, poorly explained evidence” on the FTC’s part,¹⁸³ the FTC’s statutory authority to engage in adjudicative administrative proceedings means that it can create a presumption in favor of an injunction just by raising serious and doubtful questions about the merits of the case.¹⁸⁴ In other words, the FTC is entitled to an injunction unless the FTC has “entirely failed to show a likelihood of success.”¹⁸⁵ In his dissenting opinion, Judge Kavanaugh argued that Judge Brown and Judge Tatel’s dilution of the preliminary injunction standard amounted to allowing “the FTC to just snap its fingers and temporarily block a merger.”¹⁸⁶ After *Whole Foods*, the question is no longer how much the FTC must show in order to obtain a preliminary injunction, but rather how little the FTC can show in order to obtain such an injunction.

Furthermore, Judge Brown stated that it “is not to say market definition will always be crucial to the FTC’s likelihood of success on the merits. Nor does the FTC necessarily need to settle on a market definition at this preliminary stage.”¹⁸⁷ Basically, Judge Brown believes that while the FTC must define a relevant market to prevail on the merits, it does not need to do so to at the preliminary injunction stage.¹⁸⁸ Judge Brown believes that the FTC can satisfy its burden of proof by simply showing that it has a chance of defining a market, even if it initially defines the market incorrectly. Under this standard, the FTC will be able to obtain a preliminary injunction just by speculating that a merger may reduce competition. Parties seeking to merge would be at a severe disadvantage when responding to the FTC’s requests because the FTC can put forth ambiguous market definitions and argue that it will prove the correct market definition in a later administrative proceeding. Since the FTC does not have to define the market correctly at the preliminary injunction stage,

¹⁸² FTC v. Whole Foods Mkt., Inc., 548 F.3d 1028, 1035 (D.C. Cir. 2008) (citing *H.J. Heinz*, 246 F.3d at 714-15).

¹⁸³ *Id.* at 1032.

¹⁸⁴ *Id.* at 1035.

¹⁸⁵ *Id.*

¹⁸⁶ *Id.* at 1052 (Kavanaugh, J., dissenting).

¹⁸⁷ *Id.* at 1036.

¹⁸⁸ *See id.* at 1036-37. (“[T]he FTC may have alternate theories of the merger’s anticompetitive harm, depending on inconsistent market definitions One may have such doubts without knowing exactly what arguments will eventually prevail. Therefore, a district court’s assessment of the FTC’s chances will not depend, in every case, on a threshold matter of market definition.”).

this deference to the FTC allows the FTC to challenge previously marginal cases, resulting in a greater number of merger challenges.¹⁸⁹ Going forward, this lowered preliminary injunction standard makes it far easier for the FTC to block mergers in the D.C. Circuit.

B. The Whole Foods Preliminary Injunction Standard Combined with the FTC's Ability to Pursue Administrative Litigation Is Outcome-Determinative

After *Whole Foods*, it is difficult to see how the FTC would not win at the district court level, especially when it is entitled to a presumption of injunctive relief unless it has “entirely failed to show a likelihood of success.”¹⁹⁰ Even if the district court somehow denies the preliminary injunction, the FTC can still use the threat of an administrative proceeding to force the parties to settle or to terminate the transaction. The FTC’s ability to prolong a merger challenge with an administrative trial puts enormous pressure on merging parties to either settle or terminate the transaction, even though the transaction had closed. As seen in *Whole Foods*, the choice of the FTC as the investigating agency played a big role in the outcome of the case and subsequent settlement. These divergences between the DOJ and the FTC subject merging parties to different legal obligations, and impose costs and inefficiencies on parties that may be passed on to consumers.

1. The FTC Has More Leverage in the Settlement Context

The FTC’s ability to pursue administrative litigation gives it a significant advantage that the DOJ lacks in negotiating a settlement, as few parties will want to litigate a full administrative trial and face the risk of expensive and disruptive divestitures.¹⁹¹ Unlike the FTC, the DOJ enjoys no presumption in favor of preliminary injunctive relief.¹⁹² To avoid duplication of efforts, the DOJ usually agrees with the merging parties to consolidate the hearings for preliminary and

¹⁸⁹ See *infra* Part III.B.2.

¹⁹⁰ *Whole Foods*, 548 F.3d at 1035.

¹⁹¹ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 142.

¹⁹² See *Whole Foods*, 548 F.3d at 1035. Unlike the FTC, the DOJ is still subject to the traditional equity test for a preliminary injunction. See *supra* Part III.A.1.

permanent injunctions.¹⁹³ If the DOJ fails to obtain a preliminary injunction, and barring an appeal, the parties are free to consummate their transaction.¹⁹⁴ In contrast, the FTC's ability to commence administrative litigation imposes a different timeframe and uncertainties on the merging parties, giving the FTC greater leverage in negotiating a consent agreement with the parties than the DOJ.¹⁹⁵ In its Antitrust Modernization Report, the Antitrust Modernization Commission recognized that the "mere availability" of a potential administrative trial "can harm parties by creating uncertainty as to the legal status of their transaction, a risk not faced when the DOJ brings a challenge to a merger."¹⁹⁶ The threat of administrative litigation imposes delays, uncertainties, and costs on parties whose merger is reviewed by the FTC, a risk they do not face if the merger was reviewed by the DOJ.¹⁹⁷ Rather than risking lengthy and expensive litigation, parties to a proposed transaction will be more likely to either settle or terminate the transaction if the FTC is adamant about challenging the transaction.¹⁹⁸

The divergences between the two agencies are particularly acute when the FTC decides to pursue an administrative trial post-consummation. The leverage the FTC has over merging parties is even more apparent in the consummated merger context, where the parties have little choice but to either litigate the administrative trial or settle. By then, the assets are all "scrambled" and the combined entity is already well integrated. Since the parties are already well integrated, the parties will be forced to either settle from a disadvantaged bargaining position or defend their merger in a long and costly administrative proceeding, a risk they do not face if the DOJ was challenging the merger. Thus, there is

¹⁹³ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 139.

¹⁹⁴ *Id.* According to the American Bar Association's Section of Antitrust Law, "[a]lthough the DOJ has the option of seeking permanent relief in federal court after failing to obtain a [preliminary injunction], we have not been able to find any examples of the DOJ having done so." ABA Comments re Differential Standards, *supra* note 14, at 5.

¹⁹⁵ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 140-42.

¹⁹⁶ *Id.* at 139.

¹⁹⁷ See *id.*

¹⁹⁸ See Frankel, *supra* note 75, at 182 ("Relatedly, and perhaps more importantly, it is evident (particularly to members of the defense bar) that an antitrust agency seeking a merger remedy typically has considerable negotiating leverage given that it may be difficult and costly for the defendants to fight the agency in court; indeed; the merger may not survive long enough to permit a court fight.").

much greater pressure to settle a matter when the transaction is being challenged by the FTC.

2. The FTC Has More Freedom to Challenge Marginal Cases

Commentators were surprised by the FTC's persistence in its battle with Whole Foods.¹⁹⁹ The settlement between the FTC and Whole Foods confirmed their suspicions that the case against Whole Foods was marginal at best. Despite what appears to be a significant divestiture of the Wild Oats assets, the settlement demonstrates the difficulties the FTC faces in fashioning effective post-closing relief. Not only is the burden on the parties and the divergence between the agencies particularly acute when the FTC pursues administrative litigation post-consumption, but the post-closing relief that it attains is not even necessarily in the public's interest. As seen in the Whole Foods consent agreement, post-closing relief often does not fully resolve the anticompetitive harm or restore competition, thus wasting time and resources, as well as potentially harming consumers when parties pass on the costs to them.

First, the settlement does not resolve the anticompetitive harms. After the merger closed, Whole Foods began integrating Wild Oats by rebranding Wild Oats stores, closing certain Wild Oats locations, and terminating certain leases.²⁰⁰ With this amount of "scrambling," it becomes very difficult to "unscramble the eggs" and restore the acquired firm to its former status as a competitively viable company.²⁰¹ As a result, the terms of the consent agreement are certainly much less vigorous than what the FTC would have sought before the

¹⁹⁹ The influential Wall Street Journal denounced the FTC's persistence in challenging the merger as antitrust double jeopardy and a "rigged game" in an editorial. Editorial, *Whole Foods Fiasco*, WALL ST. J., Dec. 31, 2008, at A8. According to the Journal,

[t]he Whole Foods fiasco is an embarrassment for the Bush Administration's antitrust policy. This month, eight Senators on the Judiciary Committee sounded a note of caution about the FTC's actions, and no less than Democratic antitrust scourge John Conyers has said he would like to hold hearings on abolishing the FTC's administrative proceedings. When antitrust enforcement becomes a law unto itself, it's time for some organic changes for regulators.

Id.

²⁰⁰ *FTC v. Whole Foods Mkt., Inc.*, 548 F.3d 1028, 1033 (D.C. Cir. 2008).

²⁰¹ Sher, *supra* note 18, at 52-53.

consumption of the merger.²⁰² For example, the thirty two stores that Whole Foods is to divest will only provide relief in seventeen of the twenty nine geographic markets where the FTC alleged the merger would cause competitive harm.²⁰³ Over this two-year period, Whole Foods spent at least \$16.5 million defending the merger.²⁰⁴ The FTC declined to say how much it spent but it was no doubt quite a substantial amount of taxpayer's money.²⁰⁵ Despite all this time and resources, the only relief for the twelve other affected geographic markets is the mere divestiture of the Wild Oats name.

Second, it is doubtful that the settlement will restore competition. Although the FTC hopes that reestablishing a PNOS competition under the Wild Oats name will restore the competition that was eliminated by the acquisition, and provide a "springboard for broader competition nationwide," it is unclear when this future competition will occur, if it occurs at all.²⁰⁶ The two separate organic grocers before the merger, and the combined entity since the merger has forced its competitors—much larger supermarket chains with thousands of more stores—to adopt a similar strategy of offering organic natural foods and better customer services.²⁰⁷ When the merger was first announced, many analysts hailed the move as a

²⁰² Analysis of Agreement, *supra* note 160, at 5 ("The absence of pre-consummation relief from the district court, and Whole Foods' subsequent integration activities, have made it more difficult for the Commission to obtain complete relief in this matter.").

²⁰³ *Id.* at 3.

²⁰⁴ Andrew Martin, *Wait. Why Is the F.T.C. After Whole Foods?*, N.Y. TIMES, Dec. 14, 2008, at BU8.

²⁰⁵ *Id.*

²⁰⁶ Analysis of Agreement, *supra* note 160, at 5.

²⁰⁷ See David Kesmodel, *Supervalu to Launch Organic-Foods Line*, WALL ST. J., Apr. 9, 2008, at B2. For example, Supervalu Inc., the third largest U.S. food retailer by sales, recently announced a line of organic and natural foods to compete with its rivals like Whole Foods and Trader Joe's, and also to meet consumer demand. *Id.* "The company trails other conventional grocers in launching an organics line, but its selection will be among the largest. Safeway Inc. has had success with its O Organics brand, begun in late 2005, while Kroger last August introduced an expanded array of organic products under its Private Selection label." *Id.* The successes of the other supermarkets are reflected in the decline of Whole Foods'. As Andrew Martin writes:

It was not too long ago that Whole Foods, based in Austin, Tex., was a darling of Wall Street and routinely registered double-digit growth in comparable store sales, a common industry measure of the health of stores. But the company has been battered by competition from traditional grocery stores that have expanded their offerings of organic and natural foods.

Andrew Martin, *Private Equity Firm Buys 17% of Whole Foods*, N.Y. TIMES, Nov. 5, 2008 at B12.

sensible deal and necessary if both firms wanted to continue growing as they tried to fend off much larger supermarkets.²⁰⁸ However, due to intense competition from other supermarkets, the general economic condition over the last few years, and the ensuing decrease in consumer spending, little has gone right for Whole Foods since 2007.²⁰⁹ Fortunately, amid a stabilizing economy, recent earnings data suggest the company is on the road to recovery.²¹⁰

As supermarket chains continue to enter this lucrative market, consumers have benefitted from the increased number of competitors.²¹¹ It is difficult to see how consumers will benefit from a reconstituted Wild Oats or a weakened Whole Foods, and even more difficult to see who would buy the Wild Oats name. A reconstituted Wild Oats will be a greatly weakened firm, facing off against a greater number of competitors than before the merger. Whatever the benefits of the divestiture, it is outweighed by the time and resources the FTC has put into this matter. It is difficult to believe that a divestiture of thirteen operating stores and the Wild Oats name is a victory for the FTC after a two year legal battle.

Before the *Whole Foods* decision, it was debatable if the different legal preliminary injunction standards would produce different results.²¹² But as *Whole Foods* made clear, whether a proposed transaction may proceed does depend on which

²⁰⁸ One analyst compared the merger “to two knights who decide to stop fighting each other so they can protect the castle against bigger competitors.” Martin, *supra* note 1.

²⁰⁹ See David Kesmodel, *Corporate News: Whole Foods Gets Infusion, Posts Steep Drop in Net*, WALL ST. J, Nov. 6, 2008, at B2 (“[Whole Foods] has been hit hard by the weak economy as consumers cut back on discretionary spending and buy more store brands and discounted groceries.”); Timothy W. Martin, *Corporate News: Whole Foods to Sell 31 Stores in FTC Deal*, WALL ST. J, March 7, 2009, at B5 (“A lot has changed since 2007, when the FTC said the merger would ‘mean higher prices, reduced quality and fewer choices for consumers.’ In the past year, Whole Foods has seen its profits battered by the economic recession and stiffer competition from traditional food retailers like Safeway Inc. and Supervalu Inc.”).

²¹⁰ See Paul Sonne & Timothy W. Martin, *Whole Foods Profit Jumps as Turnaround Takes Root*, WALL ST. J, Feb. 17, 2010, at B5. “The Austin, Texas-based grocery chain reported profit of \$49.7 million, or 32 cents a share, compared with \$27.8 million, or 20 cents a share [in 2009]. Same-store sales, a key measure of retail health, rose 2.5%. Total sales for the quarter ended Jan. 17 climbed 7% to \$2.6 billion from \$2.47 billion.” *Id.* Whole Foods attributed the gains to a retooled strategy of lower prices and smaller stores to gain customers back. *Id.*

²¹¹ See Editorial, *supra* note 199 (“The market for natural and organic produce has exploded, with every discount outlet from Wal-Mart to Wegmans now offering organic products.”).

²¹² See *supra* note 172; see also ABA Comments re Differential Standards, *supra* note 14, at 4.

agency is reviewing the transaction. Unlike the FTC, the DOJ enjoys no presumption in favor of preliminary injunctive relief.²¹³ If the DOJ fails to obtain a preliminary injunction, and any appeals (very unlikely) are exhausted, the parties are free to consummate their transaction. If the DOJ was the investigating agency in *Whole Foods*, Whole Foods would have been free to consummate the merger after its victory at the district court level. Unlike the FTC, the DOJ lacks the power to prolong challenged mergers until the parties settle. The effect of this divergence between the agencies is inconsistency and unpredictability in the marketplace. After *Whole Foods*, it is clear that the best possible outcome for parties seeking to merge would be for the proposed transaction to be cleared for review by the DOJ. The perception that the choice of investigating agency is outcome-determinative is no longer just perception. It is a fact.

IV. SOLUTION: MINIMIZING THE DIVERGENCES

In a dual enforcement system, it should not matter which agency is challenging the merger. The merging parties should “receive comparable treatment and face similar burdens regardless of whether the FTC or the DOJ reviews their merger.”²¹⁴ Divergences between the FTC and the DOJ undermine consistency, predictability, efficiency, and fairness in the merger review process.²¹⁵ More importantly, such divergences make it clear that the ultimate decision as to whether a transaction may proceed depends on which agency is reviewing the transaction.²¹⁶

This Note discusses two approaches to harmonizing the divergences between the FTC and the DOJ. The first approach is a judicial solution: to limit the expansive language in *Whole Foods*. The second approach is a legislative approach, specifically for Congress to amend the FTC Act to specify that the FTC is subject to the same standard for the grant of a preliminary injunction as the DOJ.

²¹³ See *FTC v. Whole Foods Mkt., Inc.*, 548 F.3d 1028, 1035 (D.C. Cir. 2008).

²¹⁴ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 138.

²¹⁵ See *id.* at 138-39.

²¹⁶ See *id.*

A. *Judicial Solution*

The most significant difference between the July and November rulings in *Whole Foods* was that Judge Tatel no longer concurred in Judge Brown's opinion but only in the judgment of the court.²¹⁷ As a result of Judge Tatel's revision, Judge Brown's opinion was no longer the majority opinion of the court, and there were questions about the precedential value of the decision.²¹⁸ However, any hope that the D.C. Circuit would limit the expansive language found in *Whole Foods* was dashed in *FTC v. CCC Holdings, Inc.*,²¹⁹ where the preliminary injunction standard articulated in *Whole Foods* helped the FTC obtain their first preliminary injunction from a federal district court in nearly seven years.²²⁰ On March 18, 2009, less than a week after *Whole Foods* settled with the FTC, Judge Collyer issued a preliminary injunction that stopped the proposed merger between CCC Holdings and Mitchell International.²²¹ The FTC had argued that the merger of CCC Holdings and Mitchell International would reduce the number of competitors in the relevant market from three to two.²²² In enjoining the CCC Holdings merger, Judge Collyer adopted Judge Brown's diluted preliminary injunction standard articulated in *Whole Foods*.²²³ As a result, Judge Collyer granted the injunction because the FTC had "raised questions going to the merits so serious, substantial, difficult and doubtful as to make them fair ground for thorough investigation, study, deliberation and determination by the FTC."²²⁴ Rather than risking an administrative trial, CCC Holdings and Mitchell International abandoned their transaction on March 11, 2009.²²⁵

²¹⁷ *Whole Foods*, 548 F.3d at 1028.

²¹⁸ See *id.* at 1061 n.8 (Kavanaugh, J., dissenting). According to Judge Kavanaugh, "this confused decision will invite years of uncertainty and litigation over what the holding of this case is—a separate but important problem with the Court's approach." *Id.*

²¹⁹ 605 F. Supp. 2d 26 (D.D.C. 2009).

²²⁰ Press Release, Fed. Trade Comm'n, *Competition Acting Director David Wales to Leave FTC*, (April 1, 2009), available at <http://www.ftc.gov/opa/2009/04/dwales.shtm>.

²²¹ *CCC Holdings*, 605 F. Supp. 2d at 30.

²²² *Id.*

²²³ *Id.* at 35-36.

²²⁴ *Id.* at 30 (citing *Heinz*, 246 F.3d at 714-15; *Whole Foods*, 548 F.3d at 1035 (Brown, J.); *id.* at 1042 (Tatel, J., concurring)).

²²⁵ See Press Release, CCC Information Services Inc., *supra* note 9.

CCC Holdings, Inc. made it clear that Judge Brown's opinion was now binding precedent.²²⁶ The FTC read the case in the same way. In a speech about the FTC's relationship to the legislative, executive, and judicial branches, FTC Commissioner Thomas Rosch stated that:

The *CCC* case confirms that *Whole Foods* established a new standard for FTC preliminary injunctions—at least for the D.C. Circuit. It is also clear, based on the express language in both the *Whole Foods* and *CCC* opinions, that this new standard will allow the FTC to more readily obtain a preliminary injunction from a federal district court. We are unlikely to see decisions like *Arch Coal* again in the D.C. District Court, where the FTC brings most of its merger enforcement actions and which is required to apply the law of the D.C. Circuit. Of course, there are 11 other regional circuit courts for which *Whole Foods* is not binding authority, but I expect that these courts will adopt the D.C. Circuit's 13(b) standard as the opportunity arises.²²⁷

After *CCC Holdings*, it appears that the best approach to harmonizing the divergences between the DOJ and the FTC would have to be a legislative approach.

B. *Legislative Solution*

The Antitrust Modernization Commission was created to undertake “a comprehensive review of U.S. antitrust law to determine whether it should be modernized.”²²⁸ It delivered its final report in 2007.²²⁹ One area of antitrust law the Commission reviewed and recommended improvements for was the enforcement process. The Commission recognized that there was “a perception, if not a reality, that the FTC and the DOJ face different standards for obtaining a preliminary injunction,” and called for Congress to remedy the growing

²²⁶ See *CCC Holdings*, 605 F. Supp. 2d at 36. According to Judge Collyer, “precedents irrefutably teach that in this context ‘likelihood of success on the merits’ has a less substantial meaning than in other preliminary injunction cases. *Heinz* not only emphasized this point but *Whole Foods* makes clear that *Heinz* remains good law.” *Id.* at 36 n.11.

²²⁷ J. Thomas Rosch, Commissioner, Fed. Trade Comm'n, Thoughts on the FTC's Relationship (Constitutional and Otherwise) to the Legislative, Executive, and Judicial Branches, Remarks before the Berlin Forum for EU-US Legal-Economics Affairs, at 20-21 (Sept. 19, 2009), available at <http://www.ftc.gov/speeches/rosch/090919roschberlinspeech.pdf>.

²²⁸ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at i.

²²⁹ *Id.*

divergences between the DOJ and the FTC.²³⁰ The Commission made three recommendations in this regard:

24. The Federal Trade Commission should adopt a policy that when it seeks injunctive relief in Hart-Scott-Rodino Act merger cases in federal court, it will seek both preliminary and permanent injunctive relief, and will seek to consolidate those proceedings so long as it is able to reach agreement on an appropriate scheduling order with the merging parties.

25. Congress should amend Section 13(b) of the Federal Trade Commission Act to prohibit the Federal Trade Commission from pursuing administrative litigation in Hart-Scott-Rodino Act merger cases.

26. Congress should ensure that the same standard for the grant of a preliminary injunction applies to both the Federal Trade Commission and the Antitrust Division of the Department of Justice by amending Section 13(b) of the Federal Trade Commission Act to specify that, when the Federal Trade Commission seeks a preliminary injunction in a Hart-Scott-Rodino Act merger case, the Federal Trade Commission is subject to the same standard for the grant of a preliminary injunction as the Antitrust Division of the Department of Justice.²³¹

The Commission believed these recommendations would eliminate the divergences between the DOJ and the FTC. If the FTC seeks both preliminary and permanent injunctive relief in the same proceeding, this practice would be consistent with the DOJ's current approach, and would eliminate the difference in burden of proof for the two agencies.²³² To avoid the appearance of inconsistency, unpredictability, and unfairness in the merger review process, the Commission recommended the elimination of the FTC's ability to commence administrative litigation in HSR Act merger cases.²³³ The Commission believed that the elimination of administrative litigation would provide the FTC with further incentive to seek permanent relief in a district court, and not in an administrative trial.²³⁴ Finally, to ensure

²³⁰ *Id.* at 141.

²³¹ *Id.* at 139-41.

²³² *Id.* at 139. Since the DOJ seeks both preliminary and permanent injunctions in the same proceeding, it has to prove the proposed transaction "would violate Section 7 of the Clayton Act by a preponderance of the evidence." In contrast, the FTC only seeks a preliminary injunction so its burden required for obtaining a preliminary injunction is lower. *Id.*

²³³ *See id.* at 140.

²³⁴ *Id.* at 141. The Commission noted that the elimination of administrative trials would only apply in HSR Act merger cases, and would not affect the FTC's ability to commence administrative trials for consummated mergers. "The proposed statutory

that courts apply the same preliminary injunction standard for both enforcement agencies, the Commission recommended that Congress amend the FTC Act to remove any standard for granting a preliminary injunction in HSR cases.²³⁵ The Commission believed that the elimination of the “public interest” language would lead courts to apply the traditional equity test used for the DOJ.²³⁶

While the Commission’s recommendations will likely play a large role in the implementation of any legislative solution, it appears the recommendation with the best chance for success would be for Congress to amend Section 13(b) of the FTC Act to specify that the DOJ and the FTC is subject to the same preliminary injunction standard. The Commission’s other two recommendations, eliminating administrative proceedings in HSR Act merger cases and requiring the FTC to seek both preliminary and permanent injunctive relief in the same proceedings, are likely to engender significant opposition from the FTC and the Obama administration. During his presidential campaign, President Obama made it clear that he would direct his administration “to reinvigorate antitrust enforcement” and “step up review of merger activity and take effective action to stop or restructure those mergers that are likely to harm consumer welfare, while quickly clearing those that do not.”²³⁷ It is unlikely that the Obama administration, which has signaled its support for vigorous antitrust enforcement, would weaken the FTC’s antitrust enforcement ability by eliminating its administrative litigation authority.²³⁸ Any such efforts to require the FTC to emulate the DOJ’s current practice would be politically very difficult. Thus, the recommendation with the best chance of success is to amend

bar would not preclude the FTC from pursuing an administrative complaint after the consummation of a merger, based on evidence that the merger has had actual, as opposed to predicted, anticompetitive effects.” *Id.*

²³⁵ *Id.* at 141-42.

²³⁶ *Id.* at 142.

²³⁷ Senator Barack Obama, Statement of Senator Barack Obama for the American Antitrust Institute (2007), available at http://www.antitrustinstitute.org/archives/files/aai-%20Presidential%20campaign%20-%20Obama%209-07_092720071759.pdf.

²³⁸ See Justin Blum, *Intel Case May Signal Increased Antitrust Enforcement*, BLOOMBERG, Dec. 17, 2009, <http://www.bloomberg.com/apps/news?pid=20670001&sid=a4BNq.ad67N0>; Stephen Labaton, *Administration Plans to Strengthen Antitrust Rules*, N.Y. TIMES, May 11, 2009, at A1; Elizabeth Williamson & Matthew Karnitschnig, *U.S. Signals More Scrutiny of Mergers, Antitrust*, WALL ST. J., May 12, 2009, at B1.

the FTC Act to specify that the FTC be subject to the same standard for the grant of a preliminary injunction as the DOJ.

CONCLUSION

To be effective, an enforcement system “must be clear, fairly administered, and not unreasonably burdensome.”²³⁹ After *Whole Foods*, our dual antitrust enforcement system is anything but that. There has been a running debate as to whether the FTC and the DOJ face different standards for obtaining a preliminary injunction, namely that the FTC enjoys a lower preliminary injunction standard than the DOJ.²⁴⁰ The D.C. Circuit Court’s opinion in *Whole Foods* puts that debate to rest in the D.C. Circuit with an empathetic “yes.”²⁴¹ The FTC’s ability to commence administrative litigation,²⁴² coupled with *Whole Foods*’ articulation of the preliminary injunction standard for the FTC,²⁴³ create divergences between the DOJ and the FTC. In a dual enforcement system, such divergences undermine consistency, predictability, efficiency, and fairness in the merger review process. More importantly, such divergences make it clear that the ultimate decision as to whether a transaction may proceed depends on which agency is reviewing the transaction.

After *Whole Foods*, the arbitrary allocation of a proposed transaction to either the FTC or the DOJ for review can result in a very different substantive outcome for the transaction. In a system of shared responsibility for enforcement of the federal antitrust laws, such an outcome is unacceptable. In such a system, merging parties should expect comparable treatment and burden, as well as a comparable outcome.²⁴⁴ Since a judicial solution is unlikely,²⁴⁵ and any attempt to eliminate the FTC’s administrative litigation authority is politically difficult,²⁴⁶ the most politically promising solution to stem the growing divergence between the DOJ and the FTC is for Congress to amend the FTC Act to specify that the same preliminary injunction standard applies to both

²³⁹ ANTITRUST MODERNIZATION REPORT, *supra* note 13, at iv.

²⁴⁰ See *supra* note 172 and accompanying text.

²⁴¹ See *supra* Part III.A.

²⁴² See *supra* Part I.A.3.

²⁴³ See *supra* Part III.A.

²⁴⁴ See ANTITRUST MODERNIZATION REPORT, *supra* note 13, at 131.

²⁴⁵ See *supra* Part IV.A.

²⁴⁶ See *supra* Part IV.B.

enforcement agencies. By doing so, Congress will ensure that the costs of merger review do not overwhelm the benefits of a fair and effective antitrust enforcement system.

Raymond Z. Ling[†]

[†] J.D. Candidate, Brooklyn Law School, 2010; B.A., New York University, 2006. I would like to thank my family and friends for all of their support and encouragement. I would also like to thank the members of the *Brooklyn Law Review* for their hard work with this Note. Special thanks to Celine Chan, Samuel Gordon, Matthew Handler, Brittany Nilson, Joseph Roy, and Andrei Takhteyev for their invaluable comments and suggestions.

Competing with Antitrust Laws

HOW NEW YORK'S POST AND HOLD LIQUOR LAW WILL LOSE AGAINST THE SHERMAN ACT

INTRODUCTION

The distribution and sale of alcohol has a colorful past in this country,¹ and in certain respects, remains controversial to this day.² While we no longer live in the “Wild West” and are thus no longer concerned with the dangers of saloons,³ alcohol is still linked to a host of societal problems.⁴ In response to the social evils tied to alcohol, the states rely on both their police powers and the authority granted to them by the Twenty-first Amendment to promote temperance.⁵ Accordingly, the states have enacted a variety of laws designed to achieve this goal.⁶

¹ See generally Sidney J. Spaeth, *The Twenty-First Amendment and State Control Over Intoxicating Liquor: Accommodating the Federal Interest*, 79 CAL. L. REV. 161, 165-80 (1991) (recounting the origins of the battle for prohibition as linked to “the ubiquitous image of the debauched saloon” as fuel for the temperance movement). For example, in the 1800s, violence broke out against saloons, and women who were advocating temperance resorted to attacking saloons, with the hatchet as their “weapon of choice.” See *id.* at 169 (describing Carry Nation’s “hatchetations” on Kansas saloons”).

² See *infra* note 12 and accompanying text.

³ See Clayton L. Silvernail, *Smoke, Mirrors and Myopia: How the States Are Able to Pass Unconstitutional Laws Against the Direct Shipping of Wine in Interstate Commerce*, 44 S. TEX. L. REV. 499, 505 (2003).

⁴ Alcohol is a leading cause of car accidents, especially car accident fatalities. See, e.g., Bureau of Transportation Services: Table 2-20: Fatalities in Crashes by Number of Vehicles and Alcohol Involvement, http://www.bts.gov/publications/national_transportation_statistics/html/table_02_20.html (last visited Jan. 4, 2010); MADD Campaign to Eliminate Drunk Driving: Statistics, <http://www.madd.org/Drunk-Driving/Drunk-Driving/Statistics/AllStats.aspx> (last visited Jan. 4, 2010). Alcohol is a necessary component of binge drinking and often a major factor in hazing among college students. See generally HANK NUWER, *WRONGS OF PASSAGE: FRATERNITIES, SORORITIES, HAZING, AND BINGE DRINKING* (1999). Excessive use of alcohol leads to severe liver damage and is actually the “[third] leading lifestyle-related cause of death for people in the United States each year.” Centers for Disease Control and Prevention: Alcohol and Public Health, <http://www.cdc.gov/alcohol/> (last visited Feb. 3, 2010).

⁵ “The transportation or importation into any State, Territory, or possession of the United States for delivery or use therein of intoxicating liquors, in violation of the laws thereof, is hereby prohibited.” U.S. CONST. amend. XXI, § 2.

⁶ See, e.g., *infra* notes 34, 99, 106, 151-152 and accompanying text.

New York, like many other states,⁷ has enacted “price posting” and “post and hold” liquor laws designed to regulate the distribution of alcohol within the state.⁸ New York’s price posting law requires manufacturers and wholesalers to file price schedules that report future prices.⁹ The law is also considered a “post and hold” law, as it requires them to make resale prices public, then hold those prices for a defined period of time, rather than allow prices to fluctuate based on market forces.¹⁰

As with many other liquor laws,¹¹ there has been controversy over the years regarding the validity of the price posting laws, especially those classified as post and hold.¹² This Note argues that New York’s post and hold law is in contravention of the Sherman Antitrust Act¹³ and hard to defend under current Twenty-first Amendment jurisprudence, which has steadily limited the broad grant of powers given to the states regarding the control of liquor within their borders.

Part I provides background information on both the history of alcohol distribution in the United States and the Sherman Act, as well as a brief overview of New York’s distribution system. Part II analyzes the evolution of the United States Supreme Court jurisprudence on the Twenty-first Amendment, illustrating how the Court has shifted its interpretation of the Amendment from a broad grant of authority to the states over liquor regulation to a balancing of competing interests.¹⁴ Part III examines the steps of a federal antitrust challenge to a state law, and describes the federal decisions that specifically address the validity of price posting laws in Oregon,¹⁵ Maryland,¹⁶ and Washington,¹⁷ which closely resemble New York’s statute. Part IV concludes that the New

⁷ For example, Oregon, Maryland, and Washington all had price posting laws. See *infra* notes 222-223, 235, 249 and accompanying text.

⁸ See N.Y. ALCO. BEV. CONT. LAW § 101-b (McKinney 2009).

⁹ *Id.* § 101-b(3)(a).

¹⁰ *Id.* § 101-b(3)(b).

¹¹ See, e.g., *infra* notes 34, 97-99, 106, 151-152 and accompanying text.

¹² See *Costco Wholesale Corp. v. Maleng*, 522 F.3d 874, 895-96 (9th Cir. 2008) (holding that Washington’s “post and hold” liquor distribution regime violated the Sherman Act); accord *TFWS, Inc. v. Schaefer*, 242 F.3d 198, 209-10 (4th Cir. 2001); *Miller v. Hedlund*, 813 F.2d 1344, 1349-51 (9th Cir. 1987); see also *infra* Part III.B.

¹³ See 15 U.S.C. § 1-7 (2006).

¹⁴ See *infra* notes 81-85 and accompanying text.

¹⁵ See *Miller*, 813 F.2d 1344.

¹⁶ See *TFWS, Inc.*, 242 F.3d 198.

¹⁷ See *Costco*, 522 F.3d 874.

York statute will not be able to survive an antitrust challenge in light of these recent cases because the state most likely cannot put up the formidable defense now required by the Supreme Court. As it stands, New York's post and hold regime should be struck down if challenged in court.

In late 2009, the New York State Law Revision Commission completed a two-year review of the New York Alcohol Beverage Control Law (the "ABC Law"), of which price posting is one narrow issue.¹⁸ In its final report, the Commission flagged Section 101-b as a potential source of legal problems for the state due to the evolving nature of Twenty-first Amendment jurisprudence, especially in light of the invalidation of several post and hold liquor laws by the federal courts.¹⁹ Thus, the state may actually be in a good position to take protective action if it so desires.

I. BACKGROUND OF ALCOHOL AND ANTITRUST LAWS

A. *A Brief History of Alcohol Distribution in the United States*

With the exception of Prohibition, there has always been prolific regulation of commerce in alcohol by the states, largely because governments viewed it as a source of revenue.²⁰ The real controversy over alcohol regulation began in the early to mid-1800s, when religious opposition began to form against alcohol consumption.²¹ As a result of the lobbying efforts of anti-saloon groups and temperance societies, many states passed laws banning saloons and the "manufacture of 'spirituous or

¹⁸ See NEW YORK STATE LAW REVISION COMMISSION REPORT ON THE ALCOHOL BEVERAGE CONTROL LAW AND ITS ADMINISTRATION 1, 23 (December 15, 2009), available at <http://www.lawrevision.state.ny.us/abcls.php> (last visited Feb. 18, 2010).

¹⁹ See *id.* at 217-20 (warning that the fact that Section 101-b survived an earlier antitrust challenge "should not make . . . the Legislature sanguine about the price posting and hold requirements.").

²⁰ See RICHARD MCGOWAN, GOVERNMENT REGULATION OF THE ALCOHOL INDUSTRY: THE SEARCH FOR REVENUE AND THE COMMON GOOD 3-6, 35 (1997) ("Throughout the history of the American alcohol industry, government has played a pivotal role in determining where, when, and how alcoholic beverages are sold. Every level of government (federal, state, and local) has revenue as well as regulatory interest in the industry.").

²¹ See W. J. Rorabaugh, *Reexamining the Prohibition Amendment*, 8 YALE J.L. & HUMAN. 285, 288 (1996) ("The temperance campaign that started in the 1820's demanded personal abstinence both as the price of church membership and as a badge of middle-class respectability."); MCGOWAN, *supra* note 20, at 41 (noting that there was no religious opposition to the alcohol industry prior to the 1850s).

intoxicating liquors.”²² Enforcement was largely unsuccessful because liquor continued to be smuggled across state lines,²³ and ultimately, the laws were repealed or struck down by state courts, at least in part, as unconstitutional.²⁴ However, when the issue reached the Supreme Court in *Mugler v. Kansas*,²⁵ the Court held that complete prohibition of alcohol sale and production was within a state’s police power.²⁶ Thirteen years later, the Supreme Court declared that *Mugler* “stood for the ‘undoubted right’ of states to regulate their internal affairs.”²⁷

While the states were given free range to extensively regulate alcohol within their borders, Prohibitionists²⁸ were

²² Silvernail, *supra* note 3, at 505; *see also* Spaeth, *supra* note 1, at 168-69. These laws were commonly referred to as “Maine laws.” *See id.* (internal quotation marks omitted); *see also* Rorabaugh, *supra* note 21, at 288-89 (noting that Maine was the first state to enact prohibition).

²³ Rorabaugh, *supra* note 21, at 289.

²⁴ There were several factors that led to the repeal of most of these early prohibition laws. First, they met strong opposition from immigrants and anti-reform groups. *See* JACK S. BLOCKER, IAN R. TYRRELL, AND DAVID M. FAHEY, ALCOHOL AND TEMPERANCE IN MODERN HISTORY: A GLOBAL ENCYCLOPEDIA 395 (2003) (explaining that immigrants and anti-reform groups became instrumental in the demise of the Maine laws by joining anti-temperance coalitions to speak out against them). Moreover, the laws were undermined by court rulings, including being struck down as unconstitutional. *See, e.g.*, *People v. Toynbee*, 2 Parker Crim. Rep. 329 (N.Y. 1855) (finding that the prohibition of the sale of intoxicating liquors was an unconstitutional interference with property in violation of individuals’ due process rights); *see also* BLOCKER, *supra* at 395 (by the end of the Civil War, most prohibition laws were “unenforced, overturned, or struck down by state courts as unconstitutional.”); *see also* RICHARD F. HAMM, SHAPING THE EIGHTEENTH AMENDMENT: TEMPERANCE REFORM, LEGAL CULTURE, AND THE POLITY, 1880-1920, at 20 (1995) (stating that most prohibition laws were rendered ineffective by the courts). Finally, the Civil War distracted proponents of temperance, to the point where “progress” was not only halted but reversed. *Id.* Even the laws that still survived after the Civil War suffered from “admittedly lax enforcement.” *Id.*

²⁵ 123 U.S. 623 (1887). Kansas was the first state to “go dry” by going further than the Maine laws and amending its constitution to forbid the manufacture and sale of alcohol within its borders. *See* Silvernail, *supra* note 3, at 505-06.

²⁶ *Mugler*, 123 U.S. at 662.

²⁷ Silvernail, *supra* note 3, at 507 (quoting *Leisy v. Hardin*, 135 U.S. 100, 122 (1890)) (“The effect of *Leisy* and prior decisions . . . was to give states carte blanche with regards to regulating intoxicating liquors within their bounds.”).

²⁸ An early group of Prohibitionists were working-class Americans—the Washingtonians—who “pledg[ed] not to drink any alcoholic liquors.” *See* HAMM, *supra* note 24, at 20. The movement spread to the middle class, who also sought to end alcohol consumption. *Id.* After the Civil War, abstinence increasingly became a religious talking point, “[i]n particular, pietists, members of evangelical sects, including Baptists, Methodists, and Presbyterians . . . saw prohibition as a needed corrective to the nation’s moral laxity and resulting social problems.” *Id.* at 22 (noting that religion was linked with ethnicity, and therefore Americans of English, Scottish, and sometimes Scandinavian descent were more likely to support Prohibition than the Irish, German, Italian, and Polish.) Two major Prohibitionist groups were the Prohibition Party, whose members came primarily from the Republican Party, and the National Women’s Christian Temperance Union, which consisted of middle-class

concerned that the states could not ban the importation of alcohol.²⁹ In response to this dilemma, they lobbied Congress for states' rights to keep alcohol out entirely, and their efforts were rewarded by the enactment of the Wilson Act in 1890.³⁰ The Wilson Act subjected imported alcohol to a state's applicable laws upon arrival.³¹ Unfortunately for the Prohibitionists,³² soon after the Wilson Act's enactment, the Supreme Court suggested that the Act did not give the states permission to prohibit the importation of alcohol.³³ The Court cemented this position seven years later, explicitly holding that laws interfering with the importation of alcohol were "wholly incompatible with and repugnant to" individuals' constitutional right to ship and receive goods to and from another state.³⁴

Perhaps deflated by the Supreme Court's interpretation of the Wilson Act, the Prohibitionists lobbied Congress yet again to give states the right to ban alcohol importation,³⁵ which led to the passage of the Webb-Kenyon Act in 1913.³⁶ The Webb-Kenyon Act specifically authorized states to keep alcohol out of their borders.³⁷ The Act survived a constitutional

women. *See id.* at 23-24. The Anti-Saloon League, "one of the most powerful political organizations in United States history," was another prohibitionist group, made up of churches and other temperance societies. Spaeth, *supra* note 1, at 170.

²⁹ *See* *Bowman v. Chicago & Northwestern Ry. Co.*, 125 U.S. 465, 500 (1888).

³⁰ *See* 27 U.S.C. § 121 (2006).

³¹ *Id.* (stating, in relevant part, that imported alcohol "shall upon arrival in [a] State . . . be subject to the operation and effect of the laws of such State . . . enacted in the exercise of its police powers").

³² *See* Silvernail, *supra* note 3, at 508 (describing the Wilson Act as a "hollow victory for the Prohibitionists" because the Supreme Court failed to interpret the Act as authority for the states to ban liquor importation).

³³ *Id.* at 509 ("[The Wilson Act] simply removed an impediment to the enforcement of the state laws [i]t imparted no power to the state not then possessed") Many people argue that the Wilson Act was intended to allow dry states to remain dry, or at least that the Supreme Court should have interpreted it in such a way. *See, e.g.*, Spaeth, *supra* note 1, at 172-73 & n.81 (explaining the impetus behind the passage of the Wilson Act, and quoting Senator Kenyon of Iowa, who expressed dismay that the Act was not used in such a way to give states the option to remain dry).

³⁴ *Vance v. W.A. Vandercook Co.*, 170 U.S. 438, 455 (1898) (striking down a law giving state agents the exclusive right to purchase imported alcohol, because the law gave the state, via its agents, opportunity to discriminate against sister states by selectively choosing which to buy from); *see also* *Rhodes v. Iowa*, 170 U.S. 412, 420 (1898) (rejecting Iowa's argument that the phrase "upon arrival" in the Wilson Act gave the state authority to seize imported alcohol the moment it crossed state lines because such an interpretation would give Iowa's law "extraterritorial operation," thus "render[ing] the act of Congress repugnant to the Constitution of the United States").

³⁵ *See* Silvernail, *supra* note 3, at 511.

³⁶ *See* 27 U.S.C. § 122 (2006).

³⁷ *Id.* (providing, in relevant part, that "[t]he shipment or transportation . . . of any [alcohol] . . . from one State . . . into any other State . . . in violation of any law of

challenge in *James Clark Distilling Co. v. Western Maryland Railway Company*,³⁸ in which the Supreme Court upheld a West Virginia law that prohibited the importation of alcohol for personal use.³⁹ It was not long after the passage of the Webb-Kenyon Act that the Prohibitionists finally achieved their desired goal with the ratification of the Eighteenth Amendment in 1919, which banned alcohol entirely.⁴⁰ However, the Prohibitionists ultimately lost their battle when the experiment of Prohibition failed miserably by lasting a mere fourteen years.⁴¹ The Twenty-first Amendment was ratified in 1933 and is the current source of constitutional authority granted to the states regarding the regulation of liquor.⁴² The Supreme Court's interpretation of the extent of this authority has fluctuated over time,⁴³ as discussed in detail in Part III. The price posting liquor laws have repeatedly been challenged as violations of the Sherman Act; thus, an elementary understanding of antitrust law, specifically the Sherman Act, will be helpful in addressing the constitutionality of New York's price posting law.

B. *A Brief Introduction to the Sherman Antitrust Act*

Antitrust laws . . . are the Magna Carta of free enterprise.⁴⁴

For better or worse,⁴⁵ the American economy is founded on free enterprise. In order for free enterprise to produce and

such State . . . is prohibited"). Interestingly, President Taft vetoed the Act as "an unconstitutional delegation by Congress to the states of the exclusive power to regulate interstate commerce in liquors." Spaeth, *supra* note 1, at 173-74. Congress, however, overrode Taft's veto. *Id.* at 174.

³⁸ 242 U.S. 311 (1917).

³⁹ *See id.* at 332.

⁴⁰ U.S. Const. amend. XVIII (repealed 1933) ("[T]he manufacture, sale, or transportation of intoxicating liquors within, the importation thereof into, or the exportation thereof from the United States . . . is hereby prohibited."); *see also* Silvernail, *supra* note 3, at 512.

⁴¹ The Eighteenth Amendment completed ratification in 1919 and was repealed by the Twenty-first Amendment in 1933. *See* U.S. CONST. amend. XVIII, Historical Notes (repealed 1933); U.S. CONST. amend. XXI.

⁴² U.S. CONST. amend. XXI; *see also* Silvernail, *supra* note 3 at 500 ("The Twenty-first Amendment gives the states the power to regulate the manufacture, distribution, and sale of intoxicating liquors within their borders").

⁴³ *See infra* notes 81-85 and accompanying text.

⁴⁴ *United States v. Topco Assocs., Inc.*, 405 U.S. 596, 610 (1972).

⁴⁵ *See* JOHN H. SHENEFIELD & IRWIN M. STELZER, *THE ANTITRUST LAWS* 6 (1993) (noting that a common critique of America's market economy is its potential for abuse and inevitable inequality).

maintain a flourishing economy, there must be competition.⁴⁶ Competition promotes consumer welfare⁴⁷ and efficiency.⁴⁸ Since competition is vital to the success of the American economy, it is no surprise that laws have been enacted to prevent private actors from subverting it.⁴⁹ The most basic, and most important, “pro-competition” law is the Sherman Antitrust⁵⁰ Act⁵¹ (the “Sherman Act”).⁵²

The Sherman Act has lofty goals: it seeks to protect and encourage producers by “diffus[ing] economic power and maximiz[ing] individual opportunity” to create a “fair” playing field, while simultaneously “maximiz[ing] efficiency and consumer welfare.”⁵³ To effect these goals, the Sherman Act “proscribes agreements in restraint of trade”⁵⁴ as well as “monopoly abuse.”⁵⁵ Relevant types of prohibited restraints of

⁴⁶ See *id.* at 7 (“The engine of free enterprise is competition.”).

⁴⁷ See *id.* The authors point out that “[n]umerous sellers, vying for customers, must produce goods and services of sufficient quality, and at acceptable prices, or be driven from the field.” *Id.* Such a system results in better (and usually more) options for consumers because sellers have an incentive to be innovative to attract new customers and increase profits, or at least maintain a consistent quality of product to keep their current customers. See *id.* at 12.

⁴⁸ See *id.* at 7 (“[The] necessity [of vying for consumers] forces [sellers] to be efficient, to buy so-called inputs—labor and materials—at the lowest possible prices, and . . . [keep] production costs . . . to a minimum.”).

⁴⁹ See *id.* (noting that one way competition can fail is when “private participants in the market subvert competition and thus prevent market forces from operating freely”).

⁵⁰ “Antitrust” laws are so named as a result of practices of the large enterprises of Standard Oil, sugar, whiskey, and others of taking the forms of “trusts,” placing “shareholder voting power in the hands of a single managing trustee.” See *id.* at 8.

⁵¹ 15 U.S.C. §§ 1-7 (2006). Additional antitrust laws were enacted to strengthen the Sherman Act. See SHENEFIELD & STELZER, *supra* note 45, at 9. Major ones include the Clayton Act, and the Federal Trade Commission Act. *Id.*

⁵² WALTER ADAMS & HORACE M. GRAY, *MONOPOLY IN AMERICA*, v (1955) (describing the Sherman Act as “the first and most important antitrust [law]”).

⁵³ SHENEFIELD & STELZER, *supra* note 45, at 13; see also ADAMS & GRAY, *supra* note 52, at 177 (“competition provides an effective technique for reconciling the dual objectives of economic welfare and economic freedom.” (internal quotation marks omitted)).

⁵⁴ SHENEFIELD & STELZER, *supra* note 45, at 14; see also 15 U.S.C. § 1. A restraint of trade refers to an action or condition that is intended to prevent free competition in business. BLACK’S LAW DICTIONARY (8th ed. 2004). The Sherman Act refers to “contract, combination, . . . or conspiracy” as opposed to the term “agreement.” 15 U.S.C. § 1. Because Professors Lopatka and Page make a compelling argument that the Supreme Court is primarily concerned with the “element of agreement” when applying this section, see John E. Lopatka & William H. Page, *State Action and the Meaning of Agreement Under the Sherman Act: An Approach to Hybrid Restraints*, 20 YALE J. ON REG. 269, 278-79 & notes 38-43 (2003), this Note will generally use the term “agreement” as well to reflect the Sherman Act’s prohibition of concerted action to unreasonably restrain trade.

⁵⁵ See SHENEFIELD & STELZER, *supra* note 45, at 17; see 15 U.S.C. § 2.

trade include vertical price restraints and horizontal price fixing—as exemplified by New York’s ABC Law § 101-b.

Vertical price restraints involve attempts by manufacturers to set the prices at which their distributors will resell the manufacturers’ goods to consumers.⁵⁶ This type of behavior, also known as resale price maintenance,⁵⁷ falls within the category of “agreements in restraint of trade.”⁵⁸ Post and hold laws have typically been treated as horizontal price fixing, which generally involves an agreement among competitors to increase, set, or maintain prices.⁵⁹ Section 1 of the Sherman Act provides that agreements in restraint of trade are illegal, and thus actors who violate Section 1 may be subject to criminal prosecution.⁶⁰ The key concept in Section 1 is concerted action, or “agreement,” because without collective action there can be no violation of the provision, no matter how anticompetitive an individual’s conduct.⁶¹

The Supreme Court has interpreted Section 1 to apply only to restraints of trade that are unreasonable,⁶² and has developed two categories of such unreasonable restraints.⁶³ First, there are restraints that are deemed unreasonable *per se*, and accordingly, these are *per se* violations of Section 1.⁶⁴

⁵⁶ See SHENEFIELD & STELZER, *supra* note 45, at 65. For a simple example: Company A manufactures whiskey and sells it to wholesalers, such as Costco or Sam’s Club, but only on the stipulation that they will sell the whiskey to their consumers at \$30 per bottle. This agreement is a vertical price restraint because Company A, as an upstream, or vertical, seller, is setting prices for a downstream seller rather than allowing market forces (supply and demand) to control the resale price of the whiskey. The same analysis would apply to a wholesaler who imposed a similar condition on downstream retail liquor stores that purchase the whiskey for future resale to individual consumers.

⁵⁷ See *id.* at 66.

⁵⁸ SHENEFIELD & STELZER, *supra* note 45, at 15; see also 15 U.S.C. § 1.

⁵⁹ See *infra* Part III.B.; see also Dep’t of Justice, U.S. Attorneys, Antitrust Resource Manual, Identifying Sherman Act Violations, available at http://www.justice.gov/usao/eousa/foia_reading_room/usam/title7/ant00008.htm (last visited Jan. 4, 2010).

⁶⁰ 15 U.S.C. § 1 (“Every contract, combination . . . or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations is declared to be illegal. Every person who shall [engage in the prohibited activity] shall be deemed guilty of a felony.”).

⁶¹ 15 U.S.C. § 1; see also Lopatka & Page, *supra* note 54, at 273; SHENEFIELD & STELZER, *supra* note 45, at 15.

⁶² Standard Oil Co. of New Jersey v. U.S., 221 U.S. 1, 60-62 (1911).

⁶³ See SHENEFIELD & STELZER, *supra* note 45, at 15-16.

⁶⁴ See *id.* at 16; see also N. Pac. Ry. v. U.S., 356 U.S. 1, 5 (1958) (in which the Supreme Court states that “there are certain agreements or practices which because of their pernicious effect on competition and lack of any redeeming virtue are conclusively presumed to be unreasonable.”).

Per se unreasonable restraints include price-fixing.⁶⁵ The second category of unreasonable restraints of trade consists of restraints that are assessed under the “rule of reason.”⁶⁶ In addition to requiring an agreement to unreasonably restrain trade, the Sherman Act also requires that the wrongful conduct (i.e., anticompetitive practices) result in “competitive injury.”⁶⁷ Competitive injury includes artificially high prices, limited output of goods or services, or exclusion of competitors.⁶⁸ In summary, a vertical price restraint that violates the Sherman Act is one that involves an agreement to restrain trade (by setting/controlling prices irrespective of market forces) that causes competitive injury. New York’s price posting scheme, of which Section 101-b is an integral part, contemplates just such a prohibited restraint.⁶⁹

C. *A Brief Overview of New York’s Price Posting Scheme*

New York maintains a “three tier” alcohol distribution system.⁷⁰ This means that, with the exception of direct shipping in wine,⁷¹ a manufacturer must sell alcohol to New York wholesalers, who in turn sell to retailers, who then sell to

⁶⁵ See ADAMS & GRAY, *supra* note 52, at 164; see *infra* notes 252-253 and accompanying text.

⁶⁶ SHENEFIELD & STELZER, *supra* note 45, at 16. In general, antitrust law is not always black-and-white, and many activities are examined by the courts to determine “whether, on balance, the conduct is procompetitive or anticompetitive.” *Id.*

⁶⁷ *Id.* at 32 (“[T]he basic inquiry concerns competitive injury . . .”). Market participants can cause competitive injury without violating the Sherman Act, however. For example, a firm with a monopoly in a market is not necessarily violating the Sherman Act despite the fact that its conduct decreases competition by excluding competitors, as long as the firm did not achieve its monopoly status by entering into agreements with other firms to establish their respective market positions, as opposed to individually competing for consumers. See *id.* at 36 (“[P]ure, lawfully attained monopoly is not prohibited.”).

⁶⁸ *Id.* at 32.

⁶⁹ See *infra* Part IV.

⁷⁰ See *Arnold’s Wines, Inc. v. Boyle*, 571 F.3d 185, 187-88 (2d Cir. 2009) (describing New York’s three-tier regulatory system); see also FTC, POSSIBLE ANTICOMPETITIVE BARRIERS TO E-COMMERCE: WINE 5-7 (July 2003) (hereinafter FTC REPORT), available at <http://www.ftc.gov/os/2003/07/winereport2.pdf> (last visited Jan. 4, 2010).

⁷¹ New York permits both out-of-state and in-state wineries to ship directly to consumers. N.Y. ALCO. BEV. CONT. LAW §§ 79-c, -d. For the interested reader, direct shipping of wine is also a controversial issue. The Supreme Court recently discussed the issue in *Granholm v. Heald*, 544 U.S. 460 (2005), and there are numerous scholarly articles available for more information. See e.g. Elizabeth Norton, *The Twenty-first Amendment in the Twenty-first Century: Reconsidering State Liquor Controls in Light of Granholm v. Heald*, 67 OHIO ST. L. J. 1465, 1471 (2006); Silvernail, *supra* note 3.

consumers.⁷² The New York State Liquor Authority (the “SLA”) is responsible for enforcing New York’s Alcohol Beverage Control Law, which is a complex set of laws that generally prohibits deviation from the “three tier” system.⁷³

This Note focuses on the “price posting” statute within the ABC Law, specifically whether compliance with the law is a violation of the Sherman Act. New York’s price posting scheme is found in Section 101-b of the ABC Law.⁷⁴ Within Section 101-b, there are requirements that both manufacturers and wholesalers file a monthly posting with the SLA that lists their products’ prices for the following pricing period.⁷⁵ After the prices are filed, the SLA produces a composite for inspection, and there is a three-day window in which wholesalers may lower their prices to the lowest posted prices for the same products.⁷⁶ After this window ends, the prices cannot be changed for the entire month without prior written permission from the SLA.⁷⁷

In light of current Twenty-first Amendment jurisprudence, Section 101-b’s mandate that prices must not be changed without the SLA’s permission (as opposed to being dependent on market forces) constitutes an unreasonable restraint of trade, thereby violating the Sherman Act.⁷⁸ The Twenty-first Amendment may serve as a defense when a liquor law is challenged as preempted by the Sherman Act.⁷⁹ Therefore, the problems of Section 101-b should not be addressed without considering the Supreme Court’s approach to the Twenty-first Amendment and the regulatory powers it gave to the states regarding alcohol. This is especially true because of conflicting language in Supreme Court jurisprudence.⁸⁰

⁷² See *Arnold’s Wines*, 571 F.3d at 187-88 (citing various provisions of New York’s Alcohol Beverage Control Law); see also FTC REPORT, *supra* note 70, at 5-7.

⁷³ See *Arnold’s Wines*, 571 F.3d at 187 n.1 (“With the exception of wineries, . . . all manufacturers’ products must pass through the three-tier system.”) (internal citations omitted).

⁷⁴ See generally N.Y. ALCO. BEV. CONT. LAW § 101-b.

⁷⁵ *Id.* § 101-b(3)(a).

⁷⁶ *Id.* § 101-b(4). The SLA also has the option to simply produce all filed price schedules for inspection, rather than creating a composite of them. *Id.*

⁷⁷ *Id.* § 101-b(3)(b).

⁷⁸ See *supra* Part I.B.

⁷⁹ See *infra* note 120 and accompanying text.

⁸⁰ See *infra* notes 89-95, 118 and accompanying text.

II. EVOLUTION OF TWENTY-FIRST AMENDMENT JURISPRUDENCE

Over time, the interpretation of the scope of power that Section 2 of the Twenty-first Amendment gives the states has varied, especially regarding the “interplay between the Commerce Clause and the Twenty-first Amendment.”⁸¹ Two main approaches to interpreting Section 2 have developed: the “absolutist” approach and the “federalist” approach.⁸² Absolutists argue that the “plain language of the Twenty-first Amendment vests complete control of regulation over intoxicating liquor to the states.”⁸³ Federalists, on the other hand, stress that the Twenty-first Amendment “does not vest in the states any new powers, but merely restores the status quo that existed prior to Prohibition.”⁸⁴ The Supreme Court has evolved from an absolutist stance, which was highly deferential to state liquor laws at the expense of other federal laws, to an approach closer to the federalist view. Today, the Court examines these liquor laws in relation to pertinent federal laws.⁸⁵

A. *Policy of Non-Interference*

The Twenty-first Amendment incited controversy not only from Prohibitionists, but also from those who believed that it conflicted with Congress’ power to regulate interstate commerce.⁸⁶ Beginning in the late 1930s, the Supreme Court made it clear that it would take a deferential approach to state laws that invoked the Twenty-first Amendment to regulate

⁸¹ Duncan Baird Douglass, *Constitutional Crossroads: Reconciling the Twenty-first Amendment and the Commerce Clause to Evaluate State Regulation of Interstate Commerce in Alcoholic Beverages*, 49 DUKE L.J. 1619, 1636-37 (2000). Douglass refers often to the intersection of the Twenty-first Amendment and the dormant Commerce Clause. *See generally id.* at 1624-38. However, as he points out, the dormant Commerce Clause is a negative implication of the affirmative powers that the Commerce Clause grants the states; they are therefore intertwined. *Id.* at 1624. For clarity and consistency, this note will refer to the Commerce Clause or commerce powers, rather than the dormant Commerce Clause.

⁸² Silvernail, *supra* note 3, at 513 (internal quotation marks omitted).

⁸³ *Id.*

⁸⁴ *Id.*

⁸⁵ Douglass, *supra* note 81, at 1636-37.

⁸⁶ *See Norton, supra* note 71, at 1471; *compare* U.S. CONST. art. I, § 8, cl. 3 (“Congress shall have Power . . . To regulate Commerce . . . among the several States”) with U.S. CONST. amend. XXI, § 2 (“The transportation or importation into any State, Territory, or possession of the United States for delivery or use therein of intoxicating liquors, *in violation of the laws thereof*, is hereby prohibited” (emphasis added)).

alcoholic beverages.⁸⁷ In a series of cases following the Amendment's ratification, the Supreme Court solidly established its absolutist approach to interpretation, as it repeatedly found that state liquor laws were not constrained by other provisions of the Constitution.⁸⁸

The Supreme Court's most extreme language regarding the extent of states' power to control liquor can be found in *Ziffrin, Inc. v. Reeves*.⁸⁹ In *Ziffrin*, an Indiana corporation contracted with Kentucky distillers to receive whiskey and then ship it to Chicago.⁹⁰ When Kentucky enacted a law prohibiting this type of arrangement and provided law enforcement with the authority to seize goods,⁹¹ the Indiana corporation claimed that the law was unconstitutional because it violated the Commerce Clause.⁹² In upholding the law, the Supreme Court explicitly held in favor of state regulation of liquor, stating "[t]he Twenty-first Amendment sanctions the right of a state to legislate concerning intoxicating liquors brought from without, *unfettered by the Commerce Clause.*" (emphasis added)⁹³ With limited exceptions,⁹⁴ the Court only

⁸⁷ See Norton, *supra* note 71, at 1471.

⁸⁸ See e.g., State Bd. of Equalization v. Young's Market, 299 U.S. 59 (1936) (holding that state liquor laws are not limited by the Commerce or Equal Protection Clauses), *adhered to by* Ziffrin v. Reeves, 308 U.S. 132 (1939); Indianapolis Brewing Co. v. Liquor Control Comm'n, 305 U.S. 391 (1939); Mahoney v. Joseph Triner Corp., 304 U.S. 401 (1938). Accord Douglass, *supra* note 81, at 1637-38; Spaeth, *supra* note 1, at 183-84.

⁸⁹ 308 U.S. 132 (1939).

⁹⁰ *Id.* at 133.

⁹¹ *Id.* at 133-35.

⁹² *Id.* at 137. The Indiana corporation also argued the law violated the Due Process and Equal Protection Clauses. *Id.*

⁹³ *Id.* at 138 (emphasis added).

⁹⁴ See Silvernail, *supra* note 3, at 517-19. Silvernail uses two cases that came about a decade after *Young's Market* to illustrate his theory that the Court began to "introduc[e] chips into the foundation upon which the absolutist interpretation of the Twenty-first Amendment is constructed." *Id.* at 519. First, Silvernail points to *United States v. Frankfort Distilleries*, 324 U.S. 293 (1945) as "the first case where the Supreme Court showed signs of reining in the broad sweeping powers it so readily bestowed upon the states in *Young's Market.*" Silvernail, *supra* note 3, at 518. *Frankfort Distilleries* was the first time the Court stated that the powers given to the states by the Twenty-first Amendment are qualified by federal powers. *Id.*; see also *Frankfort Distilleries*, 324 U.S. 293. Although the Court refused to ultimately decide whether the Sherman Act limits state powers enacted under the Twenty-first Amendment in an antitrust suit against Colorado liquor producers, wholesalers, and retailers, the Court stated:

Granting the state's full authority to determine the conditions upon which liquor can come into its territory and what will be done with it after it gets there, it does not follow from that fact that the United States is wholly without power to regulate the conduct of those who engage in interstate trade outside the jurisdiction of [the state whose law is at issue].

began to “retreat[] substantially” from this approach to state liquor laws in the 1960s and 1970s.⁹⁵

B. *Limiting the Scope of Section 2*

Upon brief review of the Supreme Court’s early Twenty-first Amendment jurisprudence, one would think that the Supreme Court had forgotten the importance of the Commerce Clause.⁹⁶ However, in the landmark case⁹⁷ of *Hostetter v. Idlewild Bon Voyage Liquor Corp.*,⁹⁸ the Supreme Court finally used the Commerce Clause to strike down a New York liquor law that prohibited transportation of alcohol within state borders, because the shipments at issue were merely passing through New York for delivery and use in a foreign country.⁹⁹ The Court emphasized that New York was not trying to prevent alcohol from being unlawfully diverted for use within the state,¹⁰⁰ perhaps indicating that the law would have been permissible had that been New York’s goal. In reaching its holding, the Court explained that “[b]oth the Twenty-first Amendment and the Commerce Clause are parts of the same Constitution. Like other provisions of the Constitution, each must be considered in the light of the other, and in the context of the issues and interests at stake in any concrete case.”¹⁰¹ This

Frankfort Distilleries, 324 U.S. at 299. The second case Silvernail uses to illustrate that the Court was “chip[ping] into” its absolutist foundation was *Nippert v. City of Richmond*, 327 U.S. 416 (1946). Silvernail, *supra* note 3, at 518-19. In *Nippert*, which was not a “Twenty-first Amendment [case],” the Supreme Court reiterated the strength of the federal government’s commerce powers, citing *Frankfort Distilleries* for its proposition that “even the commerce in intoxicating liquors, over which the Twenty-first Amendment gives the states the highest degree of control, is not altogether beyond the reach of the federal commerce power. . .” *Nippert* at 425 n.15.

⁹⁵ See Douglass, *supra* note 81, at 1638; see also Norton, *supra* note 71, at 1472.

⁹⁶ See *supra* note 93 and accompanying text.

⁹⁷ See Spaeth, *supra* note 1, at 185 (“The Court consummated its full retreat from earlier broad readings of [T]wenty-first [A]mendment power, in a pair of decisions handed down in 1964: *Hostetter v. Idlewild Bon Voyage Liquor Corp.* and *Department of Revenue v. James B. Beam Distilling Co.*”).

⁹⁸ 377 U.S. 324 (1964).

⁹⁹ *Id.* at 333-34.

¹⁰⁰ *Id.* at 333.

¹⁰¹ *Id.* at 332. The Supreme Court also uttered the oft-repeated comment that

To draw a conclusion from [early Twenty-first Amendment jurisprudence after ratification] that the Twenty-first Amendment has somehow operated to “repeal” the Commerce Clause wherever regulation of intoxicating liquors is concerned would, however, be an absurd oversimplification. . . . Such a conclusion would be patently bizarre and is demonstrably incorrect.

Id. at 331-32.

decision marked a notable shift in the Supreme Court's approach to interpreting Section 2 of the Twenty-first Amendment. For the first time, the Court clearly asserted that the Commerce Clause could limit a state's liquor laws.¹⁰²

After *Hostetter*, the Court continued its retreat from giving the states excessive discretion with respect to their liquor laws. For example, less than a decade later, the Court made clear that it would no longer give deference to state liquor laws at the expense of the Fourteenth Amendment.¹⁰³ These decisions paved the way for the Court to further restrict the over-broad scope it had originally given to Section 2.

C. *The Current Approach to Twenty-first Amendment Cases: The "Accommodation Doctrine"*¹⁰⁴

Beginning in 1980, the Supreme Court began clearly articulating its new approach to overreaching state liquor regulations. In *California Retail Liquor Dealers Ass'n v. Midcal Aluminum, Inc.*,¹⁰⁵ the Court held that Section 2 of the Twenty-first Amendment did not save a California liquor law that violated the Sherman Act by imposing a resale price maintenance scheme on wholesale wine producers.¹⁰⁶ The Court claimed that it was following early Twenty-first Amendment jurisprudence by acknowledging the extensive authority that the Amendment gave to the states to regulate liquor.¹⁰⁷

¹⁰² *Id.*; see also Douglass, *supra* note 81, at 1638.

¹⁰³ See *Craig v. Boren*, 429 U.S. 190, 210 (1976) (acknowledging again the states' broad powers over liquor under the Twenty-first Amendment but refusing to uphold a liquor law that violated the Equal Protection Clause of the Fourteenth Amendment); *Wisconsin v. Constantineau*, 400 U.S. 433, 436-37 (1971) (holding that while the states were given a broad grant of power to regulate liquor by the Twenty-first Amendment, a liquor law could not deprive a person of due process). The Court in *Craig* reaffirmed that "each provision [of the Constitution must] 'be considered in the light of the other . . .'" *Id.* at 206 (quoting *Hostetter*, 377 U.S. at 332); see also Silvernail, *supra* note 3, at 521.

¹⁰⁴ The Court's current approach of balancing state and federal interests when a state liquor regulation conflicts with a federal law that implicates the Commerce Power has come to be known as the "accommodation doctrine." See Silvernail, *supra* note 3, at 524 (internal quotation marks omitted); see also Elizabeth D. Lauzon, Annotation, *Interplay Between Twenty-first Amendment and Commerce Clause Concerning State Regulation of Intoxicating Liquors*, 116 A.L.R.5th 149 (2004).

¹⁰⁵ 445 U.S. 97 (1980). *Midcal* holds particular significance for this Note, as it provides the foundation for how to analyze whether a liquor law violates the Sherman Act and, if so, whether it is protected by the Twenty-first Amendment; therefore the case is relied upon by almost all of the subsequent cases on this issue. See *infra* Part III.

¹⁰⁶ *Midcal*, 445 U.S. at 113-14.

¹⁰⁷ *Id.* at 106-10.

Nonetheless, in stark contrast with those early cases, the Court refused to defer to a state law because it conflicted with a federal law enacted pursuant to the Commerce Power.¹⁰⁸ In reaching its decision, the Court relied on *Hostetter*'s suggestion that examining state liquor laws may call for balancing the state's interests with federal interests, but it went a step further by actually requiring this balancing approach "in appropriate situations."¹⁰⁹

Taking an even larger step away from the early Twenty-first Amendment jurisprudence, in *Capital Cities Cable, Inc. v. Crisp*,¹¹⁰ the Court expanded on both *Hostetter* and *Midcal* to set a new standard for state liquor regulations that conflict with federal laws.¹¹¹ *Capital Cities* stands for the proposition that courts presented with such a regulation must ask "whether the interests implicated by a state regulation are so closely related to the powers reserved by the Twenty-first Amendment that the regulation may prevail, notwithstanding that its requirements directly conflict with express federal policies."¹¹² In other words, the test to determine whether a state liquor law that conflicts with a federal law is valid is to ask whether the state law at issue directly serves the purposes of the Twenty-first Amendment,¹¹³ and whether those interests outweigh the interests of the countervailing federal law.¹¹⁴

In another notable opinion, *Bacchus Imports, Ltd. v. Dias*,¹¹⁵ not only did the Court affirm this *Capital Cities* standard, but it was quite dismissive of the earlier Twenty-first Amendment cases, referring to the legislative history of the Twenty-first Amendment as "obscur[e]."¹¹⁶ Most shockingly, the Court declared, "[i]t is by now clear that the [Twenty-first]

¹⁰⁸ *Id.*

¹⁰⁹ *Id.* at 110 ("[T]here is no bright line between federal and state powers over liquor. . . . Although States retain substantial discretion to establish other liquor regulations, those controls may be subject to the federal commerce power in appropriate situations. The competing state and federal interests can be reconciled only after careful scrutiny of those concerns in a 'concrete case.'").

¹¹⁰ 467 U.S. 691 (1984).

¹¹¹ *Id.* at 711-14.

¹¹² *Id.* at 714.

¹¹³ *Id.*

¹¹⁴ See *Midcal*, 445 U.S. at 110; see *supra* notes 109, 112 and accompanying text.

¹¹⁵ 468 U.S. 263 (1984).

¹¹⁶ See *id.* at 274 ("Despite broad language in some of the opinions of this Court written shortly after ratification of the Amendment, more recently we have recognized the obscurity of the legislative history of § 2." (internal citations omitted)); see also Silvernail, *supra* note 3, at 525.

Amendment did not entirely remove state regulation of alcoholic beverages from the ambit of the Commerce Clause.”¹¹⁷ Such a statement directly contradicted the Court’s language in *Ziffrin* that state liquor laws are “unfettered by the Commerce Clause.”¹¹⁸ After *Bacchus Imports*, there could be no doubt that under the accommodation doctrine the Supreme Court would take a hard look at state liquor regulations that conflicted with federal laws.¹¹⁹ Still, knowing that a court will give rigorous scrutiny to New York’s post and hold statute is barely scratching the surface of the type of analysis required to determine the validity of Section 101-b.

III. THE TWENTY-FIRST AMENDMENT VS. THE SHERMAN ACT: WHEN A LIQUOR LAW IS SUBJECT TO AN ANTITRUST CHALLENGE

When a state’s liquor law is challenged on constitutional grounds, one of the most common reactions of that state is to use the Twenty-first Amendment as a defense.¹²⁰ The same is true when a liquor law is challenged as being in violation of the Sherman Act.¹²¹ The Supreme Court will no longer give great deference to liquor laws that conflict with federal legislation, such as the Sherman Act, simply because the laws are claimed to have been enacted pursuant to the Twenty-first Amendment.¹²² As a result, courts faced with determining the validity of such a law must perform an analysis that involves wading through complex issues of antitrust law and assessing the legitimacy of states’ claimed interests in order to ultimately decide whether a state has proven that it can properly rely on the Twenty-first Amendment to shield a state law that conflicts with federal law.

¹¹⁷ *Bacchus Imports*, 468 U.S. at 275.

¹¹⁸ *Ziffrin, Inc. v. Reeves*, 308 U.S. 132, 138 (1939); see *supra* Part II.A.

¹¹⁹ See *supra* note 104.

¹²⁰ See *e.g.*, *Cal. Retail Liquor Dealers Ass’n v. Midcal Aluminum, Inc.*, 445 U.S. 97, 106 (1980); see also *Lauzon*, *supra* note 104 (explaining that under the accommodation doctrine, after a court finds that a state liquor regulation violates the Commerce Clause, the burden shifts to the state to show that the law at issue is saved by the Twenty-first Amendment).

¹²¹ See, *e.g.*, *Midcal*, 445 U.S. at 106.

¹²² See *supra* Part II.C.

A. *Assessing the Validity of Liquor Laws Challenged as Violations of the Sherman Act*

Section 101-b is a state liquor law and should be treated as other state liquor laws that have run up against the Sherman Act. The Supreme Court laid the foundation for an intricate three-part sequential test to determine whether a state's liquor regulation may be sustained when challenged as a violation of the Sherman Act.¹²³ First, a court must determine whether the regulation is preempted by the Sherman Act.¹²⁴ If the law does not violate the Sherman Act, then the challenger will clearly lose because the law has antitrust immunity.¹²⁵ However, if the court finds that the liquor regulation does indeed violate the Sherman Act, it must perform the second step of the three-part analysis and determine whether the law has antitrust immunity under the state-action doctrine.¹²⁶ The law will be sustained if the court finds that it has antitrust immunity.¹²⁷ If not, the third step in the analysis is to determine whether Section 2 of the Twenty-first Amendment will serve as a valid defense and save the law.¹²⁸ Each step in this three-part test is broken down further, and while the Supreme Court has yet to give clear guidance for how to apply the test in a given case, the federal courts have developed and applied this test to many state liquor regulations challenged on the grounds of violating the Sherman Act.¹²⁹

¹²³ *Midcal*, 445 U.S. at 102-06.

¹²⁴ *Id.* at 102. In striking down a California liquor regulation, the Supreme Court noted “[t]he threshold question is whether [the liquor regulation] . . . violates the Sherman Act.” *Id.*

¹²⁵ *Id.* at 102-03. For example, a unilateral restraint is not preempted by the Sherman Act, or in other words, has antitrust immunity. *See infra* Part III.A.1.

¹²⁶ *Midcal*, 445 U.S. at 103 (after finding that California's wine pricing scheme violated the Sherman Act, the Court then considered whether it was immune); *Parker v. Brown*, 317 U.S. 341, 368 (1943). The state-action doctrine is essentially a two-part test to determine whether the challenged liquor law should be treated as “state action,” and thus immune from the Sherman Act, despite the fact that private actors are involved in the law's enforcement. *See infra* Part III.A.2. The doctrine is intended to address the tension between serving the federal interests of the Sherman Act, e.g. promoting competition, and the rights of the states as sovereign entities. *See infra* Part III.A.2.

¹²⁷ *See supra* note 123 and accompanying text.

¹²⁸ *Midcal*, 445 U.S. at 106 (turning to an analysis of whether the Twenty-first Amendment served as a basis for upholding the challenged law after finding the law conflicted with the Sherman Act and had no antitrust immunity).

¹²⁹ *See infra* Part III.B.

1. Step One: Whether a State Liquor Law Is Preempted by the Sherman Act

As a threshold issue, a court must look at the challenged law to determine whether it conflicts with the Sherman Act.¹³⁰ Even this threshold issue is complex, requiring its own sequential two-step analysis.¹³¹ Assume that the challenged law is a restraint of trade, in that it hampers free competition.¹³² As discussed above, the Sherman Act prohibits agreements to engage in unreasonable restraints of trade.¹³³ Therefore, a court must determine two things: (1) whether the required element of agreement has been met; and (2) whether the restraint is unreasonable. Whether the first element, a finding of agreement, will be satisfied largely depends on whether the challenged law may be classified as a unilateral¹³⁴ or a hybrid¹³⁵ restraint.¹³⁶ If the restraint is deemed unilateral, then the law has antitrust immunity because it is a sovereign act by the state that the Sherman Act was not intended to prohibit.¹³⁷ If, however, the law is deemed a hybrid restraint, the court will proceed to the second step in the analysis: determining whether the restraint actually violates the Sherman Act.¹³⁸ Thus, applying the rule to Section 101-b, the two-step test for whether it conflicts with the Sherman Act consists of asking (1)

¹³⁰ *Midcal* 445 U.S. at 102.

¹³¹ *See TFWS, Inc. v. Schaefer*, 242 F.3d 198, 207 (4th Cir. 2001).

¹³² *See supra* note 54 (defining restraint of trade).

¹³³ *See supra* Part I.B.

¹³⁴ A unilateral restraint is typically a state law, or governmental action, that forces private individuals to engage in anticompetitive behavior simply by complying with the law; there is no agreement among the individuals. *See infra* Part.III.A.1.a; *see also* *Fisher v. City of Berkeley*, 475 U.S. 260, 266-67 (1986) (finding a rent control ordinance to be a unilateral restraint and noting that the landlords whose prices were restricted by the ordinance had made no agreement to put a ceiling on rent prices); *Lopatka & Page, supra* note 54, at 273; A unilateral restraint is in direct contrast with a private restraint, in which private individuals agree to restrain trade, and there is no related governmental regulation shaping their behavior. *See id.* at 284-85.

¹³⁵ A hybrid restraint is not easily defined, but as a general matter involves a state regulation in which private individuals have some discretion as to whether they will comply with the regulation in a way that consists of anticompetitive behavior that would violate the Sherman Act, if not immune. *See infra* Part.III.A.1.a. In other words, hybrid restraints involve a mixture of government and private action. *See Lopatka & Page, supra* note 54, at 287.

¹³⁶ *See TFWS, Inc.*, 242 F.3d at 207 (describing this step in the analysis); *see also Lopatka & Page, supra* note 54, at 284-85.

¹³⁷ *See Parker v. Brown*, 317 U.S. 341, 352 (1943). The Sherman Act was intended to sanction private parties that agree to restrain trade, not to prevent states from taking affirmative action to regulate commerce. *Id.*

¹³⁸ *See TFWS, Inc.*, 242 F.3d at 207.

whether Section 101-b is a unilateral or hybrid restraint in order to find whether the element of agreement has been satisfied; and (2) if Section 101-b is a hybrid restraint, whether it constitutes an unreasonable restraint of trade.

a. Step 1(a): Is There an Agreement?

The Sherman Act expresses the concept of agreement as a “contract, combination . . . or conspiracy.”¹³⁹ However, as scholars have noted, “the meaning of agreement is . . . notoriously complex.”¹⁴⁰ Indeed, it has been argued that the term’s meaning varies depending on whether a restraint is characterized as unilateral or hybrid.¹⁴¹ Since unilateral restraints are automatically immune from preemption by the Sherman Act,¹⁴² the determination of whether a restraint is unilateral or hybrid may effectively result in the invalidation of a law.¹⁴³ Therefore, attempting to draw a line between the two categories is critical.

The distinction between unilateral and hybrid restraints of trade is not always clearly articulated by the courts.¹⁴⁴ What can be gleaned from the cases is that the less discretion private individuals have in affecting competition by complying with the law, the more likely it is that a court will find the law to be a unilateral restraint.¹⁴⁵ In contrast, the more discretion private market participants are given by the law, the more likely it is that a court will deem the law a hybrid restraint.¹⁴⁶ In other words, the restraint’s classification turns on the issue of control.

¹³⁹ 15 U.S.C. § 1; Lopatka & Page, *supra* note 54, at 271.

¹⁴⁰ Lopatka & Page, *supra* note 54, at 288 (“[Agreement] is a term of art whose peculiar contours vary with the Court’s understanding of a particular restraint’s likely competitive effects.”).

¹⁴¹ *See id.* at 297 (“[T]he definition of agreement in the context of hybrid restraints differs from the definition of agreement in the contexts of private restraints and purely governmental restraints.”).

¹⁴² *See supra* note 137 and accompanying text.

¹⁴³ Lopatka & Page, *supra* note 54, at 272.

¹⁴⁴ *See id.* at 269 (stating “the Supreme Court’s precedents are not entirely consistent” in distinguishing between unilateral and hybrid restraints).

¹⁴⁵ *Id.* at 283-84.

¹⁴⁶ Looking forward briefly, a hybrid restraint of trade that conflicts with the Sherman Act may fail to qualify as immune under the state action doctrine. There are several steps before declaring that a challenged law is not immune. However, a law deemed to be a unilateral restraint escapes the lengthy scrutiny that a hybrid restraint will receive, especially if that hybrid restraint is deemed a per se violation of the Sherman Act. *See infra* notes 167-71 and accompanying text.

While unilateral restraints take away private control over competitive decision-making, hybrid restraints allow private individuals to retain at least some degree of control. For example, in *Fisher v. City of Berkeley*, the Supreme Court declared that Berkeley's Rent Control Ordinance setting maximum rent prices that landlords could charge was a unilateral restraint because the Ordinance removed price-setting control from the landlords and gave it to the City.¹⁴⁷ The Court went on to characterize hybrid restraints as using "nonmarket mechanisms [to] merely enforce private marketing decisions," stating that "the regulatory scheme may be attacked" when "private actors" are given "a degree of private regulatory power."¹⁴⁸ The *Fisher* Court took the concept of "hybrid restraints" from a concurrence in an earlier case;¹⁴⁹ the Court then used two cases to illustrate the concept.¹⁵⁰ Although the opinions in those cases did not reference hybrid restraints, they nonetheless serve as guidance since the Court has clearly pointed to them (albeit ex post) as examples of hybrid restraints.

First, the Court in *Fisher* pointed to *Schwegmann Bros. v. Calvert Distillers Corp.*,¹⁵¹ in which the Court had struck down a Louisiana statute authorizing distributors to enter resale price contracts with retailers selling their products and to enforce those price-fixing agreements against not only those retailers, but other retailers selling the distributors' products who were not party to the price-fixing agreement.¹⁵² The Court

¹⁴⁷ *Fisher v. City of Berkeley*, 475 U.S. 260, 269 (1986) (stating the Ordinance "place[d] complete control over maximum rent levels exclusively in the hands of the Rent Stabilization Board. Not just the controls themselves but also the rent ceilings they mandate have been unilaterally imposed on the landlords by the city."). Similarly, the First Circuit compared a Massachusetts law limiting liquor storeowners to a maximum of three liquor store licenses to the Rent Control Ordinance in *Fisher*, because it did not give any control over competitive decision-making to private individuals. See *Mass. Food Ass'n v. Mass. Alcoholic Beverages Control Comm'n*, 197 F.3d 560, 565-66 (1st Cir. 1999) (in finding that the law at issue was not preempted by the Sherman Act, the court stated "[t]he Massachusetts statute . . . does not authorize or direct any private agreements or permit any competitor to determine the price or location of another. . . . As in *Fisher*, the restrictions have been 'unilaterally imposed by government . . . to the exclusion of private control.'" (quoting *Fisher*, 475 U.S. at 266)).

¹⁴⁸ *Fisher*, 475 U.S. at 267-68 (internal quotation marks omitted).

¹⁴⁹ *Id.* (citing *Rice v. Norman Williams*, 458 U.S. 654, 665 (1982) (Stevens, J., concurring)).

¹⁵⁰ See *Fisher*, 475 U.S. at 268.

¹⁵¹ 341 U.S. 384 (1951).

¹⁵² *Schwegmann Bros. v. Calvert Distillers Corp.*, 341 U.S. 384, 395 (1951). At the time this Louisiana statute was in effect, the Miller-Tydings Act amended Section 1 of the Sherman Act to allow agreements setting minimum resale prices for certain commodities in intrastate transactions; that Act is now repealed. 15 U.S.C. § 1,

in *Fisher* distinguished the Rent Control Ordinance, pursuant to which Berkeley set maximum rents, from the Louisiana law that allowed the distributors to set minimum resale prices, thus leaving some control in the hands of private individuals.¹⁵³ Next, the *Fisher* Court referred to *Midcal*.¹⁵⁴ According to the *Fisher* Court, the resale price maintenance scheme in *Midcal* was a hybrid restraint because “[t]he trade restraint condemned in *Midcal* entailed a similar degree of free participation by private economic actors.”¹⁵⁵

Even if *Schwegmann Bros.* and *Midcal* were not decided based on the concept of “hybrid restraints,” the Supreme Court has come to rely on them as examples of the concept, stressing that each case turned on the fact that private individuals had discretion to affect competition, and their decisions would be enforced by the state.¹⁵⁶ In particular, both *Schwegmann Bros.* and *Midcal* dealt with price restraints, and the *Fisher* Court stressed the importance of the fact that the private market participants were the ones setting prices rather than the state, which merely enforced them.¹⁵⁷ This seems to indicate that where a law authorizes price restraints, and the state does not set the prices itself as it did in *Fisher*, it is especially prone to being characterized as a hybrid restraint. Assuming that the challenged law is found to be a hybrid restraint, the element of agreement has been satisfied and the next step in the analysis is to assess whether the restraint is a violation of the Sherman Act.¹⁵⁸

b. Step 1(b): Is There a Violation of the Sherman Act?

As noted above, the Sherman Act only prohibits *unreasonable* restraints of trade.¹⁵⁹ Of course, a state would

Amendments (showing enactment of 50 Stat. 693 (1937), and its repeal, 89 Stat. 801 (1975)).

¹⁵³ See *Fisher*, 475 U.S. at 268-69.

¹⁵⁴ See *supra* notes 105-109 and accompanying text.

¹⁵⁵ *Fisher*, 475 U.S. at 268.

¹⁵⁶ See, e.g., *id.*

¹⁵⁷ *Id.* at 268-69.

¹⁵⁸ See *supra* note 138 and accompanying text. It may seem counterintuitive, if not unjust, to find an illegal agreement under Section 1 of the Sherman Act where individuals are complying with a state liquor law *without actually agreeing to restrain trade*. However, the Supreme Court has affirmatively stated that a violation of the Sherman Act may be found *in the absence of private agreement* if the state compels activity that would otherwise be a per se violation. *324 Liquor Corp. v. Duffy*, 479 U.S. 335, 345 (1987).

¹⁵⁹ See *supra* Part I.B.

prefer that a law be deemed a unilateral restraint and therefore not subject to Sherman Act preemption.¹⁶⁰ Even if the restraint is hybrid, though, the state will still have the opportunity to show that the restraint is not unreasonable.¹⁶¹ Post and hold statutes, like other price restraints, have generally been treated as per se violations.¹⁶² A state statute is a per se violation of the Sherman Act if it “mandates or authorizes conduct that necessarily constitutes a violation of the antitrust laws in all cases, or if it places irresistible pressure on a private party to violate the antitrust laws in order to comply with the statute.”¹⁶³ If, as in the cases involving “post and hold” statutes, the restraint does violate the Sherman Act,¹⁶⁴ the court must then determine whether the law is immune under the state-action doctrine espoused by *Parker v. Brown*.¹⁶⁵ Considering the treatment of other post and hold liquor laws,¹⁶⁶ Section 101-b should certainly reach this level of analysis.

2. Step Two: State-Action Doctrine

Assuming that Section 101-b of the ABC Law is deemed a hybrid price restraint that violates the Sherman Act, New York would certainly argue that it is entitled to antitrust immunity under the state-action doctrine. The state-action doctrine originated in *Parker*, where the Supreme Court pointed out that the Sherman Act was not intended to prohibit the states from taking affirmative action that restrains competition.¹⁶⁷ Unilateral restraints are not subject to the scrutiny that hybrid restraints receive because unilateral restraints qualify as this type of affirmative state action.¹⁶⁸ Although hybrid restraints do not automatically qualify for

¹⁶⁰ See *supra* note 137 and accompanying text.

¹⁶¹ See *supra* Part I.B. Determining whether a restraint violates the Sherman Act is typically fact-specific, but it can be superficially described as two basic approaches. While some restraints of trade are per se unreasonable, and thus in conflict with the Sherman Act without further analysis, other restraints are assessed under the “rule of reason,” which essentially consists of determining whether a restraint is reasonable in light of the circumstances. See *id.*

¹⁶² See *infra* Part III.B.

¹⁶³ *Rice v. Norman Williams Co.*, 458 U.S. 654, 661 (1982).

¹⁶⁴ See *infra* Part III.B.

¹⁶⁵ 317 U.S. 341 (1943) (finding California’s regulation of the state’s 1940 raisin crop to be proper regulation of state industry not interfering with federal commerce powers).

¹⁶⁶ See *infra* Part III.B.

¹⁶⁷ *Parker v. Brown*, 317 U.S. 341, 352 (1943).

¹⁶⁸ See *supra* note 137 and accompanying text.

such immunity, the Supreme Court in *Midcal* held that a hybrid restraint will be immune from the Sherman Act under *Parker's* state-action doctrine if it satisfies two requirements.¹⁶⁹ First, the restraint must be “clearly articulated and affirmatively expressed as state policy.”¹⁷⁰ Second, the restraint must be “actively supervised by the State itself.”¹⁷¹

a. Step 2(a): Is the Restraint Clearly Articulated and Affirmatively Expressed as State Policy?

The first hurdle of *Midcal*, i.e., whether the state has “clearly articulated and affirmatively expressed [the restraint] as state policy,” has typically been overcome.¹⁷² This makes sense considering that the state does not have to do much to satisfy the standard. For example, in *Midcal*, the Court said that the California wine-pricing scheme (already deemed a hybrid restraint) satisfied the first prong of the immunity test because “[t]he legislative policy [was] forthrightly stated and clear in its purpose to permit resale price maintenance.”¹⁷³ So long as the legislature makes its intent to displace competition clear, the first prong will most likely be satisfied.¹⁷⁴

b. Step 2(b): Is the Restraint Actively Supervised by the State Itself?

It is the second *Midcal* prong that has more often prevented a challenged restraint from establishing immunity.¹⁷⁵ In *Midcal*, the Supreme Court struck down California’s wine-pricing scheme because it was not actively supervised by the

¹⁶⁹ Cal. Retail Liquor Dealers Ass’n v. Midcal Aluminum, Inc., 445 U.S. 97, 105 (1980).

¹⁷⁰ *Id.* (internal quotation marks omitted).

¹⁷¹ *Id.* (internal quotation marks omitted).

¹⁷² *Id.* (internal quotation marks omitted). See generally *Costco Wholesale Corp. v. Maleng*, 522 F.3d 874, 902 (9th Cir. 2008); *TFWS, Inc. v. Schaefer*, 242 F.3d 198, 210-11 (4th Cir. 2001).

¹⁷³ *Midcal*, 445 U.S. at 105.

¹⁷⁴ But see *Freedom Holdings, Inc. v. Spitzer*, 357 F.3d 205, 230 (2d Cir. 2004) (finding that New York’s Contraband Statutes failed prong one of *Midcal* because the state’s articulated interest in using them as a method of revenue production was not legitimate); *infra* notes 286-288 and accompanying text. The Supreme Court has not spoken on the issue, but if the Second Circuit’s approach is correct, then a clearly stated intent to displace competition may nonetheless fail to satisfy *Midcal*’s first prong if the court finds the policy behind the intent illegitimate.

¹⁷⁵ *Midcal*, 445 U.S. at 105-06; see *Costco Wholesale Corp.* 522 F.3d at 902-03; *TFWS, Inc.*, 242 F.3d at 210-11.

state, stating in a frequently-quoted phrase that “[t]he State neither establishes prices nor reviews the reasonableness of the price schedules . . . [t]he State does not monitor market conditions or engage in any ‘pointed reexamination’ of the program.”¹⁷⁶ In striking down New York ABC Law § 101-bb, a “markup” statute, the Court similarly found a lack of active supervision.¹⁷⁷ Rather, “[t]he State ha[d] displaced competition among liquor retailers without substituting an adequate system of regulation.”¹⁷⁸ The Supreme Court seems to have established two ways for a state to protect an anticompetitive liquor regulation. If the regulation allows private individuals to set prices, i.e., gives private individuals discretion or control over prices, the state must have a system in place to ensure that these prices are reasonable.¹⁷⁹ Alternatively, the consequences of allowing individuals to set prices must be reasonable.¹⁸⁰ If the state is unable to show reasonableness, the restraint will not be immune, and the state must then rely on the Twenty-first Amendment’s grant of power for redemption.

3. Step Three: The Twenty-first Amendment Defense

After a state law is deemed a hybrid restraint and fails to obtain antitrust immunity, the Twenty-first Amendment is the final obstacle to invalidation,¹⁸¹ or in other words, Section 101-b’s last resort for protection. The courts have developed yet another two-part test for determining whether the Twenty-first

¹⁷⁶ *Midcal*, 445 U.S. 97, 105-06 (1980) (after finding the scheme to be a hybrid restraint preempted by the Sherman Act). *See, e.g., Costco Wholesale Corp.*, 522 F.3d at 889 (quoting *Midcal*); *Miller v. Hedlund*, 813 F.2d 1344, 1351 (9th Cir. 1987) (quoting *Midcal*).

¹⁷⁷ New York ABC Law section 101-bb required retailers to “markup” the “posted” wholesale price for liquor by 112 percent (but allowed wholesalers to sell to retailers at less than the “posted” price). *324 Liquor Corp. v. Duffy*, 479 U.S. 335, 335, 337 (1987).

¹⁷⁸ *Id.* at 344-45.

¹⁷⁹ *See, e.g., 324 Liquor*, 479 U.S. at 344-45; *Midcal*, 445 U.S. at 105-06.

¹⁸⁰ The Court has indicated that it is worried about state authorization of private price-fixing that essentially fosters cartelization, with no check on the individual market participants’ power. *See 324 Liquor*, 479 U.S. at 342 (discussing the possibility that “industrywide resale price maintenance . . . may facilitate cartelization”); *Midcal*, 445 U.S. at 106 (“The national policy in favor of competition cannot be thwarted by casting such a gauzy cloak of state involvement over what is essentially a private price-fixing arrangement.”). While the Court expresses its concerns, it does not elaborate as to what kinds of consequences would be considered reasonable, and neither do the federal courts applying the Supreme Court’s jurisprudence in this area.

¹⁸¹ *See Midcal*, 445 U.S. at 106.

Amendment serves as a valid defense to a restraint that violates the Sherman Act.¹⁸² First, the restraint must be intended to serve a legitimate state policy.¹⁸³ Second, the state must show that the restraint “substantiates” that policy.¹⁸⁴ At this point, under the accommodation approach,¹⁸⁵ the Twenty-first Amendment will only protect a challenged restraint if it directly serves the policies of the Amendment and those policies outweigh the goals of the Sherman Act.¹⁸⁶ If the restraint can pass this final test, it will have survived its antitrust challenge.¹⁸⁷

a. Step 3(a): Was the Restraint Intended to Serve a Legitimate State Policy?

Even if Section 101-b purports to serve legitimate state concerns, it will not necessarily be saved by the Twenty-first Amendment. In essence, there must be a balancing of a state’s legitimate interests and the federal interest in the Sherman Act.¹⁸⁸ For example, in *Bacchus Imports*, the Supreme Court refused to uphold a discriminatory state liquor tax because it “was [not] designed to promote temperance or carry out any other purpose of the Twenty-first Amendment.”¹⁸⁹ While *Bacchus* involved a challenge based on discrimination in violation of the Commerce Clause as opposed to an antitrust challenge, it is the clearest statement of the type of reasoning that has been applied in the liquor antitrust cases.¹⁹⁰ Specifically, if a state liquor law does not promote the goals of the Twenty-first Amendment, it will be invalidated.

¹⁸² Both prongs must be satisfied for the restraint to survive an antitrust challenge. *Id.* at 113-14.

¹⁸³ *See id.*

¹⁸⁴ *Id.* The word “substantiate” appears to be used as a term of art by the Supreme Court and the federal courts to assess whether the restraint at issue effectuates the policy asserted in support of it. For consistency, this Note will use this term as well with the same intended meaning.

¹⁸⁵ *See supra* Part II.C.

¹⁸⁶ *See supra* note 114 and accompanying text. The goals of the Sherman Act are fairness among producers, economic efficiency, and consumer welfare. *See supra* note 53 and accompanying text.

¹⁸⁷ *See supra* notes 123-128 and accompanying text.

¹⁸⁸ *See, e.g., Midcal*, 445 U.S. at 108-14.

¹⁸⁹ *Bacchus Imports, Ltd. v. Dias*, 468 U.S. 263, 276 (1984); *see Douglass, supra* note 81, at 1641-42.

¹⁹⁰ *See, e.g., 324 Liquor Corp. v. Duffy*, 479 U.S. 335, 347-49 (1987); *Midcal*, 445 U.S. at 113-14.

Therefore, in order to serve a legitimate state policy, the restraint must be based on a concern that relates directly to the Twenty-first Amendment, *and* that concern must outweigh the federal interests served by the Sherman Act.¹⁹¹ So far, legitimate state interests have included temperance and protecting small retailers.¹⁹² Economic protectionism is an example of a state interest that is *not* legitimate because it is not a “core concern[]” of the Twenty-first Amendment.¹⁹³ The fact that a restraint purports to serve legitimate state interests, however, will not be enough to save the restraint if it is not *effectively* serving those interests.¹⁹⁴

b. Step 3(b): Does the Restraint Substantiate the State’s Legitimate Concerns?

If a court finds that Section 101-b serves clearly articulated interests that are expressed as state policies, and those interests outweigh the interests of the Sherman Act, the court will have one final inquiry before it may declare Section 101-b valid. The Supreme Court has held that “*unsubstantiated* state concerns . . . simply are not of the same stature as the goals of the Sherman Act.”¹⁹⁵ Courts have repeatedly struck down restraints that were put into place to serve legitimate interests for failing to actually promote these interests.¹⁹⁶ For example, in *324 Liquor Corp. v. Duffy*,¹⁹⁷ the Court acknowledged the legitimacy of New York’s desire to protect small retail establishments, but struck down the state’s “markup” statute because the state failed to show that the restraint helped those retailers.¹⁹⁸ The Court went even further than this, however, by pointing out that the *Midcal* Court had cited evidence showing that other states with similar laws had experienced increased failure of firms and decreased growth of small retail establishments.¹⁹⁹ While the Supreme Court has not stated a clear rule for when an interest is substantiated, it has

¹⁹¹ See *supra* note 115 and accompanying text.

¹⁹² See *Midcal*, 445 U.S. at 113-14.

¹⁹³ Lauzon, *supra* note 104.

¹⁹⁴ See *Midcal* 445 U.S. at 113.

¹⁹⁵ *Id.* (emphasis added).

¹⁹⁶ See, e.g., *324 Liquor Corp.*, 479 U.S. 335, 350 (1987); *Midcal*, 445 U.S. at 113-14.

¹⁹⁷ 479 U.S. 335 (1987).

¹⁹⁸ *Id.* at 350.

¹⁹⁹ *Id.*

clearly shown that it is willing to not only require empirical evidence from the state but that it will also look at evidence to the contrary.²⁰⁰ This type of approach is a far cry from the original deference applied to state liquor laws.²⁰¹

B. Post and Hold Cases Challenged as Violations of the Sherman Act

There have been several “post and hold” cases in the federal courts since the Supreme Court’s decision in *Midcal*, which can be considered the most instructive case for analyzing whether a liquor law that is challenged as preempted by the Sherman Act may be saved by the Twenty-first Amendment. With one exception, each time a federal circuit has considered the validity of a post and hold law, it has found that it was preempted by the Sherman Act,²⁰² did not qualify for antitrust immunity under *Parker*,²⁰³ and was not saved by the Twenty-first Amendment because the law’s purported goals were not substantiated.²⁰⁴ These cases illustrate the analysis outlined above and are indicative of how a court will treat ABC Law Section 101-b.

Each of the cases discussed in this section involved restraints considered “post and hold” laws, like ABC Law Section 101-b.²⁰⁵ The only case in which a federal circuit upheld a challenged post and hold restraint was *Battipaglia v. New York State Liquor Authority*,²⁰⁶ decided twenty-four years ago. The challenged restraint addressed by the Second Circuit in *Battipaglia* was none other than ABC Law Section 101-b.²⁰⁷ The majority held that Section 101-b did not violate the Sherman Act, and alternatively that if it did violate the Sherman Act, it was entitled to prevail because of the Twenty-first

²⁰⁰ *Id.*

²⁰¹ *See supra* Part II.A.

²⁰² *See Costco Wholesale Corp. v. Maleng*, 522 F.3d 874, 895-96 (9th Cir. 2008); *TFWS, Inc. v. Schaefer*, 242 F.3d 198, 209-10 (4th Cir. 2001); *Miller v. Hedlund*, 813 F.2d 1344, 1349-50 (9th Cir. 1987).

²⁰³ *See TFWS, Inc.*, 242 F.3d at 211; *Miller*, 813 F.2d at 1351-52; *see infra* notes 254-255 and accompanying text.

²⁰⁴ *See infra* notes 256-259 and accompanying text. Two of the cases, *Miller v. Hedlund* and *TFWS, Inc. v. Schaefer* were remanded on this issue and in both instances, the post and hold laws ultimately were struck down for failing to substantiate the states’ interests. *See infra* notes 229-233, 243-247 and accompanying text.

²⁰⁵ *See supra* Part I.C. for a reminder of what Section 101-b requires.

²⁰⁶ 745 F.2d 166 (2d Cir. 1984).

²⁰⁷ *Id.* at 167.

Amendment.²⁰⁸ First, the court distinguished *Midcal*, claiming that *Midcal* involved a “resale price maintenance” scheme in which wine producers could dictate prices charged by downstream sellers,²⁰⁹ and thus was not dispositive because New York “merely requires wholesalers to post and adhere to their own unilaterally determined prices and nothing more.”²¹⁰ The majority noted that courts had disagreed over whether compliance with a state law could be grounds for the finding of “agreement” as required by the Sherman Act, but declined to choose a side.²¹¹ Instead, the court held that there was no preemption because this was a facial attack, which required proof that Section 101-b was a per se violation of the Sherman Act in all instances.²¹² Section 101-b was characterized as the “exchange of specific information,” an activity that should be subject to the rule of reason²¹³ antitrust analysis, rather than be deemed a per se violation.²¹⁴ The majority then found that even if Section 101-b violated the Sherman Act, it should prevail anyway because it was intended to serve a strong state interest in preventing price discrimination, and the state had not intended to reduce competition.²¹⁵

In response to the majority in *Battipaglia*, Judge Winter, in dissent, argued that Section 101-b is a per se violation of the Sherman Act. Under Judge Winter’s analysis, not only does Section 101-b contemplate the exchange of price information, but it also requires adherence to publicly announced prices, which was always held to be illegal irrespective of reasonableness.²¹⁶ Judge Winter then went on to opine that the element of “agreement” should be found because *Midcal* does

²⁰⁸ *Id.* at 170. The majority declined to answer whether Section 101-b would be immune under the state-action doctrine. *Id.*

²⁰⁹ *Id.* at 172; see also *supra* note 56 and accompanying text.

²¹⁰ *Battipaglia*, 745 F.2d at 172.

²¹¹ *Id.* at 173 (internal quotation marks omitted). However, the court commented that “state compulsion of individual action is the very antithesis of an agreement.” *Id.*

²¹² *Id.* at 174-75.

²¹³ See *supra* note 66 (explaining that not all anticompetitive activity results in a per se violation of the Sherman Act).

²¹⁴ *Battipaglia*, 745 F.2d at 175.

²¹⁵ *Id.* at 178-79. The court noted that Section 101-b could create disincentives to reducing prices, but that the plaintiffs challenging the law had not argued this or provided any evidence that it was occurring. *Id.* at 178.

²¹⁶ *Id.* at 179 (Winter, J., dissenting).

apply to Section 101-b, contrary to what the majority reasoned.²¹⁷ After determining that Section 101-b was thus preempted by the Sherman Act, Judge Winter found that while the intentions of New York were clearly stated and affirmatively expressed, as required by *Midcal*'s first prong, Section 101-b was not immune under the state-action doctrine because New York does not actively supervise whether Section 101-b carries out its intended policies.²¹⁸ In concluding, Judge Winter commented that temperance would be a valid interest under the Twenty-first Amendment but that the Amendment should not apply in the case before the court because, in his opinion, the law was intended to allow liquor dealers to "seek out their profit-maximizing price/output level[s]."²¹⁹ Accordingly, he did not address whether Section 101-b substantiated the state's purported interest in preventing price discrimination.²²⁰

A few years after *Battipaglia* came *Miller v. Hedlund*,²²¹ in which the Ninth Circuit Court of Appeals found that several features of Oregon's liquor distribution regime violated the Sherman Act.²²² The problematic provisions included: a requirement to post future prices at least ten days before the prices were to go into effect, a requirement that permissible price decreases remain in effect for a specified period, and a requirement that the posted price not be increased because of transportation costs.²²³ In essence, this was a post and hold regime because of the requirements to post resale prices in advance and adhere to those prices. In considering whether the regulations violated the Sherman Act, the court relied on *Schwegmann* and *Midcal* to find that they constituted hybrid restraints.²²⁴ After determining that they were also per se violations of the Act because "[a]n agreement to adhere to

²¹⁷ *Id.* In essence, Judge Winter determined that this was a hybrid restraint, although he did not use the language. *See id.* (explaining that Section 101-b contemplated a combination of state and private action).

²¹⁸ *Battipaglia*, 745 F.2d at 180 (Winter, J., dissenting). Judge Winter stated that New York does not set the prices, review them for reasonableness, monitor the liquor industry's market conditions, or review the scheme. *Id.* He quoted *Midcal*: "the national policy in favor of competition is thwarted by casting a . . . gauzy cloak of state involvement over what is essentially a private price-fixing agreement." *Id.* (quoting *Cal. Retail Liquor Dealers Ass'n v. Midcal Aluminum, Inc.*, 445 U.S. 97, 106 (1980)).

²¹⁹ *Battipaglia*, 745 F.2d at 180 (Winter, J., dissenting).

²²⁰ *Id.*

²²¹ 813 F.2d 1344 (9th Cir. 1987).

²²² *Id.* at 1351.

²²³ *Id.* at 1347. Oregon's ban on volume discounts was also challenged, but not at issue on appeal. *Id.* at 1348 n.3.

²²⁴ *Miller*, 813 F.2d at 1350-51.

previously announced prices . . . is unlawful per se under the Sherman Act,”²²⁵ the court proceeded to apply the test for *Parker’s* state-action immunity.²²⁶ Again, the court relied on *Midcal* and denied immunity to the regulations, stating that Oregon failed to actively supervise them.²²⁷ Specifically, the court noted that Oregon neither set the prices nor determined their reasonableness.²²⁸ Finally, the court considered the state’s Twenty-first Amendment defense, stressed the importance of balancing the state’s claimed interests served by these regulations against the Sherman Act’s interests in fostering competition, and ultimately remanded because the factual record had not been developed on this issue.²²⁹

In the case’s conclusion, the District Court of Oregon assessed the state’s purported interests in the price posting regime, asking whether the regime in fact substantiated those interests.²³⁰ Oregon argued that its intent was to prevent price discrimination.²³¹ However, the court found that there was no evidence that price posting helped the state identify instances of price discrimination; instead, the court found that the price posting laws “authoriz[ed], facilitat[ed], and induc[ed] horizontal price fixing.”²³² Consequently, Oregon was enjoined from enforcing its post and hold laws since they were not shielded by the Twenty-first Amendment.²³³

About a decade later, another post and hold case was decided, this time in Maryland. In *TFWS, Inc. v. Schaefer*,²³⁴ the owner of a large retail liquor store in Maryland sued the State Comptroller on the grounds that the state’s liquor regulatory scheme, which required liquor wholesalers to file price schedules with the state and adhere to those prices for at least

²²⁵ *Id.* at 1349. The Supreme Court held in *324 Liquor* that a per se violation may be found in the absence of a private agreement if the state compels activity that would be a per se violation. *324 Liquor Corp. v. Duffy*, 479 U.S. 335, 345-46 (1987). In other words, the concept of agreement is treated differently for hybrid restraints in that the individuals complying with the law do not actually have to agree to fix prices in the normal sense of the word “agree.” See *supra* note 141; see generally *supra* Part III.A.1.a.

²²⁶ *Miller*, 813 F.2d at 1351-52.

²²⁷ *Id.* at 1351-52.

²²⁸ *Id.* at 1351-52 n.6.

²²⁹ *Id.* at 1352.

²³⁰ See generally *Miller v. Hedlund*, 717 F. Supp. 711 (D. Or. 1989).

²³¹ *Id.* at 712.

²³² *Id.* at 715-16.

²³³ *Id.* at 716.

²³⁴ 242 F.3d 198 (4th Cir. 2001).

a month after posting, violated the Sherman Act.²³⁵ The Fourth Circuit Court of Appeals first declared that “[t]he post-and-hold system is a classic hybrid restraint”²³⁶ because it requires private parties (wholesalers) to set prices, which are not reviewed for reasonableness, thus giving those parties a great deal of “private regulatory power.”²³⁷ Next, the court relied on *Miller’s* analysis to hold that the law was a per se violation of the Sherman Act.²³⁸ The court explained that the post and hold regime “mandate[d] activity that is essentially a form of horizontal price fixing, which has been called ‘the paradigm of an unreasonable restraint of trade.’”²³⁹ Maryland, like Oregon, was unable to establish state-action immunity for the post and hold laws, with the court relying on *Midcal*²⁴⁰ and *324 Liquor*²⁴¹ to explain that the state failed to set prices, review the privately-set prices for reasonableness, monitor market conditions in the liquor industry, or “engage in any ‘pointed reexamination’ of the [post and hold regime].”²⁴²

With respect to Maryland’s Twenty-first Amendment defense, the Fourth Circuit remanded the case to develop the record, in order to determine whether the post and hold pricing scheme substantiated Maryland’s avowed interest in promoting

²³⁵ *Id.* at 201-02. Also at issue in the case was a ban on volume discounts, which the Court struck down. *See id.* at 202, 210. New York law also bans volume discounts, N.Y. ALCO. BEV. CONT. LAW § 101-b(2)(a), and it is reasonable to believe that this would be struck down as well, based on the reasoning that applies to the post and hold law. *See infra* note 237.

²³⁶ *TFWS Inc.*, 242 F.3d at 208.

²³⁷ *Id.* at 208-09. The Court also noted that “[t]he volume discount ban is a part of the hybrid restraint because it reinforces the post-and-hold system by making it even more inflexible.” *Id.* at 209. The court later went on to hold that the volume discount ban was also a per se violation of the Sherman Act. *Id.* at 210.

²³⁸ *Id.* at 209-10. The Court commented that “[s]everal district courts have reached the same result,” citing *Beer & Pop Warehouse v. Jones*, 41 F. Supp. 2d 552, 560-62 (M.D. Pa. 1999) (“holding that Pennsylvania post-and-hold pricing statute for beer was a per se violation of the Sherman Act”) and *Canterbury Liquors & Pantry v. Sullivan*, 16 F. Supp. 2d 41, 47-48 (D. Mass. 1998) (“holding that Massachusetts post-and-hold liquor pricing scheme was a per se violation of § 1 [of the Sherman Act]”). *Id.* at 210. The court acknowledged *Battipaglia’s* approach, see *supra* notes 206-215 and accompanying text, and then declined to follow, saying that no other court has followed it and a “leading commentator on antitrust law” had agreed with Judge Winter. *TFWS Inc.*, 242 F.3d at 210; *supra* note 216 and accompanying text.

²³⁹ *TFWS*, 242 F.3d at 209 (quoting *N.C.A.A. v. Bd. of Regents of the Univ. of Okla.*, 468 U.S. 85, 100 (1984)).

²⁴⁰ *See supra* notes 176-179 and accompanying text.

²⁴¹ *See supra* notes 176-179 and accompanying text.

²⁴² *TFWS, Inc.*, 242 F.3d at 211 (internal citation omitted) (quoting *324 Liquor Corp. v. Duffy*, 479 U.S. 335, 344-45 (1987)).

temperance.²⁴³ The case went through several stages in both the district court and back up to the Fourth Circuit before it was finally resolved in 2007, in favor of TFWS (the liquor storeowner).²⁴⁴ After extensive evidentiary findings (and disputes) involving expert testimony on both theoretical and empirical studies²⁴⁵ the district court determined that the state's evidence that the scheme promoted temperance was tenuous, and thus outweighed by the federal interest in fostering competition.²⁴⁶ In other words, Maryland was unable to save its post and hold laws because although it had a valid interest in promoting temperance, that interest was not substantiated, as required by *Midcal* and *324 Liquor*.²⁴⁷

The most recent federal decision to find that a post and hold law violated the Sherman Act was in the Ninth Circuit Court of Appeals, this time addressing a Washington law in *Costco Wholesale Corp. v. Maleng*.²⁴⁸ Washington had a post and hold system similar to both Oregon and Maryland, in that wholesalers were required to file prices and adhere to them for a specified period after they went into effect.²⁴⁹ The *Costco* court's analysis was not as clear as that of *Miller* or *TFWS*. For example, it appeared that the court wanted to collapse the inquiry of whether the post and hold law was a hybrid restraint with the inquiry of whether the post and hold system was actively supervised by the state for purposes of antitrust immunity under the state action doctrine.²⁵⁰ While the *Costco* court questioned the clarity of the unilateral-hybrid restraint versus an active supervision analysis, it ultimately followed the approach of *Miller* and *TFWS*, first concluding that the law

²⁴³ See *TFWS, Inc.*, 242 F.3d at 212-13. The court acknowledged that temperance is an interest contemplated by the Twenty-first Amendment. *Id.* at 213.

²⁴⁴ See *TFWS, Inc. v. Schaefer*, No. WDQ-99-2008, 2007 WL 2917025, at *10 (D. Md. Sept. 27, 2007).

²⁴⁵ See, e.g., *TFWS, Inc. v. Schaefer*, 183 F. Supp. 2d 789, 791-94 (D. Md. 2002); *TFWS, Inc. v. Schaefer*, 315 F. Supp. 2d 783, 783-84 & n. 1 (D. Md. 2004); *TFWS, Inc.* 2007 WL 2917025, at *2-8.

²⁴⁶ *TFWS, Inc.*, 2007 WL 2917025, at *10.

²⁴⁷ See *supra* notes 176-179 and accompanying text.

²⁴⁸ 522 F.3d 874 (9th Cir. 2008).

²⁴⁹ See *id.* at 883; see *supra* note 223 and accompanying text; see *supra* note 235 and accompanying text.

²⁵⁰ See *Costco*, 522 F.3d at 887-88. The court was fairly reasonable in concluding that this is a "doctrinally confusing area." *Id.* at 888. However, it is clear that a law may be considered a hybrid restraint, in that it gives a degree of regulatory power to individuals, see *supra* Part III.A.1.a., but that the law may also be immune because the state reviews the individuals' exercise of that power for reasonableness. See *supra* Part III.A.2.b.

was a hybrid restraint. The court explained that while the wholesalers are not required to match others' prices, "the logical result of the restraints is a less uncertain market, a market more conducive to collusive and stabilized pricing, and hence a less competitive market."²⁵¹ In other words, Washington set up a system that facilitated price-fixing by private parties.²⁵² The court then concluded that the law was also a per se violation of the Sherman Act because it was "highly likely to facilitate horizontal collusion among market participants."²⁵³

Moving on to the antitrust immunity issue, i.e., state-action doctrine, the court applied *Miller* to the case and held that Washington, like Oregon, failed to actively supervise its post and hold scheme,²⁵⁴ and thus had not established immunity for the scheme.²⁵⁵ Finally, the court considered whether the post and hold restraint was saved by the Twenty-first Amendment defense.²⁵⁶ This time, a factual record had been developed on the issue and the district court had already decided against Washington.²⁵⁷ The court affirmed the decision against the state, agreeing with the district court that temperance was a "valid and important interest" under the Twenty-first Amendment, but Washington failed to show that the post and hold regulation promoted temperance.²⁵⁸ In doing so, the court repeated the district court's finding that "there was little empirical evidence documenting the relationship between such pricing schemes and consumption."²⁵⁹

Although Section 101-b was facially challenged and upheld in *Battipaglia*,²⁶⁰ it has become clearer over time that contrary to *Battipaglia*'s characterization of Section 101-b as the "exchange of price information" that should be subject to a rule of reason analysis,²⁶¹ it is in fact a post and hold provision and thus a "classic hybrid restraint."²⁶² It is inappropriate to

²⁵¹ *Costco*, 522 F.3d at 888, 893-94.

²⁵² *See id.* at 894-95.

²⁵³ *Id.* at 895-96. Like the court in *TFWS*, the *Costco* court discussed *Battipaglia* and then declined to follow it. *See id.* at 893-94; *supra* note 238.

²⁵⁴ *Costco*, 522 F.3d at 901 n.22.

²⁵⁵ *See supra* Part III.A.2.b. (explaining that states are often barred from immunity due to the absence of active supervision)

²⁵⁶ *Costco*, 522 F.3d at 901-04.

²⁵⁷ *Id.* at 902-03.

²⁵⁸ *Id.*

²⁵⁹ *Id.* at 903.

²⁶⁰ *See supra* notes 206-215 and accompanying text.

²⁶¹ *Battipaglia v. New York State Liquor Auth.*, 745 F.2d 166, 174-75 (2d Cir. 1984).

²⁶² *TFWS, Inc. v. Schaefer*, 242 F.3d 198, 208 (4th Cir. 2001).

conclude that Section 101-b will be struck down on an antitrust challenge without any evidence regarding the Twenty-first Amendment defense.²⁶³ Nevertheless, the following section will apply the analysis outlined above²⁶⁴ to Section 101-b to demonstrate why it is unlikely that the State of New York can protect Section 101-b in an antitrust challenge.

IV. ABC LAW SECTION 101-b WILL LIKELY LOSE ON AN ANTITRUST CHALLENGE

Given the evolution of the jurisprudence in the Supreme Court and the federal courts regarding liquor laws that mandate anticompetitive behavior, it seems very unlikely that Section 101-b would be able to withstand another antitrust challenge. Section 101-b, like the post and hold restraints at issue in *Miller*,²⁶⁵ *TFWS*,²⁶⁶ and *Costco*,²⁶⁷ requires not only the filing of prices, but also adherence to those prices. To recap the requirements of Section 101-b, manufacturers and wholesalers must file a monthly posting with the State Liquor Authority (“SLA”) that will go into effect for the following pricing period after a three-day window in which prices may be reduced to match the lowest posted price for the same product.²⁶⁸ Once the prices are in effect, they cannot be changed without the SLA’s prior written permission.²⁶⁹ Since Section 101-b is very similar to the post and hold restraints challenged and struck down by *Miller*, *TFWS*, and *Costco*, it should be analyzed in the same manner.

First, Section 101-b is preempted by the Sherman Act because it is a hybrid restraint and a per se violation of the Act. As noted above, while the Second Circuit declared that Section 101-b simply requires exchanging price information, the majority conveniently overlooked the holding aspect of Section

²⁶³ See *Miller v. Hedlund*, 813 F.2d 1344, 1352 (9th Cir. 1987); see also *supra* note 243 and accompanying text.

²⁶⁴ See *supra* Part III.A (illustrated in Part III.B.).

²⁶⁵ See *supra* notes 221-233 and accompanying text.

²⁶⁶ See *supra* notes 234-247 and accompanying text.

²⁶⁷ See *supra* notes 248-259 and accompanying text.

²⁶⁸ See *supra* notes 75-77. At first glance, the statute may seem beneficial to consumers as it allows manufacturers and wholesalers to decrease their prices to the lowest price posted. The harm is that generally there is an ongoing possibility that, in a free market, prices for a good will fluctuate; here, after the three-day window ends the manufacturers and wholesalers are unable to lower prices even if the market would justify a reduction. In other words, consumers lose the benefits of competition as the manufacturers and wholesalers simply have no incentives to compete via price changes after the posting period goes into effect (because they know they cannot be undercut).

²⁶⁹ See *supra* notes 75-77.

101-b,²⁷⁰ which is the primary reason this type of law is problematic. The holding aspect of Section 101-b “logical[ly] result[s] . . . [in] a less uncertain market, . . . and hence a less competitive market.”²⁷¹ In 2004, the Second Circuit stated, “[w]here the anticompetitive effects of a state statute obviate the need for private parties to act on their own to create an anticompetitive scheme, the statute may be attacked as a ‘hybrid’ restraint.”²⁷² Section 101-b is therefore a hybrid restraint because it delegates private regulatory power to the distributors and wholesalers by allowing them to set the prices, which the State merely enforces. In other words, Section 101-b is not a unilateral restraint because of the degree of power given to private market participants, and it is also not a private restraint because of the State’s authorization.²⁷³ It is also significant that Section 101-b involves a price restraint because a price restraint is especially prone to being deemed a hybrid restraint.²⁷⁴ Consistent with the Second Circuit’s requirements under its more recent approach to the issue of hybrid restraints, New York’s price posting regime reduces the need for liquor dealers and wholesalers to create their own anticompetitive scheme. Thus, Section 101-b is a hybrid restraint subject to preemption.

Section 101-b is also a per se violation of the Sherman Act and is thus preempted. By forcing manufacturers and wholesalers to hold to their announced prices, the state “mandates activity that is essentially a form of horizontal price fixing.”²⁷⁵ As the *Costco* court explained, horizontal collusion allows market participants to maximize profits via price (and production) coordination at the expense of consumers by increasing prices (and decreasing production).²⁷⁶ Requiring adherence to posted prices makes price cuts irrevocable, and thus “much less likely.”²⁷⁷ Furthermore, as the *Miller* court

²⁷⁰ The dissent, meanwhile, emphasized the significance of the holding requirement. See *supra* note 216 and accompanying text.

²⁷¹ *Costco Wholesale Corp. v. Maleng*, 522 F.3d 874, 892-94 (9th Cir. 2008).

²⁷² *Freedom Holdings, Inc. v. Spitzer*, 357 F.3d 205, 223-24 (2d Cir. 2004) (holding that New York’s Contraband Statutes were hybrid restraints subject to preemption by the Sherman Act for enforcing price-fixing among major tobacco producers).

²⁷³ See *supra* Part III.A.1.a.

²⁷⁴ See *supra* notes 157-158 and accompanying text.

²⁷⁵ *TFWS, Inc. v. Schaefer*, 242 F.3d 198, 208 (4th Cir. 2001).

²⁷⁶ *Costco*, 522 F.3d at 896.

²⁷⁷ *Id.*

explicitly stated, “[a]n agreement to adhere to previously announced prices . . . is unlawful *per se* under the Sherman Act.”²⁷⁸ While the majority in *Battipaglia* refrained from deciding whether Section 101-b could be preempted by the Sherman Act without actual agreement between the manufacturers and/or wholesalers, that issue has since been decided.²⁷⁹ In *324 Liquor*, the Supreme Court held that a *per se* violation may be found in the absence of a private agreement if the state compels activity that would otherwise be a *per se* violation.²⁸⁰ Indeed, the Second Circuit has since acknowledged that the Supreme Court does not require actual agreement as a prerequisite to preemption under the Sherman Act.²⁸¹ Thus, Section 101-b is a hybrid restraint for delegating regulatory power to private individuals, and it is a *per se* violation because adhering to posted prices is illegal under the Sherman Act.

Second, Section 101-b is most likely not immune under the state-action doctrine. As required by prong one of the *Midcal* test for antitrust immunity,²⁸² New York has “clearly articulated and affirmatively expressed as state policy” its intent to promote temperance and orderly market conditions by prohibiting price discrimination with Section 101-b.²⁸³ In *Midcal*, California satisfied prong one of the test when it clearly stated its goal of permitting price resale maintenance as legislative policy.²⁸⁴ The Supreme Court similarly found prong one satisfied in *324 Liquor*, where New York also clearly intended to allow price resale maintenance.²⁸⁵ Interestingly, in *Freedom Holdings, Inc. v. Spitzer*,²⁸⁶ the Second Circuit found that New York failed to satisfy prong one of *Midcal* when it claimed an interest in revenue production was the underlying

²⁷⁸ *Miller v. Hedlund*, 813 F.2d 1344, 1349 (9th Cir. 1987).

²⁷⁹ *See 324 Liquor Corp. v. Duffy*, 479 U.S. 335, 345-46 (1987).

²⁸⁰ *Id.*

²⁸¹ *See Freedom Holdings, Inc. v. Spitzer*, 357 F.3d 205, 224 n.17 (2d Cir. 2004); *see 324 Liquor Corp.*, 479 U.S. at 345-46 & n.8.

²⁸² *See supra* Part III.A.2.a. for a refresher on *Midcal*'s first prong to establish immunity under the state-action doctrine.

²⁸³ *Battipaglia v. New York State Liquor Auth.*, 745 F.2d 166, 176 (2d Cir. 1984) (quoting *Cal. Retail Liquor Dealers Ass'n v. Midcal Aluminum, Inc.*, 445 U.S. 97, 105). In dissent, Judge Winter stated that New York's policy of “creating a cartel” with Section 101-b was “one clearly articulated and affirmatively expressed” by the state,” which satisfied the first part of the *Midcal* test. *Id.* at 180 (Winter, J. dissenting) (quoting *Midcal*, 445 U.S. 97 at 105); *see also* N.Y. ALCO. BEV. CONT. LAW § 101-b(1).

²⁸⁴ *See Midcal*, 445 U.S. at 105.

²⁸⁵ *See 324 Liquor Corp.*, 479 U.S. at 344.

²⁸⁶ 357 F.3d 205 (2d Cir. 2004).

goal of enforcing a price-fixing scheme among major tobacco producers.²⁸⁷ The Second Circuit explained, “an ancillary function of the first *Midcal* prong is to establish the legitimate State policy underlying the decision to displace the Sherman Act.”²⁸⁸ Even if the legitimacy of New York’s interests are assessed at this stage of the analysis, as opposed to waiting until the Twenty-first Amendment defense is raised,²⁸⁹ New York will still likely satisfy prong one of *Midcal* because its interests in promoting temperance and orderly market conditions involve public and economic interests beyond mere revenue production for the state.

While Section 101-b will probably pass the first inquiry under *Midcal*, it most likely will fail *Midcal*’s second prong, which requires that New York “actively supervise” the implementation of Section 101-b.²⁹⁰ Post and hold restraints similar to Section 101-b have repeatedly failed to satisfy prong two of *Midcal* because the states responsible for the laws “neither establishe[d] prices nor review[ed] the reasonableness of the price schedules,” and the states failed to “monitor market conditions or engage in any ‘pointed reexamination’ of the program[s].”²⁹¹ In finding that ABC Law Section 101-bb was not actively supervised by New York, the Supreme Court in *324 Liquor* reasoned, “[t]he State has displaced competition among liquor retailers without substituting an adequate system of regulation.”²⁹² Judge Winter, in his dissent from the *Battipaglia* majority, stated that New York “does nothing whatsoever to establish the actual prices charged, review their reasonableness, monitor market conditions, or engage in reexamination of the program.”²⁹³ As it had done in *324 Liquor*, New York persists in displacing competition without an

²⁸⁷ See *id.* at 230.

²⁸⁸ *Id.*

²⁸⁹ See *supra* Part III.A.3.

²⁹⁰ See *supra* Part III.A.2.b.

²⁹¹ Cal. Retail Liquor Dealers Ass’n v. Midcal Aluminum, Inc., 445 U.S. 97, 105-06 (1980) (citation omitted); see also Costco Wholesale Corp. v. Maleng, 522 F.3d 874, 901 n.22 (9th Cir. 2008); TFWs, Inc. v. Schaefer, 242 F.3d 198, 211 (4th Cir. 2001); Miller v. Hedlund, 813 F.2d 1344, 1351-52 & n.6 (9th Cir. 1987).

²⁹² 324 Liquor Corp. v. Duffy, 479 U.S. 335, 344-45 (1987).

²⁹³ Battipaglia v. New York State Liquor Auth., 745 F.2d 166, 180 (2d Cir. 1984) (Winter, J., dissenting). Indeed, the New York State Law Revision Commission reported that the SLA does not monitor posted prices. See NEW YORK STATE LAW REVISION COMMISSION REPORT ON THE ALCOHOL BEVERAGE CONTROL LAW AND ITS ADMINISTRATION [Hereinafter COMMISSION REPORT PART ONE], 34 (September 30, 2009), available at <http://www.lawrevision.state.ny.us/abcls.php> (last visited Feb. 18, 2010).

adequate system of regulation by giving manufacturers and wholesalers discretion over prices and enforcing them without regard to their reasonableness. Accordingly, Section 101-b will most likely not be immune, for failing prong two of *Midcal*.

Finally, it is very unlikely that Section 101-b will prevail if New York asserts the Twenty-first Amendment defense. Not only must New York have a legitimate policy supporting Section 101-b, the law must also be effective in serving that policy.²⁹⁴ In determining whether New York's interests are legitimate, a court must find that they are "closely related" to the goals of the Twenty-first Amendment,²⁹⁵ and that New York's interests outweigh the federal interests of the Sherman Act, which has been described as "the Magna Carta of free enterprise."²⁹⁶ As an initial matter, New York's stated interest in promoting temperance is certainly a legitimate state interest.²⁹⁷ New York also has expressed intent to prohibit price discrimination for the purpose of orderly markets,²⁹⁸ which is also likely a legitimate state interest.²⁹⁹ However, the Twenty-first Amendment will likely fail to protect Section 101-b because New York will probably not be able to meet its burden of showing that Section 101-b actually promotes temperance, prevents price discrimination, or promotes orderly markets.

In order to show that Section 101-b substantiates its purported goals, New York will have to spend considerable time and money to produce persuasive evidence. With respect to showing that Section 101-b promotes temperance, perhaps New York could prepare analytic state studies on consumption, possibly distinguishing between New York and another state without a post and hold restraint in place. However, a

²⁹⁴ See *supra* Part III.A.3.

²⁹⁵ *Capital Cities Cable, Inc. v. Crisp*, 467 U.S. 691, 714 (1984); see *supra* note 112-113 and accompanying text.

²⁹⁶ *Midcal*, 445 U.S. at 110; see *supra* Part III.A.3.

²⁹⁷ See *supra* notes 192, 243, and 258 and accompanying text.

²⁹⁸ N.Y. ALCO. BEV. CONT. LAW § 101-b(1) (McKinney 2009).

²⁹⁹ With respect to New York's interest in promoting orderly markets, this argument was addressed in a footnote by the *Costco* court. See *Costco Wholesale Corp. v. Maleng*, 522 F.3d 874, 902 n.23 (9th Cir. 2008). Washington cited *North Dakota v. United States*, 495 U.S. 423, 432 (1990) as support for its argument that it had an interest in orderly markets, but the court explained the concept of "orderly markets" was hard to define and thus there could be no clear error by the district court in deciding that this interest was not substantiated by the challenged post and hold restraint. *Id.* With respect to prohibiting price discrimination, the *Miller* court apparently accepted that this was a legitimate state interest as well, as the court went on to inquire whether the interest was substantiated, ultimately finding it was not. See *Miller v. Hedlund*, 813 F.2d 1344, 1352; see also *supra* notes 231-233.

challenger may rebut such evidence and a court does not have to give deference to the state's evidence.³⁰⁰ With respect to preventing price discrimination and promoting orderly markets, New York could present state agency reports and/or congressional studies regarding effects of Section 101-b on market conditions, and empirical economic evidence.³⁰¹ Of course, these studies must first be performed, assuming no such studies on this precise issue have been prepared as of yet.³⁰² A challenger may also produce conflicting studies, again giving a court the choice of whose evidence to accept.³⁰³ Finally, New York will most likely need to produce expert witness testimony as well,³⁰⁴ which may also be rebutted.

This is not to say that it is impossible for New York to save Section 101-b if it is challenged. Rather, it is to emphasize the amount of effort that New York will have to invest to show that Section 101-b should be sustained, and that even with extensive evidence, there is no guarantee that a court will find in New York's favor. Unless New York is able to develop a record showing Section 101-b fosters its stated interests, Section 101-b will be struck down on an antitrust challenge. *Unsubstantiated* state interests, no matter how closely related to the Twenty-first Amendment, cannot outweigh the Sherman Act's policy of promoting competition.³⁰⁵

³⁰⁰ See *supra* note 200 and accompanying text (explaining that a court will examine evidence to the contrary).

³⁰¹ See *supra* notes 243-247 and accompanying text.

³⁰² This seems to be a fair assumption considering the New York State Law Revision Commission's recent findings:

The SLA is unable to determine industry's compliance with the law. Price posting information is not monitored so it is no surprise that the SLA would fail to detect abuses in the industry. Because it does not monitor the information, it is unable to demonstrate that the objectives of the post and hold process are achieved.

COMMISSION REPORT PART ONE at 34.

³⁰³ For example, in *Miller*, while Washington argued that its post and hold restraint prevented price discrimination, the court agreed with the challenger that rather than prevent price discrimination, the price posting laws simply "authoriz[ed], facilitat[ed], and induc[ed] horizontal price fixing." *Miller v. Hedlund*, 717 F. Supp. 711, 715-16 (D. Or. 1989).

³⁰⁴ See *supra* notes 243-247 and accompanying text.

³⁰⁵ See *supra* note 195 and accompanying text.

CONCLUSION

Given the problems that alcohol has caused in the past and continues to cause today,³⁰⁶ it is no surprise that New York wants to have special regulations imposed on the liquor industry. However, it is unreasonable to be overly concerned with the regulation of alcohol distribution at the expense of the Sherman Act, and the goals of the Sherman Act should not be discarded. Quite the contrary, these goals are just as important to the promotion of social welfare as the desire to prevent excessive consumption and price discrimination.³⁰⁷ However, if New York insists upon sacrificing the pro-competition policy of the Sherman Act, it must take a more proactive role in implementing Section 101-b,³⁰⁸ which would probably take no more effort than putting up a strong defense under the Twenty-first Amendment. Whether New York wishes to create or find evidence conclusively showing that Section 101-b actually promotes temperance, prevents price discrimination, or promotes orderly markets, or whether New York wishes to take a more active role in supervising its price posting system, one thing is clear: some sort of action should be taken to prevent the law invalidation in the event of an antitrust challenge. Despite confusing and sometimes inconsistent individual opinions regarding the Twenty-first Amendment's protection of liquor regulations,³⁰⁹ it has become increasingly clear over time that the current state of the law will not permit Section 101-b to stand if challenged.

Tammy E. Linn[†]

³⁰⁶ See *supra* notes 1-5 and accompanying text.

³⁰⁷ See *supra* notes 44-53 and accompanying text.

³⁰⁸ If New York takes this action, then Section 101-b would likely qualify for antitrust immunity under the state-action doctrine as an actively supervised hybrid restraint. See *supra* Part III.A.2.b.

³⁰⁹ See *supra* Part II.

[†] J.D. Candidate, Brooklyn Law School, 2010. I would like to thank Professor Robert Pitler for introducing me to this legal issue and for providing invaluable support throughout the last three years. I would also like to thank the editors and staff of *Brooklyn Law Review*, especially Andrei Takhteyev, Joseph Roy, William Vandivort, and Melissa Palombo for their helpful suggestions. Finally, I would like to give a special thank you to Adam Greenberg and my mom, Robin Linn, for having endless patience during the writing process of this Note.

License to Kill

MDY V. BLIZZARD AND THE BATTLE OVER COPYRIGHT IN WORLD OF WARCRAFT

I. INTRODUCTION

Copyright law grants a limited bundle of exclusive rights to copyright owners.¹ These rights include the exclusive right to reproduce and distribute the work.² However, these rights are limited as the law distinguishes between protecting one's intellectual property in a product and protecting a right to the product in and of itself.³

In *MDY Industries, LLC v. Blizzard Entertainment, Inc.*⁴ the District Court of Arizona upheld Ninth Circuit precedent that gutted this distinction, finding that the purchaser and user of the video game, World of Warcraft ("WoW"), is a licensee of that game, not an "owner."⁵ By finding that a WoW user was a mere licensee and not an "owner" of the software, the *MDY* court concluded that the user was not protected by

¹ See 17 U.S.C. § 106 (2006). "The purpose of copyright is to grant authors a limited property right in the form of expression of their ideas." NAT'L COMM'N OF NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS, FINAL REPORT 16 (Library of Congress 1979), available at <http://digital-law-online.info/CONTU/contu1.html> [hereinafter CONTU REPORT].

² Brief for Public Knowledge as Amici Curiae Supporting Neither Party at 5, *MDY Indus., LLC v. Blizzard Entm't, Inc.*, 2008 WL 2757357 (D. Ariz. 2008) (No. CV06-0255-PHX-DGC) available at <http://www.publicknowledge.org/pdf/pk-amicus-20080502.pdf>. [hereinafter Public Knowledge]; see also 17 U.S.C. § 106.

³ Indeed, Congress specifically recognized this distinction when codifying the Copyright Act.

Ownership of a copyright, or of any of the exclusive rights under a copyright, is distinct from ownership of any material object in which the work is embodied. Transfer of ownership of any material object, including the copy or phonorecord in which the work is first fixed, does not of itself convey any rights in the copyrighted work embodied in the object; nor, in the absence of an agreement, does transfer of ownership of a copyright or of any exclusive rights under a copyright convey property rights in any material object.

17 U.S.C. § 202 (2006).

⁴ No. CV-06-02555-PHX-DGC, 2008 WL 2757357 (D. Ariz. July 14, 2008).

⁵ *Id.* at *8-10. The Copyright Act encompasses video games and other similar computer programs. See 17 U.S.C. § 102.

the Copyright Act's Section 117(a)(1) safe harbor provision, which allows "owners" to copy software to a computer's Random Access Memory ("RAM") as an "essential step" in using the program.⁶ Thus, the *MDY* court held that when a user played WoW using a popular third-party application known as WoWGlider ("Glider"), the user exceeded his license in the End User License Agreement ("EULA") and Terms of Use ("TOU"), and created infringing copies of the game in the computer's RAM.⁷ Because of these infringing copies, the court held MDY Industries, the owner of Glider, liable for contributory and vicarious copyright infringement⁸ resulting in \$6,000,000 in damages.⁹

In addition to snuffing out Glider use, the *MDY* decision disrupted the delicate balance between a copyright holder's ability to protect its intellectual property and a consumer's right to use his particular copy without being held liable for copyright infringement. Indeed, the *MDY* decision facilitated a "chilling extension of control" by copyright holders over their software.¹⁰

While the *MDY* court followed a line of Ninth Circuit precedent under *Wall Data Inc. v. Los Angeles County Sheriff's Department*¹¹ and *MAI Systems Corp. v. Peak Computer Inc.*,¹² which gave conclusive weight to the software provider's EULA when determining whether a purchaser owned a piece of

⁶ *Id.* at *8, *10.

⁶ *Id.* at *1, *10; RAM is a form of computer data storage in which information can be temporarily recorded. LEE HOLLAR, LEGAL PROTECTION OF DIGITAL INFORMATION, ch.2, sec. II.C.1 (2002), <http://digital-law-online.info/lpdi1.0/treatise20.html>. Whenever software is loaded into RAM, a copy is created. *Id.*; see also *MAI Sys. Corp. v. Peak Computer Inc.*, 991 F.2d 511, 517-18 (9th Cir. 1993). When the computer is turned off, the data is lost. HOLLAR, *supra*, at ch.2, sec. II.C.1.

⁸ *MDY Indus., LLC*, 2008 WL 2757357 at *10.

⁹ MDY paid Blizzard a stipulated judgment pending appeal. Benjamin Druanske, *MDY Agrees to Pay Blizzard \$6m in Damages of Warcraft Bot Lawsuit, Pending Appeal*, VIRTUALLY BLIND, Sept. 29, 2008, <http://virtuallyblind.com/2008/09/29/mdy-blizzard-damages/>. Outside the scope of this note, but of significant interest, is the court's ruling on the remaining issues left unresolved by summary judgment. The court found that MDY violated Sections 1201(a)(2) and 1201(b)(1) of the Digital Millennium Copyright Act ("DMCA"), that MDY's owner was personally liable for MDY's DMCA and copyright violations, and that Blizzard was entitled to a permanent injunction against Glider sales. *MDY Indus., LLC v. Blizzard Entm't, Inc.*, 616 F. Supp. 2d 958, 962-68 (D. Ariz. 2009).

¹⁰ The Patry Copyright Blog, <http://williampatry.blogspot.com/2008/07/strange-copyright-world-of-warcraft.html> (July 15, 2008, 08:48 EDT).

¹¹ 447 F.3d 769 (9th Cir. 2006).

¹² 991 F.2d 511 (9th Cir. 1993).

software,¹³ this Note argues the court's holding was ultimately incorrect for three reasons. First, the *MDY* court should have followed an alternative line of Ninth Circuit precedent under *United States v. Wise*¹⁴ and *Vernor v. Autodesk, Inc.*,¹⁵ which more equitably allocates rights between software providers and software purchasers. Specifically, *Wise* and *Vernor* utilized the First Sale Doctrine, which focuses on the economic realities of the underlying transaction surrounding the software purchase to determine whether a purchaser is an "owner" or licensee of the software instead of granting the software provider's EULA conclusive weight.¹⁶ Under this precedent, the *MDY* court should have found for the third-party application maker and held that WoW users are software "owners," not licensees. Second, courts should be informed by John Locke's theory of labor desert when analyzing whether a WoW user is a licensee or an "owner."¹⁷ Lockean labor desert theory argues that ownership rights are created by the investment of time and labor in creating a good such as a WoW user's self-created character, or avatar. Third, the Copyright Act's underlying policies favoring progress and innovation counsel in favor of more substantial protections for WoW users' rights.

Part II of this Note discusses the background of WoW and Glider. Part III then discusses the *MDY* case and the precedent developed under *Wall Data* and *MAI* that led to the court's decision. Part IV argues that *MDY* was wrongly decided. It first examines the contrary Ninth Circuit precedent under *Wise* and *Vernor*. Second, it discusses why courts should afford software purchasers and their time investments greater, though not absolute, protection under Lockean labor desert theory. Third, it argues the policies of copyright law require greater protection of WoW users' rights. This Note concludes by summarizing why the *MDY* decision was incorrect and how the case should have been decided.

¹³ See *MDY Indus. LLC*, 2008 WL 2757357, at *8.

¹⁴ 550 F.2d 1180 (9th Cir. 1977).

¹⁵ 555 F. Supp. 2d 1164 (W.D. WA 2008).

¹⁶ See *United States v. Wise*, 550 F.2d 1180, 1188-90 (9th Cir. 1977); *Vernor v. Autodesk*, 555 F. Supp. 2d 1164, 1169-70 (W.D. Wash. 2008).

¹⁷ JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* 306 (Peter Laslett ed., Cambridge Univ. Press 1988) (1690).

II. BACKGROUND

A. *World of Warcraft*

1. World of Warcraft and MMORPGs Generally

WoW is a massive multiplayer online role-playing game (“MMORPG”) released in 2004 by Blizzard Entertainment, Inc. (“Blizzard”), based upon the Warcraft universe,¹⁸ a Tolkien-esque fantasy world explicated in a series of video games that Blizzard created.¹⁹ WoW is the most successful MMORPG ever²⁰—it has over eleven million players²¹ and generates over \$1.5 billion in annual revenue for Blizzard.²² WoW generates its revenue mainly through a monthly fee.²³ Of course, Blizzard profits from the player’s initial purchase of WoW’s physical software package as well.²⁴ Blizzard has also released two sizable expansion packs²⁵ for WoW, which have further contributed to WoW’s success.²⁶

In an MMORPG, hundreds or thousands of players exist in the same virtual world²⁷ at the same time, which creates an

¹⁸ WorldofWarcraft, Game Guide, What is WoW, www.worldofwarcraft.com/info/basics/guide.html (last visited Jan. 2, 2010). For a discussion of the “Warcraft universe” as well as WoW’s background history, see World of Warcraft Europe, Warcraft History Library, <http://www.wow-europe.com/en/info/story/index.html#history> (last visited Jan. 2, 2010).

¹⁹ *MDY Indus., LLC*, 2008 WL 2757357 at *1.

²⁰ *Id.*

²¹ Press Release, Blizzard, World of Warcraft® Subscriber Base Reaches 11.5 Million Worldwide (Nov. 21, 2008) (<http://www.blizzard.com/us/press/081121.html>).

²² *MDY Indus., LLC*, 2008 WL 2757357 at *1.

²³ World of Warcraft, Game Guide, General F.A.Q., www.worldofwarcraft.com/info/faq/general.html (last visited Jan 20, 2009). The monthly fee is required to support customer service and WoW content updates. *Id.*

²⁴ See Blizzard, Blizzard Store, <http://www.blizzard.com/store/browse.xml?f=p:110000034,p:110000018,p:110000044> (last visited Jan. 20, 2009).

²⁵ See World of Warcraft, Cataclysm F.A.Q., <http://www.worldofwarcraft.com/cataclysm/faq/> (last visited Mar. 12, 2010).

²⁶ In 2007, Blizzard released its first expansion pack, “The Burning Crusade,” WorldofWarcraft, Game Guide, Intro to WoW, www.worldofwarcraft.com/info/beginners/index.html (last visited Jan. 29, 2009), and in 2008 Blizzard released the second, “Wrath of the Lich King.” *Id.* Blizzard currently plans to release a third WoW expansion pack (“Cataclysm”), Press Release, Blizzard World of Warcraft: Cataclysm Unveiled (Aug. 21 2009) (<http://us.blizzard.com/en-us/company/press/pressreleases.html?090821>), sometime in 2010. Posting of Adam Holisky to WOW.com, <http://www.wow.com/2009/08/23/world-of-warcraft-cataclysm-targeted-for-a-2010-release-date/> (Aug. 23, 2009, 1:03PM).

²⁷ “Virtual worlds are persistent, dynamic computer-based environments in which interconnected users interact with each other and the virtual environment around them.” Steven Horowitz, *Competing Lockean Claims to Virtual Property*, 20 HARV. J.L. & TECH 443, 443-44 (2007).

engaging, interactive atmosphere that more closely approximates human reality than a closed game in which only one or a small number of players are controlled by actual humans and a computer controls the rest of the environment.²⁸ To illustrate the distinction, imagine if humans controlled the ghosts chasing Ms. Pac-Man in the popular arcade game, thus eliminating any advantage gained by memorizing the computer program's built-in instructions for controlling the ghosts' movement.

Events that change the virtual world and affect other players constantly occur, even when those other players are not playing the game.²⁹ An MMORPG's immersive, interactive atmosphere adds to the intricacies of the game. These intricacies are far more complicated than other game genres such as "first-person shooter" games in which the objective is to kill computer-generated monsters or other virtual people.³⁰ Instead, MMORPGs envelop a player in an entire virtual world where users do more than kill. For example, players can role-play with different identities and connect to other users in a virtual community.³¹ The absence of a set chain of events and an open-ended storyline creates a history for the player, giving new and significant meaning to every adventure the user undertakes.³²

MMORPGs further attract players by allowing for thousands of hours of game play and providing an "infinite variety" of tasks, goals, and achievements for them to experience throughout the virtual world.³³ Indeed, because the game never ends for the player, it is impossible to "win" at an MMORPG. Most MMORPGs, including WoW, provide regular monthly content updates that add new creatures to kill, items to acquire, and dungeons to explore.³⁴

²⁸ World of Warcraft, Game Guide, General F.A.Q., *supra* note 23.

²⁹ *Id.*

³⁰ Jack Balkin, *Virtual Liberty: Freedom to Design and Freedom to Play in Virtual Worlds*, 90 VA. L. REV. 2043, 2043 (2004).

³¹ *See id.*

³² *Id.* at 2057 ("[An MMORPG] player is in a very different situation than someone who operates a pinball machine. . . [they] can take on multiple personas. . . they can create their own stories. . . and they can build things and form communities.").

³³ World of Warcraft, Game Guide, General F.A.Q., *supra* note 23., www.worldofwarcraft.com/info/faq/general.html (last visited Mar. 3, 2010).

³⁴ *See id.*

2. Avatar Creation and Improvement

Upon first logging into WoW, a player creates an avatar³⁵ and chooses from two rival factions: Horde or Alliance.³⁶ A player's faction choice is significant because a player can only speak to members of his own faction, and other facets of the game, such as the "level"³⁷ of skill a player may attain, the quests a player is eligible to attempt, and dungeons a player may enter, are organized by faction.³⁸ Once a player chooses a faction, the player must then choose a "class."³⁹ Players may further customize their avatar's appearance by choosing the avatar's race, gender, skin color, facial structure, and hair color/style.⁴⁰ WoW also enables players to make their avatars unique by adding facial markings, piercings, facial hair, or tusks.⁴¹

Players improve their avatars by killing monsters and completing quests.⁴² Once a player kills enough monsters or finishes enough quests, the avatar will gain a level or "level up."⁴³ "Leveling" a character requires a great deal of time.⁴⁴

³⁵ An avatar represents a player's physical representation in a virtual world. MICHAEL LUMMIS & ED KERN, *WORLD OF WARCRAFT MASTER GUIDE SECOND EDITION STRATEGY GUIDE 4* (Brady Games 2006).

Avatars are "onscreen characters controlled (and often designed) by the players." Theodore Westbrook, Note, *Owned: Finding a Place for Virtual World Property Rights*, 2006 MICH. ST. L. REV. 779, 780 (2006).

³⁶ Both horde and alliance players have five customizable race choices each with different strengths and weaknesses. World of Warcraft, Races, <http://www.worldofwarcraft.com/info/races/index.html> (last visited Jan. 2, 2010).

³⁷ In WoW, players are assigned a level that reflects how powerful an avatar is. See LUMMIS & KERN, *supra* note 35, at 5. Avatars begin at level one and the maximum level is level 80. World of Warcraft, Game Guide, Characters F.A.Q., <http://www.worldofwarcraft.com/info/faq/characters.html> (last visited Jan. 2, 2010). As one kills monsters and gains experience points, the player will reach the next level, or "level up," thus increasing the avatar's stats, abilities, and enabling the avatar to accomplish challenges it was not able to accomplish before. See LUMMIS & KERN, *supra* note 35, at 4-5.

³⁸ See World of Warcraft, Game Guide What is WoW, *supra* note 18.

³⁹ World of Warcraft, Game Guide, Classes F.A.Q., www.worldofwarcraft.com/info/faq/classes.html (last visited Jan. 20, 2009). When beginning the game, players choose between warrior, mage, rogue, druid, hunter, warlock, priest, paladin, rogue, and shaman classes. See *id.* Within each class, one is able to specialize in different talent trees. See *id.* This provides for greater diversity of skills among classes and allows a player to experience WoW game play from different perspectives See *id.*

⁴⁰ WorldofWarcraft.com, Game Guide Characters F.A.Q., <http://www.worldofwarcraft.com/info/faq/characters.html> (last visited Apr. 7, 2010).

⁴¹ *Id.*

⁴² See LUMMIS & KERN, *supra* note 35, at 5.

⁴³ See *id.*

⁴⁴ See Extreme Leveling, <http://www.extremeleveling.com/> (last visited Jan. 20, 2009) (Illustrating that creating a level 60 character often requires 19 days, or 456

Players may complete quests to earn experience points, or repeatedly kill a certain type of monster to “level” faster, which most gamers find rather dull compared to high-level game content. Avatars may also improve by acquiring high-level items⁴⁵ through professional skills such as crafting,⁴⁶ killing enemy bosses,⁴⁷ earning reputation awards,⁴⁸ engaging in player versus player combat, and purchasing items through an auction house.⁴⁹

Because items such as weapons and armor are needed to level an avatar and accomplish other in-game quests, these items are of great importance. They are so important that players often choose to purchase items in “real world” dollars instead of earning them within the game because many months of game play may be required to attain them.⁵⁰ In other MMORPG games, players resort to “camping”⁵¹ and “kill stealing”⁵² to obtain these items. However, in developing WoW, Blizzard took elaborate steps to prevent these cheating

hours of in-game time). However, Blizzard has recently reduced that time. *See* World of Warcraft, Game Guide, The Gods of Zul’Aman Patch 2.3, <http://www.worldofwarcraft.com/info/underdev/implemented/2p3.html> (last visited Jan. 20, 2009); Because players often find the required time commitment to level a maximum level character to be enormous, more experienced WoW players began offering guides, for a fee, to greatly reduce this time. *See* Extreme Leveling, *supra*.

⁴⁵ High-level items are separated into four separate classifications. In order of increasing rarity and power, they are uncommon items, rare items, epic items, and the coveted legendary items. *See* LUMMIS & KERN, *supra* note 35, at 14.

⁴⁶ Players can create powerful weapons and armor through professions such as blacksmithing, engineering, leatherworking, and tailoring. *See id.* at 246 (Brady Games 2006).

⁴⁷ World of Warcraft, Game Guide. Items F.A.Q., <http://www.worldofwarcraft.com/info/items/basics.html> (last visited Jan. 20, 2009).

⁴⁸ By increasing one’s reputation with a faction within the WoW virtual world, players gain access to reputation rewards that enable one to acquire high-level items. WorldofWarcraft, Game Guide, Reputations, <http://www.worldofwarcraft.com/info/basics/reputation.html> (last visited Jan. 2, 2010).

⁴⁹ An auction house serves as a clearing house for items that players acquire who would rather sell the items than use the items. World of Warcraft.com, Game Guide, Auction Houses, <http://www.worldofwarcraft.com/info/basics/auctionhouses.html> (last visited Jan. 2, 2010). Players may bid on or buy-out weapons, armor, or other in-game goods. *Id.*

⁵⁰ Leandra Lederman, “Stranger than Fiction”: Taxing Virtual Worlds, 82 N.Y.U. L. REV. 1620, 1628 (2007).

⁵¹ Camping is when a player monopolizes a group of monsters, killing them over and over again, in order to level-up or acquire loot. *See* LUMMIS & KERN, *supra* note 35, at 4.

⁵² World of Warcraft, Game Guide, Gameplay F.A.Q., www.worldofwarcraft.com/info/faq/gameplay.html (last visited Mar. 4, 2010). Kill stealing is rushing to kill a monster another player was attempting to kill in order to gain the experience or loot from that monster before the other player. *See* LUMMIS & KERN, *supra* note 35, at 5.

mechanisms by creating a pseudo first-in-time property right, where the first player or group to damage the virtual monster will receive both the experience and the loot.⁵³ Blizzard has also dispersed the dropping of high-level items across many monsters in the world, resulting in relatively little advantage in camping or racing to kill a specific monster type.⁵⁴

3. World of Warcraft as a Social Network

WoW is not only a complex video game, it is also a social network.⁵⁵ One of Blizzard's main goals in creating WoW was to encourage in-game socializing.⁵⁶ Players can join a guild⁵⁷ to socialize with other players, as well as to make group hunting easier.⁵⁸ "Guilds are an integral part of the game, allowing like-minded players to join together to achieve goals, not to mention getting to wear a really cool tabard."⁵⁹ In regular WoW play, players may group with up to five other players to complete quests.⁶⁰ One feature of the game, called an "instance,"⁶¹ allows for the creation of a sub-world within the larger WoW world. In

⁵³ WorldofWarcraft.com, Game Guide, Gameplay F.A.Q., *supra* note 52.

⁵⁴ *See id.* However, WoW has rare or "elite monsters" which appear from time to time in the game which are more difficult to kill, but almost always drop a high-level item. *See* LUMMIS & KERN, *supra* note 35, at 14. Once they are spotted, players will likely rush to kill the monster to retrieve its high value items.

⁵⁵ *See* David Sheldon, Comment, *Claiming Ownership, but Getting Owned: Contractual Limitations on Asserting Property Interests in Virtual Goods*, 54 UCLA L. REV. 751, 757 & n.27 (2007) ("[P]layers of the game enjoy a form of comity rarely seen in the real world; higher-level players go out of their way to tutor newbies and accompany them on quests. Deep friendships are forged. Relationships begin that flower into marriage, with Tauren brides and Undead grooms tying the knot in some virtual tavern in Thunder Bluff." (quoting Steven Levy, *Is World of Warcraft a Game?*, NEWSWEEK, Sept. 18, 2006, at 48) (alteration in original)); *see also* Balkin, *supra* note 30, at 2078 ("Some players already invest enormous amounts of time in these worlds; they make friends there and form attachments.").

⁵⁶ *See* World of Warcraft, Game Guide, What is WoW, *supra* note 18.

⁵⁷ A guild is an in-game association of players. *See* LUMMIS & KERN, *supra* note 35, at 121. Guilds may provide lower level players with a network to complete quests, receive discounted or free items from other players in the guild, and provide higher level players with a network to complete more difficult game content. *See id.* at 122-23.

⁵⁸ *See id.*

⁵⁹ *Id.* at 121. A "tabard is a wearable item that proudly displays your guild's chosen symbol and colors." World of Warcraft, Guilds, <http://www.worldofwarcraft.com/info/basics/guilds.html> (last visited Mar. 12 2010).

⁶⁰ World of Warcraft, Game Guide, What is WoW, *supra* note 18.

⁶¹ "An instance is a personal copy of the dungeon for you and your party. The only players in [the] instance will be yourself and the members of your party—no one else can enter your dungeon instance." World of Warcraft, Game Guide, Instancing, <http://www.worldofwarcraft.com/info/basics/instancing.html> (last visited Jan. 20, 2009).

an “instance,” players may band together in groups of up to forty players to kill monsters and complete quests that would otherwise be impossible to complete alone.⁶²

4. World of Warcraft’s Terms of Use and End User License Agreement

Because WoW has millions of players with deep social connections, a complicated reward structure, and a user-created in-game economy, the game requires rules to protect other players’ in-game rights. Thus, Blizzard provides a EULA⁶³ and TOU⁶⁴ to regulate player conduct. Under Section 2(A) of the TOU, Blizzard banned the use of “cheats, automation software (bots), hacks, mods or any other unauthorized third-party software designed to modify the World of Warcraft experience.”⁶⁵ Most significantly, under Section 4(a) of the EULA, Blizzard provided that “[a]ll title, ownership rights and intellectual property rights in and to the Game and all copies thereof. . . are owned or licensed by Blizzard.”⁶⁶ Further, Section 4 of the TOU provides, “[a]ll rights and title in and to the Service. . . are owned by Blizzard or its licensors.”⁶⁷ Indeed, Section 11 of the TOU specifies:

[y]ou may not purchase, sell, gift or trade any Account, or offer to purchase, sell, gift or trade any Account, and any such attempt shall be null and void. Blizzard owns, has licensed, or otherwise has rights to all of the content that appears in the Game. You agree that you have no right or title in or to any such content, including without limitation the virtual goods or currency appearing or originating in the game . . . you may not sell in-game items or currency for “real” money, or exchange those items or currency for value outside of the game.⁶⁸

Last, Section 7 of the TOU provides that “you acknowledge and agree that you shall have no ownership or other property interest in the Account, and you further acknowledge and agree that all rights in and to the Account are and shall forever be

⁶² *See id.*

⁶³ World of Warcraft, World of Warcraft End User License Agreement, <http://www.worldofwarcraft.com/legal/eula.html> (last visited July 29, 2008).

⁶⁴ World of Warcraft, World of Warcraft Terms of Use, www.worldofwarcraft.com/legal/termsofuse.html (last visited July 29, 2008).

⁶⁵ *Id.*

⁶⁶ World of Warcraft, World of Warcraft End User License Agreement, *supra* note 63.

⁶⁷ World of Warcraft, World of Warcraft Terms of Use, *supra* note 64.

⁶⁸ *Id.*

owned by and inure to the benefit of Blizzard.”⁶⁹ These EULA and TOU provisions grant the copyright holder, Blizzard, title to anything the user procures within the game, thus arming Blizzard with a powerful weapon against any claim the user may have to his virtual commodities.⁷⁰ The more restrictive Blizzard makes its EULA, and the more rights Blizzard attempts to withhold from its customers, the more difficult it becomes for WoW users to claim any property rights over their in-game commodities.

Blizzard inhibits virtual property rights and bans account sales in order to protect itself from black market transactions.⁷¹ According to Blizzard, there are two problems with black market transactions. First, if players were allowed to buy a high-level avatar, the player would spend less money on the subscription fees required to level that avatar through game play.⁷² Second, with more users playing high-level characters, Blizzard would need to create more high-level content to keep those players satisfied.⁷³ For Blizzard to be profitable, it must retain a high-level of monthly subscribers who spend a great deal of time experiencing the virtual world. If black market transactions were allowed, Blizzard would receive less money from monthly subscription fees and would have to expend greater resources on content updates, because it would have to update WoW’s content and storyline more frequently to keep it new and challenging.⁷⁴ Blizzard’s costs would increase while its revenues would decrease.

Thus, to give force to these EULA and TOU provisions, Blizzard penalizes violating players who lessen the gaming experience for other users.⁷⁵ Penalties include a warning for a minor account violation, a brief suspension for moderately severe violations, and account closure for the most severe

⁶⁹ *Id.*

⁷⁰ See Joshua A. T. Fairfield, *Virtual Property*, 85 B.U. L. REV. 1047, 1082 (2005).

⁷¹ Jaime J. Kayser, Note, *The New New-World: Virtual Property and the End User License Agreement*, 27 LOY. L.A. ENT. L. REV. 59, 73-74 (2007).

⁷² *Id.* at 73.

⁷³ *Id.*

⁷⁴ See *id.*

⁷⁵ Blizzard, Account Penalties, http://us.blizzard.com/support/article.xml?locale=en_US&articleId=20221&rhtml=y (Blizzard takes “disciplinary action . . . against disruptive players who are causing damage to other’s play experiences or the service itself.”) (last visited Jan. 20, 2009).

violations.⁷⁶ Blizzard has a reputation for aggressively responding to EULA and TOU violations.⁷⁷ In November 2006, Blizzard banned 105,000 accounts for selling virtual items for “real world” currency.⁷⁸ More strikingly, Blizzard considered canceling accounts of guild leaders trying to recruit for a guild catering to “gay, lesbian, bisexual, and transgendered individuals,” out of a concern that other game players might respond inappropriately.⁷⁹ Blizzard ultimately allowed the guild to continue recruiting,⁸⁰ however, Blizzard clearly takes a proactive approach to ensure that players’ “real world” rights do not transfer to their “virtual world” pursuits. A prime example of Blizzard’s aggressive defense of itself through its EULA and TOU is the penalties it meted out to players running Glider.

B. MDY Industries and WoWGlider

MDY Industries is the creator and owner of Glider,⁸¹ a third-party program⁸² that plays WoW while the user is away from his computer.⁸³ Glider’s sophistication allows it to undertake several surprisingly complex tasks. “It grinds, it loots, it skins, it heals, it even farms soul shards . . . without you.”⁸⁴ Since MDY began selling Glider in June 2005, “it has sold some 100,000 copies.”⁸⁵ MDY advertises and tailors Glider not to new players, but to experienced ones who want to level-up a new avatar quickly.

⁷⁶ See *id.* Blizzard will close accounts when “a player has excessively and/or grossly violated [its] policies” and when a player “insists on negatively affecting other players’ enjoyment of the game or harming the service itself.” *Id.* Blizzard rarely closes accounts. *Id.*

⁷⁷ See Sheldon, *supra* note 55, at 769.

⁷⁸ *Id.*

⁷⁹ *Id.* at 769-70.

⁸⁰ *Id.* at 770.

⁸¹ MDY Indus., LLC v. Blizzard Entm’t, Inc., No. CV-06-02555-PHX-DGC, 2008 WL 2757357 (D. Ariz. July 14, 2008) at *1.

⁸² A third-party program is any program developed by someone other than the original software developer, which modifies the original program. Blizzard, Hacks and Third-party Programs, http://us.blizzard.com/support/article.xml?locale=en_US&articleId=21133 (last visited Jan. 2, 2010).

⁸³ Glider, www.mmoglider.com/default.aspx?LS=54694 (last visited Jan. 20, 2009).

⁸⁴ *Id.*

⁸⁵ MDY Indus., LLC, 2008 WL 2757357, at *1; Glider can be purchased for \$25.00. Glider, Frequently Asked Questions, www.mmoglider.com/FAQ.aspx (last visited Jan. 20, 2009).

Best priest just quit your guild, but got no good recruits? Want to find out if you should have picked a mage instead of a warlock, but don't want to spend all that hard . . . game time again? Want to get some rogue-riffic revenge on those guys sneaking up on you in Battlegrounds? Those are the problems that the Glider solves.⁸⁶

WoW players can tailor Glider to their own preferences, instructing it to accomplish specific tasks, such as killing a particular monster.⁸⁷ Once the player instructs Glider, the program works automatically, allowing the player to return to his computer later and resume playing with the added experience and valuable items Glider earned in the meantime.⁸⁸ All a player has to do after launching Glider is to locate an area of monsters to kill, indicate to Glider the radial area the player wants his avatar to patrol, and specify the monsters the player wants to kill.⁸⁹

Players who really want to take full advantage of Glider can “dual box,”⁹⁰ which allows a player to have one Glider account active on more than one computer at the same time. Before Blizzard took an active interest in Glider use, Glider had become so widely utilized within avid gaming circles that software developers created third-party add-ons for Glider itself.⁹¹ Thus, there was a third-party program for the third-party program.

MDY recognizes that Glider, its program, is against Blizzard's TOU,⁹² and tells its customers as much: “If you are detected using Glider, your account will be suspended for 72 hours and very likely banned completely.”⁹³ MDY further warns its customers that they use Glider at their own risk.⁹⁴ MDY also has a community forum that, in part, is used to advise its users of account closings that may be due to Glider use.⁹⁵ These bans often occur in waves when Blizzard changes its monitoring

⁸⁶ Glider, Frequently Asked Questions, *supra* note 85.

⁸⁷ *See id.*

⁸⁸ *See id.*

⁸⁹ *See id.*

⁹⁰ Dual boxing, or multiboxing describes one player using multiple computers at one time to be active on more than one account at one time. WoWWiki, Multiboxing, <http://www.wowwiki.com/Multiboxing> (last visited Jan. 20, 2009).

⁹¹ Glider, Forums, Best Addons for use with Glider, <http://vforums.mmoglider.com/showthread.php?t=80> (last visited Jan. 6, 2010).

⁹² Glider, Frequently Asked Questions, *supra* note 85.

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ Glider, Forums, Ban Wave in Progress, May 20, 2008, vforums.mmoglider.com/showthread.php?t=148301 (last visited Jan. 6, 2010).

“Warden”⁹⁶ program or updates the game client and scans a user’s computer⁹⁷ before Glider changes its detection evasion coding.⁹⁸

In order to facilitate a player’s Glider use, Glider has defense mechanisms to lower Blizzard’s detection rate of the program.⁹⁹ Glider is able to evade detection when Blizzard searches a user’s computer for illegal third-party programs.¹⁰⁰ This feature is what makes Glider such a difficult problem for Blizzard to solve, thus causing Blizzard to divert resources from improving the game to combat Glider.¹⁰¹

The legal implications of the Glider program under the Copyright Act arose in *MDY Industries, LLC v. Blizzard Entertainment, Inc.*,¹⁰² the subject of Part III.

⁹⁶ Warden is a program used by Blizzard to detect when a player is using an unauthorized third-party program. *MDY Indus., LLC v. Blizzard Entm’t, Inc.*, No. CV-06-02555-PHX-DGC, 2008 WL 2757357, at *11 (D. Ariz. July 14, 2008). Warden detects Glider in two ways. *Id.* First, it will scan the player’s computer to locate the Glider program. *Id.* If Warden does detect Glider, Blizzard will deny the player access to the game server. *Id.* Second, Warden scans the user’s computer while playing WoW as well. *Id.* Again, if Warden detects Glider, “Blizzard revokes access to the game.” *Id.*

⁹⁷ Players consent to these computer scans under Section 17(A) of the TOU. World of Warcraft, World of Warcraft Terms of Use, *supra* note 64. Players who do not consent are unable to launch the game. The enforceability of these online click-wrap agreements is outside the scope of this note. *See generally* Kaustuv M. Das, *Forum Selection Clauses in Consumer Clickwrap and Browsewrap Agreements and the “Reasonably Communicated” Test*, 77 WASH. L. REV. 481 (2002). In Blizzard’s TOU, Blizzard capitalized the entire provision. However, that may not matter. For example, to prove a point, PC Pitstop, an internet site offering antivirus and internet speed scans, provided in one of its EULAs that if anyone emailed a certain email address listed in its EULA, the sender would receive \$1000. Larry Magid, *It Pays to Read License Agreements*, PC PITSTOP NEWSL., Feb. 16, 2005, <http://www.pcpitstop.com/spycheck/eula.asp>. After four months and three thousand downloads, someone finally wrote in to receive the money. *Id.* Further illustration of the absurdity of click-wrap contracts can be found in Google’s Terms of Service that includes a clause barring any person not of legal age from using “any of Google’s Web properties.” Chris Soghoian, *Google: No Kids Allowed*, CNET NEWS, Mar. 27, 2008, http://news.cnet.com/8301-13739_3-9902548-46.html. Recently, Apple also mistakenly included a clause in its EULA for the Windows version of Safari (an internet browser), providing that the browser was not to be installed on a PC. Jeff Hinman, *I’m a EULA. I’m a Contract. Apple Fumbles, Exposes EULA Dangers*, LEGALITY, Apr. 30, 2008, available at <http://www.thelegality.com/2008/04/30/i%E2%80%99m-a-eula-i%E2%80%99m-a-contract-apple-fumbles-exposes-eula-dangers/> (last visited Jan. 6, 2010).

⁹⁸ *See* On Warden, <http://onwarden.blogspot.com/2008/05/may-20th-ban-wave-wow-242.html> (May 20, 2008, 6:14PM).

⁹⁹ *See* Glider, Frequently Asked Questions, *supra* note 85.

¹⁰⁰ *See id.*

¹⁰¹ *MDY Indus., LLC*, at *15.

¹⁰² No. CV-06-02555-PHX-DGC, 2008 WL 2757357 (D. Ariz. July 14, 2008).

III. *MDY INDUSTRIES, LLC v. BLIZZARD ENTERTAINMENT, INC.*

The main copyright issue that arose in MDY Industries was whether a user infringed Blizzard's copyright in WoW on the ground that whenever a user launched Glider in conjunction with WoW, the user created unauthorized "copies" of those programs in the computer's RAM because Glider use violated WoW's EULA and TOU.

A. *Facts/Claims*

On October 25, 2006, Blizzard representatives traveled to the home of MDY Industries's founder Michael Donnelly and advised him that MDY's Glider sales violated Blizzard's copyright in WoW.¹⁰³ Blizzard told Donnelly that if he did not agree to stop selling Glider, they would immediately file a lawsuit against him and MDY.¹⁰⁴ Donnelly refused to stop selling the program, and Blizzard filed suit in Arizona federal district court.¹⁰⁵

Blizzard claimed that Glider diminished WoW's value, influenced players to deactivate their WoW accounts, and decreased Blizzard's revenue.¹⁰⁶ Due to WoW's meticulously orchestrated competitive balance, Blizzard asserted that players who used Glider were able to unfairly complete tasks

¹⁰³ *Id.* at *2.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *Id.* at *1. Players have terminated their accounts due to other players' Glider use. See EDWARD CASTRONOVA, EFFECTS OF BOTTING ON WORLD OF WARCRAFT® 5 (Nov. 13, 2007), http://virtuallyblind.com/files/mdy/blizzard_msj_exhibit_7.pdf. First, players that do not use Glider feel that it is unfair for Glider users who are violating the EULA and TOU to advance through the game more quickly than the users who do not use Glider. See *id.* Second, Glider users, by playing more than humanly possible, negate Blizzard's intent for certain items to cost a certain amount by flooding the in-game WoW economy. See *id.* at 6-8. Thus, a player who does not use Glider will only realize a marginal market return on any in-game goods the user decides to sell. See *id.* at 9. Third, an additional market distortion comes in the form of gold farming, where Glider users sell their in-game currency for "real world" money. See *id.* at 7. Because the average player realizes less of a return from his farming due to Glider users, a user may be forced to buy gold from Glider users to purchase in-game goods, thus decreasing the amount of real-world dollars available for WoW subscription fees, which may force account cancellations. *Id.* at 11. Fourth, Glider use also increases Blizzard's costs of providing WoW by requiring greater customer service costs arising from Glider use complaints, and increasing the cost to technologically eliminate Glider. *Id.* at 14-16. Last, Blizzard markets WoW as an immersive, role-playing, social, in-game experience. *Id.* at 16-17. Glider use is detrimental to this vision in that it incentivizes players to use the game while not at their computers. *Id.* at 18.

throughout the game more quickly than Blizzard intended, and that Glider users lessened the gaming experience for players who did not use Glider.¹⁰⁷ Blizzard further alleged that Glider facilitated “gold farming,”¹⁰⁸ and the selling of in-game gold to other users.¹⁰⁹ Gold farming, like the use of third-party programs, is also expressly prohibited by the TOU.¹¹⁰

Specifically, the copyright issue that arose in *MDY* was that whenever a user launched Glider in conjunction with WoW, the user created unauthorized “copies” of those programs in the computer’s RAM because Glider was against WoW’s EULA and TOU.¹¹¹ The Ninth Circuit determined in *MAI* that a work copied from software to RAM was sufficiently “fixed in a tangible medium of expression” so that it could be considered a “copy” for purposes of the Copyright Act, because it was present for a period longer than a “transitory duration.”¹¹²

Nevertheless, software “owners” are permitted to copy software to RAM. Section 117 of the Copyright Act permits the “owner” of a computer program to “copy” software to RAM if the copy was created as an “essential step” in using the program.¹¹³ However, the users in the *MAI* case were not entitled to such a defense because the users were not software

¹⁰⁷ *MDY Indus.*, 2008 WL 2757357 at *1.

¹⁰⁸ Gold farming is “an Internet-age phenomenon in which players in less developed countries collect and sell virtual gold . . . to wealthier gamers in the developed world. This enables gamers who have the means to buy virtual gold to get ahead in the games without actually having to accomplish the grunt work.” Dave Rosenberg, *China Bans Online ‘Gold Farming’*, CNET NEWS, June 29, 2009, http://news.cnet.com/8301-13846_3-10275180-62.html. Gold farming will often result in bans from the game. *Id.* Gold farming is a one billion dollar industry. Posting of Michael Sacco to WoW.com, <http://www.wow.com/2009/06/29/china-bans-gold-farming> (June 29, 2009 7:40 PM). “Some half a million people in developing nations are working at least part time [farming gold.]” Posting of Mark Hefflinger, to digitalmediawire.com (Aug. 25, 2008, 11:59AM). While gold farming has been big business, China recently banned online gold farming. *See* Rosenberg, *supra*. Approximately four of every five gold farmers live in China. *See* Sacco, *supra*.

¹⁰⁹ *MDY Indus., LLC*, 2008 WL 2757357, at *1.

¹¹⁰ *Id.* at *1; *see also* World of Warcraft, World of Warcraft Terms of Use, *supra* note 64 (Section 9(B)(vii)).

¹¹¹ World of Warcraft, World of Warcraft Terms of Use, *supra* note 64 (Section 17(A)).

¹¹² *MAI Systems Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 518-19 (9th Cir. 1993) (holding that a computer maintenance company running MAI’s program on its client’s computers as part of a repair job had created an unauthorized “copy” of the software in the RAM of the client’s computer).

¹¹³ *MDY Indus.*, 2008 WL 2757357 at *6. “It is not an infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided . . . that such a new copy or adaptation is created as an essential step in the utilization of the computer program. . . .” 17 U.S.C. § 117 (2006).

“owners.”¹¹⁴ Thus, the *MAI* court ruled that users created unauthorized “copies” by merely using software that was then copied into the computer’s RAM.¹¹⁵ Therefore, the unauthorized copying in *MAI* constituted copyright infringement by a non-owner, as non-owners are not entitled to a Section 117 defense.¹¹⁶

Congress intended to render “owners” free from copyright liability for the lawful purchase and use of software, when the software’s use in its intended manner involves the copying of software to RAM.¹¹⁷ If an “owner” exceeded his license and unlawfully copied software to RAM, the software provider may have a remedy in contract, but not in copyright.¹¹⁸

In *MDY*, Blizzard argued that the court should find MDY liable for contributory¹¹⁹ and vicarious copyright infringement¹²⁰ because the individuals who purchased WoW were not “owners” of the game, but instead were licensees, who may not take advantage of the Section 117 safe harbor.¹²¹ Under

¹¹⁴ *MAI Systems Corp.*, 991 F.2d at 518 n.5.

¹¹⁵ *Id.* at 518-19.

¹¹⁶ *Id.* at 518 n.5 & 518-19.

¹¹⁷ “Because the placement of a work into a computer is the preparation of a copy, the law should provide that persons in rightful possession of copies of programs be able to use them freely without fear of exposure to copyright liability.” CONTU REPORT, *supra* note 1, at 13. Indeed, the situation in MDY is exactly the situation that CONTU commission intended to protect in its recommendations to Congress concerning the Copyright Act. *Id.* The commission stated,

Obviously creators, lessors, licensors, and vendors of copies of programs intend that they be used by their customers, so that rightful users would but rarely need a legal shield against potential copyright problems. It is easy to imagine, however, a situation in which the copyright owner might desire, for good reason or none at all, to force a lawful owner or possessor of a copy [of a program] to stop using a particular program. One who rightfully possesses a copy of a program, therefore, should be provided with a legal right to copy it to that extent which will permit its use by that possessor.

CONTU REPORT *supra* note 1, at 13.

¹¹⁸ CONTU REPORT, *supra* note 1, at 13-14 (“Should proprietors feel strongly that they do not want rightful possessors of copies of their programs to prepare such adaptations, they could . . . make such desires a contractual matter.”).

¹¹⁹ “A person commits contributory copyright infringement by ‘intentionally inducing or encouraging direct infringement.’” *MDY Indus., LLC*, 2008 WL 2757357 at *3 (quoting *MGM Studios Inc. v. Grokster*, 545 U.S. 913, 930 (2005)).

¹²⁰ “A person commits vicarious infringement ‘by profiting from direct infringement while declining to exercise a right to stop or limit it.’” *Id.* at *3 (quoting *MGM Studios Inc. v. Grokster*, 545 U.S. at 930). The court also ruled on summary judgment for MDY’s alleged infringement of Sections 1201(a)(2) and 1201(b)(1) of the Digital Millennium Copyright Act, *see id.* at *10-14, tortious interference with contract, *see id.* at *14-16, and unjust enrichment, *see id.* at *17. These rulings are outside the scope of this Note.

¹²¹ *MDY Indus., LLC*, 2008 WL 2757357 at *3.

Blizzard's argument, if the purchasers were licensees, they would not be entitled to a Section 117 defense, and thus MDY may be liable for the underlying direct copyright violations.

Blizzard framed its argument in the context of the Ninth Circuit's decisions in *MAI* and *Wall Data*, which held that when software providers utilize a EULA to restrict a purchaser's property interest in software to that of a licensee, the computer program purchasers are not "owners" of the software and are precluded from utilizing Section 117's shield.¹²²

Specifically, the *Wall Data* court held that if the copyright holder clearly stated that it only granted the purchaser a license to the software copy, and imposed significant restrictions on that purchaser's ownership interests in terms of redistribution or copying, the purchaser was only licensed to use the software and could not be considered an "owner" under Section 117.¹²³ In *Wall Data*, the Los Angeles County Sheriff's Department contracted with Wall Data, a developer and seller of computer programs, to purchase eight CD-ROMs that contained Wall Data's terminal emulation program,¹²⁴ "RUMBA."¹²⁵ Each CD-ROM contained two hundred fifty licenses, for a total of two thousand licenses.¹²⁶ However, the parties disagreed as to the relationship between the copies of the software and the license. The Sheriff's Department claimed that it purchased 2,000 *copies* of RUMBA, while Wall Data contended that the Sheriff's Department only bought 2,000 *licenses* of RUMBA.¹²⁷ Subsequently, the Sheriff's Department purchased additional RUMBA licenses, which brought it to a total of 3,663 licenses.¹²⁸ In order to facilitate the opening of its new detention facility, the Sheriff's Department decided to simultaneously install the RUMBA software onto all

¹²² *See id.*

¹²³ *Wall Data Inc. v. Los Angeles Cty. Sheriff's Dep't.*, 447 F.3d 769, 785 (9th Cir. 2006).

¹²⁴ A terminal emulator is a program that makes a computer "appear to look like another, usually older type of terminal so that a user can access programs originally written to communicate with the other terminal type." SearchNetworking.com, What is Terminal Emulation?, http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci213121,00.html (last visited Jan. 20, 2009).

¹²⁵ *Wall Data*, 447 F.3d at 774.

¹²⁶ *Id.*

¹²⁷ *Id.* at 774 n.2.

¹²⁸ *Id.* at 774.

of its 6,007 computers in the new facility, exceeding the 3,663 purchased licenses.¹²⁹

Wall Data learned of the Sheriff's Department's actions and sued for copyright infringement.¹³⁰ Wall Data claimed that because the Sheriff's Department over-installed the RUMBA software onto its computers, it violated the terms contained in Wall Data's shrink-wrap,¹³¹ click-wrap,¹³² and volume license agreement.¹³³ Therefore, because Wall Data clearly stated that it only granted the purchaser a license to the software copy, and imposed significant restrictions on that purchaser's ownership interests in terms of redistribution or copying, the court considered the purchaser an unprotected licensee rather than an "owner" who could avail itself of a Section 117 safe harbor defense.¹³⁴

Under the *MAI* and *Wall Data* precedent, the *MDY* court awarded Blizzard summary judgment¹³⁵ on its claims for contributory and vicarious copyright infringement. Section 106 of the Copyright Act grants the "owner" of a copyright the exclusive right to "reproduce" the copyrighted work or to prepare derivative works¹³⁶ based upon the work, or to distribute copies of work to the public.¹³⁷ Further, under Section

¹²⁹ *Id.* at 774-75.

¹³⁰ *Id.*

¹³¹ A "shrink-wrap license" is a "form on the packing or on the outside of the CD-ROM containing the software which states that by opening the packaging or CD-ROM wrapper, the user agrees to the terms of the license." *Id.* at 774 n.4.

¹³² A "click-through license" is a "form embedded in computer software which requires the person initially installing the software onto a computer to affirmatively click a box or an 'accept' button indicating that the user accepts the terms of the license in order to complete the software installation and to use the software after it is installed." *Id.* at 775 n.5.

¹³³ *Id.* at 775.

¹³⁴ *Id.* at 785.

¹³⁵ Summary judgment may be granted if "there is no genuine issue as to any material fact" and "the moving party is entitled to judgment as a matter of law." FED. R. CIV. P. 56(c). A party seeking summary judgment "always bears the initial responsibility of informing the district court of the basis for its motion, and identifying those portions of . . . [the record] which it believes demonstrate the absence of a genuine issue of material fact." *Celotex Corp. v. Catrett*, 477 U.S. 317, 323 (1986) (citations omitted).

¹³⁶ A "derivative work" is a "work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a 'derivative work.'" 17 U.S.C. § 101 (2006).

¹³⁷ *Id.* § 106; see also *MDY Indus., LLC v. Blizzard Entm't, Inc.*, No. CV-06-02555-PHX-DGC, 2008 WL 2757357, at *2 (D. Ariz. July 14, 2008).

501 of the Copyright Act, anyone who violates one of the exclusive rights granted to the copyright holder in Section 106 is an infringer of the copyright.¹³⁸

Applying the *Wall Data* reasoning to *MDY* and looking to the restrictions on ownership Blizzard dictated in its EULA and TOU, the *MDY* court determined that first, Blizzard's EULA stated that it granted a limited license,¹³⁹ and second, Blizzard imposed significant restrictions on the transfer and use of the game client software.¹⁴⁰ In further support of Blizzard's argument in favor of classifying a WoW purchaser as a licensee instead of an "owner," Blizzard pointed out the notices on WoW's box, the paper copy of the EULA in WoW's box, and the online notices that appeared when the user installed the WoW game client, which all notified the purchaser of his limited rights in the game as licensees.¹⁴¹ Therefore, when users launched WoW using Glider, they exceeded Blizzard's license and created infringing copies of the game.¹⁴² Just as in *Wall Data*, the Court refused to afford MDY a Section 117 defense because of Blizzard's restrictive EULA language.¹⁴³

Although *Wall Data*'s result may arguably be sound under its facts, *MDY*'s facts make clear *Wall Data*'s faulty underlying reasoning. Ultimately, allowing a copyright holder to restrict a consumer's legal rights under the Copyright Act by merely including restrictive language in a click-wrap or shrink-wrap contract is inequitable. As seen in today's marketplace, because a shrink-wrap EULA may be the beginning and end of the inquiry in determining ownership, video game manufacturers, music companies, and other software providers only need to include restrictive boilerplate language in their EULAs and TOUs to hold purchasers who exceed these license terms liable for copyright infringement.¹⁴⁴ Illustrating the unfairness of this result, as one commentator suggested, "[n]either the traditional norms of contract law nor the policies behind the protection of intellectual property support

¹³⁸ 17 U.S.C. § 501; see also *MDY*, 2008 WL 2757357, at *2.

¹³⁹ *MDY*, 2008 WL 2757357, at *8.

¹⁴⁰ *Id.* at *9.

¹⁴¹ *Id.*

¹⁴² *Id.* at *3.

¹⁴³ *Id.* at *8-9.

¹⁴⁴ Sherwin Siy, *MDY v. Blizzard: Cheating at WoW May Be Bad, but It's Not Copyright Infringement*, May 5, 2008, <http://www.publicknowledge.org/node/1546> (last visited Mar. 1, 2010).

enforcement of agreements that exist primarily to frustrate public legislation.”¹⁴⁵ Here, Congress intended Section 117 of the Copyright Act to protect software purchasers from copyright liability for making incidental copies of software to RAM.¹⁴⁶ A software provider should not be allowed to thwart that legislative purpose by providing in its EULAs that they sell nothing, and license everything.

Further illustrating *MAI*'s and *Wall Data*'s problematic owner-licensee distinction is the experience of an ordinary purchaser from that purchaser's perspective. One who purchases a copy of *WoW* from her local electronics retailer and leaves the store with the software, never obligated to return the software to the store as if she had temporarily leased the software, would never think she only purchased a software license. On the contrary, the customer would think she purchased a copy of software.¹⁴⁷ The person could dispose of the software copy as she chooses by throwing it in the trash, giving it to a friend, or installing it on her computer.¹⁴⁸ All of these activities are consistent with ownership powers.¹⁴⁹

Admittedly, allowing software providers to limit a purchaser's rights has several benefits. For example, it is a simple rule to administer. If the software provider implements restrictive language in its EULA, the court need not look elsewhere to discern the purchaser's rights. Further, if the software provider only wants to sell licenses to its software and courts begin holding that the providers are actually selling ownership rights to the particular copy, software providers may stop selling certain software altogether or may adjust prices or other terms.

Nevertheless, the court's method of looking only to the software provider's restrictive EULA language as the dispositive issue in classifying a purchaser as an "owner" or a licensee of the software is ultimately inequitable.

¹⁴⁵ Elizabeth I. Winston, *Why Sell What You Can License? Contracting Around Statutory Protection of Intellectual Property*, 14 *GEO. MASON L. REV.* 93, 94 (2006).

¹⁴⁶ CONTU REPORT, *supra* note 1, at 13-14.

¹⁴⁷ *MDY*, 2008 WL 2757357, at *9.

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

IV. *MDY* WAS INCORRECT

The *MDY* court's ruling was incorrect for three reasons. First, the First Sale Doctrine under *Wise* more equitably allocates the rights between software purchasers and software providers by focusing on the economic realities of the underlying transaction to discern whether a software purchaser is an "owner" or licensee of the software. Second, Lockean labor desert theory counsels in favor of more substantial, though not absolute, protection of the users' rights when considering how purchasers utilize their software after the transaction. Third, copyright law's underlying policies suggest greater protection for purchasers because the illusory rights the Ninth Circuit afforded to software users stymie creative development far out of proportion to what Congress intended under the Copyright Act. In *MDY*, the court afforded no rights to software users, but instead merely deferred to what Blizzard provided to the purchasers in its EULA.¹⁵⁰ This is exactly the situation that Congress feared and enacted Section 117 to prevent.¹⁵¹

A. *First Sale Doctrine*

1. Precedent

MDY argued that notwithstanding the *Wall Data* decision, *WoW* purchasers were software "owners" rather than licensees under the First Sale Doctrine as articulated in *Wise*,¹⁵² and the copying of software to RAM was an "essential step" in using the game client software.¹⁵³ According to *MDY*, under Section 117 of the Copyright Act, the software purchasers were authorized to copy the game client software to RAM through the license they acquired when they bought the game.¹⁵⁴ Thus, by creating the RAM copies, *Glider* users did not infringe upon Blizzard's copyright; they only breached a contract.¹⁵⁵ Therefore,

¹⁵⁰ *Id.* at *8-9.

¹⁵¹ See *supra* note 118 and accompanying text.

¹⁵² *MDY*, 2008 WL 2757357 at *3.

¹⁵³ *Id.* at *8.

¹⁵⁴ *Id.* at *3.

¹⁵⁵ *Id.* This distinction is significant because breach of contract damages are generally limited to the value of the actual loss caused by the breach. 24 RICHARD A. LORD, WILLISTON ON CONTRACTS § 65:1, at 213 (4th ed. 2002). In contrast, copyright damages include the copyright owner's actual damages and any additional profits of the infringer or statutory damages. See 17 U.S.C. § 504 (2006).

Glider users did not infringe Blizzard's copyright even when using Glider in violation of Blizzard's EULA and TOU.¹⁵⁶

Wise and *Vernor* defended against the frustration of legislative intent through restrictive private contractual language as seen in *MDY*. In *Wise*, Woodrow Wise, Jr. operated a business that distributed lists of copyrighted movies that he sold to film enthusiasts for home use.¹⁵⁷ Each list included a provision stating, "used film for sale. Sold from one private movie collector to another for home showing only. No rights given or implied."¹⁵⁸ Witnesses who testified against Wise stated that the movie studios that held the copyright to these films did not sell the films to the purchasers, but only licensed their use for specific purposes for a limited time.¹⁵⁹ The studio licenses provided that the studio retained all rights in and title to the movies, and the license further restricted the licensees to only use the movies for their personal use.¹⁶⁰ Further, the copyright holders distributed the films pursuant to a theatrical license agreement, which stated "[t]he distributor grants the Exhibitor and the Exhibitor accepts a limited license under the respective copyrights of the motion picture . . . to exhibit said motion picture."¹⁶¹ The United States criminally prosecuted Wise for copyright infringement due to Wise's unauthorized film sales in violation of these restrictive licensing terms.¹⁶² Just as in *MDY*, Wise argued in his defense that he was an "owner" and not a licensee of the films.¹⁶³

In contrast to the Ninth Circuit's restrictive decisions in *MAI* and *Wall Data*, the *Wise* Court invoked the First Sale Doctrine.¹⁶⁴ The First Sale Doctrine provides that the Copyright

¹⁵⁶ *MDY*, 2008 WL 2757357, at *8.

¹⁵⁷ *United States v. Wise*, 550 F.2d 1180, 1183-84 (9th Cir. 1977).

¹⁵⁸ *Id.* at 1184.

¹⁵⁹ *Id.*

¹⁶⁰ *Id.*

¹⁶¹ *Id.* at 1190.

¹⁶² *Id.* at 1185.

¹⁶³ *Id.*

¹⁶⁴ "The first sale doctrine is a narrow limitation on a copyright holder's rights." *Vernor v. Autodesk*, 555 F. Supp. 2d 1164, 1168 (W.D. Wash. 2008). Under the Copyright Act, a copyright holder has the exclusive right to copy his work, 17 U.S.C. § 106(1) (2006); to prepare derivative works, *id.* § 106(2); and to distribute copies of his work, *id.* § 106(3). The first sale doctrine was first articulated in *Bobbs-Merrill Co. v. Strauss*, where a book publisher attempted to restrict resale of a book through a license agreement prohibiting resale for less than one dollar. *Bobbs-Merrill Co. v. Strauss*, 210 U.S. 339, 341 (1908). Defendants sold the book for 89 cents. *Id.* at 342. The court concluded, "[i]n our view the copyright statutes . . . do not create the right to impose, by

Act shall not “forbid, prevent or restrict the transfer of any copy of a copyrighted work the possession of which has been lawfully obtained.”¹⁶⁵ Thus, when a copyright owner first sells a copy of its copyrighted work, the owner is thereafter precluded from using his exclusive right of distribution to prevent the resale of that same copy.¹⁶⁶ The copyright holder still holds the exclusive right to reprint and copy its work, but the purchaser earns the right to sell the transferred copy. Indeed, “the copyright is distinct from the property which is copyrighted, and the sale of one does not constitute a transfer of the other.”¹⁶⁷ Therefore, under the First Sale Doctrine, if the purchaser breaches a contract by selling a copy of a copyrighted work he may be held liable for breach of contract, but not for copyright infringement.¹⁶⁸

Further, in contrast to the *Wall Data* and *MDY* decisions, which merely considered the software providers’ restrictive EULA language in determining ownership, the *Wise* court looked outside the “four corners” of the contract to discern the rights for which the parties actually bargained.¹⁶⁹ The *Wise* court found that most of *Wise*’s purchases were licenses because the transfer contracts between *Wise* and the film studios transferred only the rights to show or distribute the films for a limited period of time, and *Wise* was to return the films at the end of the license.¹⁷⁰ Even though some of the licenses did not expressly specify the copyright holder reserved title, the court concluded that such a clause was not necessary “where the general tenor of the entire agreement [was] inconsistent with such a conclusion.”¹⁷¹ Based on this reasoning, the *Wise* court found sales in two instances¹⁷² regardless of other limitations on use.¹⁷³

notice . . . a limitation at which the book shall be sold at retail by future purchasers, with whom there is no privity of contract.” *Id.* at 350.

¹⁶⁵ *Wise*, 550 F.2d at 1187; *see also* 17 U.S.C. § 109.

¹⁶⁶ *Id.* at 1187.

¹⁶⁷ *Id.* at 1187 n.9. “[O]wnership of a thing is always separate from ownership of the intellectual property embedded in a thing. Ownership of a book is not ownership of the intellectual property of the novel that the author wrote. The book purchaser owns the physical book, nothing more.” *Fairfield*, *supra* note 70, at 1096.

¹⁶⁸ *Wise*, 550 F.2d at 1187 n.10.

¹⁶⁹ *Id.* at 1190.

¹⁷⁰ *Id.*

¹⁷¹ *Id.* at 1191.

¹⁷² *Id.* at 1191-92.

¹⁷³ The limitations on use in these contracts were quite severe. *Id.* at 1192. In one agreement, Warner Brothers sold a print of “Camelot” to Vanessa Redgrave

The First Sale Doctrine as applied in *Wise* related to movie sales, but it has been recently applied by a Ninth Circuit district court to sales of computer software packages like the transactions in *MDY*.¹⁷⁴ In *Vernor*, Plaintiff Timothy Vernor, an eBay entrepreneur, lawfully purchased a used Autodesk software package at a garage sale, and auctioned it on eBay.¹⁷⁵ Included in the package was Autodesk's license agreement.¹⁷⁶ Autodesk sent notice to eBay and claimed that Vernor's sales violated Autodesk's copyright in its software;¹⁷⁷ eBay cancelled the auction.¹⁷⁸ Vernor sent eBay a counter-notice¹⁷⁹ asserting that the software package sale was lawful.¹⁸⁰ After no response from Autodesk, eBay resumed the auction.¹⁸¹

whereby Ms. Redgrave was to pay \$401.59 for the print. *Id.* According to the contract terms, Ms. Redgrave was required to have the print in her possession "at all times"; she was not allowed to sell, lease, license or loan the print; and was restricted from exhibiting it for profit. *Id.* The *Wise* court nevertheless determined this purchase to be a sale instead of a license. *Id.*

¹⁷⁴ See *Vernor v. Autodesk*, 555 F. Supp. 2d 1164 (W.D. Wash. 2008). Following *Wise*, the first sale doctrine has also recently been applied to protect a seller of promotional music CDs against a copyright infringement action. See *UMG Recordings, Inc. v. Augusto*, 558 F. Supp. 2d 1055, 1058 (C.D. Cal. 2008). In *UMG*, a promotional CD contained a license restricting its transfer, but did not contemplate the return of the CDs. *Id.* The court held that licensing language is not dispositive in creating a license; instead, "courts must analyze the 'economic realities' of the transaction." *Id.* at 1060 (citing to *Microsoft Corp. v. DAK Indus.*, 66 F.3d 1091, 1095 (9th Cir. 1995)). The court emphasized that "perpetual possession" without the copyright holder intending the item to be returned is a hallmark of ownership and a sale. *Id.* Thus, UMG still had the exclusive right to distribute and make copies of the copyrighted music, but the copy that the consumer purchased could be resold. See Matthew Schroettig, "Damn The Man!" *The Ability To Sell Second-Hand CDs*, THE LEGALITY, Oct. 16, 2008, <http://www.thelegality.com/2008/10/16/%E2%80%9Cdamn-the-man%E2%80%9D-the-ability-to-sell-second-hand-cds/> (last visited Mar. 3, 2010).

¹⁷⁵ *Vernor*, 555 F. Supp. 2d at 1165.

¹⁷⁶ *Id.* at 1165 n.1. "The License Agreement grants a nonexclusive, nontransferable license to use the enclosed program . . . according to the terms and conditions herein." The license imposed several restrictions on software purchasers such as limiting the number of computers on which the software may be installed, limiting the number of users, "software copying and copying of documentation" and prohibiting "rent, lease, or transfer [of] all or part of the Software, Documentation, or any rights granted hereunder to any other person without Autodesk's prior consent." *Id.* at 1166 (citation omitted) (internal quotation marks omitted).

¹⁷⁷ *Id.* at 1165.

¹⁷⁸ *Id.*

¹⁷⁹ Internet content providers such as eBay enjoy protection from secondary liability for copyright infringement through a take-down notice regime. 17 U.S.C. § 512(c) (2006). Generally, once a content owner sends notice to the content provider of the allegedly infringing content, the content provider will be immune from liability so long as the provider promptly disables access to the material, notifies the user that posted the allegedly infringing material, and did not have actual or constructive knowledge of the allegedly infringing material. *Id.*

¹⁸⁰ *Vernor*, 555 F. Supp. 2d at 1165.

¹⁸¹ *Id.*

Several years later, Vernor purchased three more Autodesk software packages at an office sale from CTA, an architecture firm.¹⁸² Again, the same process occurred: Vernor auctioned one of the software packages, Autodesk sent notice to eBay to cancel the auction, to which Vernor would respond with his own counter-notice, and the auction was reinstated.¹⁸³ However, when Autodesk objected to Vernor's fourth eBay software package auction, eBay suspended Vernor's eBay account for repeatedly infringing its policies by selling the copyrighted software.¹⁸⁴ Vernor filed a declaratory judgment action to establish the legality of the sales.¹⁸⁵

As in *Wise*, the *Vernor* court held that Vernor's sales were immunized under the First Sale Doctrine.¹⁸⁶ According to the court, "[t]he First Sale Doctrine permits a person who owns a lawfully-made copy of a copyrighted work to sell or otherwise dispose of the copy."¹⁸⁷ Thus, because Vernor lawfully owned the software packages when he purchased them from CTA, he could sell or dispose of them as he wished.¹⁸⁸ The *Vernor* court recognized that the first sale extinguished the copyright holder's ability to further control that copy's distribution.¹⁸⁹

The critical question for the *Vernor* court, as in *MDY*, was whether Autodesk sold the software packages to CTA or merely authorized a license.¹⁹⁰ Without a sale, Vernor would not have acquired ownership of the copy within the meaning of Section 109(a) of the Copyright Act, and therefore could not rely on the First Sale Doctrine.¹⁹¹ But if the transactions were sales instead of licenses, breaching the terms of the license would "give rise, at most, to a breach of contract claim."¹⁹²

¹⁸² *Id.*

¹⁸³ *Id.* at 1165-66.

¹⁸⁴ *Id.* at 1166.

¹⁸⁵ *Id.*

¹⁸⁶ *Id.* at 1168.

¹⁸⁷ *Id.* "Notwithstanding the provisions of section 106(3) [17 USCS § 106(3)], the owner of a particular copy or phonorecord lawfully made under this title, or any person authorized by such owner, is entitled, without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy or phonorecord." *Id.* (quoting 17 U.S.C. § 109(a)).

¹⁸⁸ *Id.* "For example, the first sale doctrine permits a consumer who buys a lawfully made DVD copy of 'Gone With the Wind' to resell the copy, but not to duplicate the copy." *Id.*

¹⁸⁹ *Id.* (citing *United States v. Wise*, 550 F.2d 1180, 1187 (9th Cir. 1977)).

¹⁹⁰ *Id.* at 1169.

¹⁹¹ *Id.* at 1168 (citing *Quality King Distribs., Inc. v. Lanza Research Int'l, Inc.*, 523 U.S. 135, 146-47 (1998)); see also *Wise*, 550 F.2d at 1188-89.

¹⁹² *Vernor*, 555 F. Supp. 2d at 1169 (citations omitted)

In contrast to *Wall Data*, where the court only looked to the EULA language to determine whether the purchaser owned or was licensed to use the software,¹⁹³ the *Vernor* court, following *Wise's* reasoning, stated that there was no “bright line rule” to easily determine a purchaser’s classification as an “owner” or a licensee.¹⁹⁴ Further, the *Vernor* court explicitly refused to grant conclusive weight to the restrictive language used by the providers in the transaction.¹⁹⁵ Instead, the *Vernor* court analyzed the entirety of both the agreement and the transaction to determine whether the purchase should be considered a sale.¹⁹⁶ Thus, in contrast with *MDY*, the *Vernor* court gave proper deference to Congress’ intent of protecting software users from copyright liability under Sections 109 and 117 of the Copyright Act by looking to what the parties actually transacted for, and not solely what the software provider claimed it was selling.¹⁹⁷

The *Vernor* court concluded that the critical factor in determining whether a purchase was a license or a sale was whether the purchaser was required to return the purchased copy to the copyright holder.¹⁹⁸ Therefore, even though CTA’s purchases of Autodesk’s software contained limitations on its use, because CTA purchased the copies for a one-time payment at the time of sale and the contract allowed CTA to retain possession of the program, the purchases constituted a sale.¹⁹⁹

The *Vernor* court explicitly noted the conflicting Ninth Circuit decisions in *MAI* and *Wise*,²⁰⁰ and its language suggests that it believes *MAI* was incorrectly decided.²⁰¹ Indeed, the *Vernor* court pointed out that the *MAI* court neither cited *Wise*, the previously binding Ninth Circuit precedent, nor supplied

¹⁹³ *Wall Data Inc. v. L.A. County Sheriff’s Dept.*, 447 F.3d 769, 785 (9th Cir. 2006).

¹⁹⁴ *Vernor*, 555 F. Supp. 2d at 1169 (“[t]he label placed on the transaction is not determinative.”).

¹⁹⁵ *Id.*

¹⁹⁶ *See id.*

¹⁹⁷ *See id.* at 1169-70.

¹⁹⁸ *Id.* at 1170. Indeed, “[e]ven a complete prohibition on the further transfer of the print (as in the Redgrave Contract), or a requirement that the print be salvaged or destroyed, was insufficient to negate a sale where the transferee was not required to return the print.” *Id.*

¹⁹⁹ *Id.*

²⁰⁰ *Id.* at 1171-74.

²⁰¹ *Id.* at 1171-72 (“In a single footnote, without analysis or explanation, the [*MAI*] court declared that ‘since *MAI* licensed its software, its customers do not qualify as ‘owners’ of the software and are not eligible for protection under § 117.’ The court did not cite *Wise*.”) *Id.* at 1171.

any reasoning to support its determination that the purchaser was only a licensee of the software.²⁰² In contrast with *Wise*'s reasoning, the *MAI* court looked only at the limiting terms of the license to see whether the purchase was a sale or a license, and failed to assess the "general tenor" of the agreement in making its determination.²⁰³ Due to the restrictions on the license, if the *Vernor* court followed *MAI* and its progeny, CTA would have merely received a license for the software packages.²⁰⁴ Nonetheless, the *Vernor* court followed *Wise*'s reasoning, and held that because Autodesk sold the software packages to CTA and Mr. Vernor lawfully purchased the software from CTA, Mr. Vernor was an "owner" of the copy and was entitled to a Section 117 defense.²⁰⁵

2. *MDY* and the First Sale Doctrine

The *MDY* court, without elaboration, acknowledged the *MAI* and *Wall Data* decisions as binding precedent and refused to follow *Wise*, thus undermining Congress's intent in protecting software purchasers from extensive copyright liability under Section 117 of the Copyright Act.²⁰⁶

The district court, while noting *Wise*, doubted the outcome of *MDY*'s facts under *Wise*.²⁰⁷ The court stated that under *Wise*, a transaction is a license when the purchaser never receives title in the transaction.²⁰⁸ Blizzard, in its EULA, provides "that Blizzard explicitly retains title to 'all copies' of the game client software."²⁰⁹ However, this reasoning is flawed in that it focuses solely on the copyright holder's restrictive EULA terms and fails to consider the economic realities of the transaction, as *Wise* requires.²¹⁰

Applying the reasoning of *Wise* to *MDY*'s facts, when a user purchases a copy of WoW, the user obtains one copy of the

²⁰² *Id.*

²⁰³ *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 517-18 (9th Cir. 1993).

²⁰⁴ *Vernor*, 555 F. Supp. 2d at 1172.

²⁰⁵ *Id.* at 1174-75.

²⁰⁶ See *supra* note 117 and accompanying text.

²⁰⁷ *MDY Indus., LLC v. Blizzard Entm't, Inc.*, No. CV-06-02555-PHX-DGC, 2008 WL 2757357, at *10 n. 7 (D. Ariz. July 14, 2008).

²⁰⁸ *Id.*

²⁰⁹ *Id.*

²¹⁰ *United States v. Wise*, 550 F.2d 1180, 1191 (9th Cir. 1977).

software for a one-time payment²¹¹ from either a retail store or a website.²¹² The user is not required to return the purchased copy, is not required to pay Blizzard to retain possession of the copy, and may destroy the copy if the purchaser wishes.²¹³ These are not characteristics of mere licenses, but are consistent with ownership powers. The *Wise* inquiry is concerned with the economic and social realities of the transaction, not just the restrictive language that copyright holders provide in their contract terms to limit users' rights.²¹⁴ Under *Wise*, a WoW purchaser would likely be classified as a software "owner" and, therefore, would be entitled to a Section 117 defense against copyright infringement. Thus, the District Court of Arizona erred in ignoring *Wise* and its progeny.

B. *Lockean Labor Desert Theory*

In this section, I argue that Lockean labor desert theory should at least influence courts towards classifying a WoW user as a software "owner" instead of a licensee for the purpose of a Section 117 defense, given an MMORPG player's time and labor investment into the virtual world. Because MMORPG users invest substantial amounts of time and money into their avatars' development, they should be afforded more substantial rights in their ability to use the game as they wish, free from fear of liability for copyright infringement.

"Video games are big business."²¹⁵ "Millions of people play these games," and their subscription fees make the operators very profitable.²¹⁶ In addition to the subscription fees the operators receive, many other MMORPGs, though not WoW, receive advertising dollars from major corporations such as Intel and McDonald's that cater to the gaming community.²¹⁷ While the Copyright Act protects the profits of owners, the game players whose time and effort enable those profits

²¹¹ See Public Knowledge, *supra* note 2, at 15.

²¹² See *id.* at 17.

²¹³ See *id.* at 18.

²¹⁴ *Wise*, 550 F.2d at 1191.

²¹⁵ Kenneth Hwang, *Blizzard Versus BNETD: A Looming Ice Age for Free Software Development?*, 92 CORNELL L. REV. 1043, 1046 (2006); "Virtual worlds are becoming more important in the lives of average citizens. These virtual worlds produce real effect in the real world." Kayser, *supra* note 71, at 85.

²¹⁶ Kayser, *supra* note 71, at 62.

²¹⁷ Gregory Lastowka & Dan Hunter, *The Laws of Virtual Worlds*, 92 CAL. L. REV. 1, 8 (2004).

deserve consideration as well.²¹⁸ Because MMORPGs are qualitatively different from most property purchases in how the players interact with their purchase by “living” in the virtual world and investing an extraordinary amount of time and effort into the world, an operator should not be able to merely utilize a EULA to impose unilateral restrictions upon its virtual world inhabitants.

An MMORPG player’s immense time investment into the virtual world is unlikely to be substantial enough to acquire rights the software provider withheld from the user under the EULA or TOU. The EULA and TOU are binding contracts, and the user activity does not occur at the time of the transaction, but instead occurs *ex post*. However, the user’s labor investment into the virtual world after the transaction should at least influence courts towards classifying a user as an “owner” instead of a licensee for the purpose of a Section 117 defense.²¹⁹ In other words, a WoW user’s substantial in-game time investment should not insulate the user from breaching Blizzard’s EULA or TOU contracts, but the user’s *ex post* treatment of the software should be a considerable factor in the court’s inquiry as to whether a user is an “owner” or a licensee of that particular software.

Commentators recognize the internet as a space separate and apart from the “real world,” and view internet commodities as a type of quasi-property.²²⁰ But as illustrated in *MDY*, virtual world disputes have “real world” consequences.²²¹

²¹⁸ *MDY Indus., LLC v. Blizzard Entm’t, Inc.*, No. CV-06-02555-PHX-DGC, 2008 WL 2757357, at *9 (D. Ariz. July 14, 2008). For an interesting article concerning players’ rights, see Raph Koster, *Declaring the Rights of Players*, Aug. 27, 2000, www.raphkoster.com/gaming/playerrights.shtml (advocating that players and avatars receive certain inalienable rights).

²¹⁹ “Indeed, if a user’s claim to a virtual product were strong enough, courts might be justified in ignoring the terms of a EULA that limited virtual property rights.” Horowitz, *supra* note 27, at 444. “Through time, effort, and often monetary expenditures, players’ avatars build status in their perspective communities, amass virtual property (usually taking the form of weapons or armor), and gain characteristics advantageous to game play.” Westbrook, *supra* note 35, at 780. Indeed, “many [players] spend[] hundreds of hours per year logged in.” *Id.*

²²⁰ See Jessica Vascellaro, *Yahoo Posts Loss as New Chief Plots Strategy*, WALL ST. J., Jan 29, 2009 at B1 (stating that Yahoo is “a fantastic Internet property. . . . It really doesn’t deserve everybody trying to pick it and pull it apart.”).

²²¹ Horowitz, *supra* note 27, at 443 (describing far more interesting scenarios such as a Chinese gamer killing someone for stealing an online item, and Anshe Chung, who became the first person to become a millionaire through acquiring virtual property); see also Ross Miller, *WoW Character Sells for Nearly \$10,000*, Joystiq, Sept. 17, 2007 available at www.joystiq.com/2007/09/17/wow-character-sells-for-nearly-10-000/ (describing a WoW character with “arguably the best gear in the game” that sold for \$9,700).

Nevertheless, as of now, America does not honor virtual property rights,²²² due in part to the lack of any virtual property litigation or legislation.²²³ All of the cases where it appeared the court would have to consider virtual property rights have settled.²²⁴ Many articles discuss the possibility of virtual property rights, and one of the arguments most frequently advocated in favor of recognizing these rights is one based on John Locke's theory of labor desert.

Lockean labor desert theory allocates property rights to those who invest their time and effort in distinguishing an object from a commons.²²⁵ When a person mixes her labor with an object from a commons, the person makes that object her property²²⁶ so long as her labor contributed the greatest part of

²²² "A virtual property right is a property right in a virtual product." Horowitz, *supra* note 27, at 444; *see also* Westbrook, *supra* note 35, at 782. "[C]omputer code enables [virtual items] to resemble real chattels in their 'rivalrousness, persistence, and interconnectivity.' That is, '[i]f I hold a pen, I have it and you don't . . . If I put the pen down and leave the room, it is still there . . . And finally, you can all interact with the pen.'" Lederman, *supra* note 50, at 1631.

²²³ Kayser, *supra* note 71, at 65.

²²⁴ Westbrook, *supra* note 35, at 805. However, a Chinese court acknowledged that virtual property is entitled to some protection, ordering an online gaming company to return the user's virtual items after a hacker stole the items when he hacked the game company's servers. *See* Will Knight, *Gamer Wins Back Virtual Booty in Court Battle*, *NewScientist.com*, December 23, 2003, <http://www.newscientist.com/article/dn4510-gamer-wins-back-virtual-booty-in-court-battle.html> (last visited Mar. 1, 2010). In that case, Li Hongchen, a twenty-four year old gamer, spent over two years and \$1,210 buying virtual goods in the online game, "Red Moon." *Id.* A hacker infiltrated Red Moon's servers and raided Hongchen's account. *Id.* Hongchen told the Chinese news site Xinhuanet, "I exchanged the equipment with my labour, time, wisdom and money, and of course they are my belongings." *Id.* Hongchen argued "that the developer inadequately protected his virtual belongings from theft by hackers." Westbrook, *supra* note 35, at 805. Indeed, "the line between online games and the real world have [sic] begun to blur. Some gamers already trade game goods and characters for real money through online auction sites like eBay." Knight, *supra*; *see also* Thomas Claburn, *Virtual Property Rights Are No Game*, *INFORMATION WEEK*, Dec. 16, 2006, *available at* www.informationweek.com/story/showArticle.jhtml?articleID=196604327 (describing the *Bragg v. Linden Research* case, 487 F. Supp. 2d 593 (E.D. Pa. 2007), which later settled, where "Bragg claim[ed] that Linden Lab froze \$8,000 worth of virtual assets and refused to reimburse him" when Bragg acquired the assets by "taking advantage of a loophole in its code"). The *Bragg* case is different in that *Second Life*, unlike *WoW*, allows players to own the items they acquire. *Id.* A final adjudication in this case would have been significant in that it would provide some clarification on what gamers who possess virtual items actually own.

²²⁵ Horowitz, *supra* note 27, at 451.

²²⁶ "Whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joined to it something that is his own, and thereby makes it his Property." JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* 306 (Peter Laslett ed., Cambridge Univ. Press 1988) (1690).

the asset's value.²²⁷ When one person labors to acquire a good, that person is entitled to reap its benefit over one who expended no labor.²²⁸ Under the "Enough as Good" proviso, Locke limited application of this theory to situations where "there is enough, and as good left in common for others."²²⁹

In the case of a WoW purchaser, commentators suggest that there are two competing Lockean claims.²³⁰ The operator, Blizzard, has a Lockean claim because it created and operated the commons, which in this case, is the virtual world of WoW.²³¹ Because Blizzard is responsible for creating the mechanisms by which WoW players seek to exercise a Lockean claim, a MMORPG player's Lockean claim may not be assertive enough to claim full ownership rights over her virtual property. The virtual world creator not only created the software, but in Lockean terms, the creator also supplied the "raw materials" that the users gathered to create or claim the items that they call "property."²³² Thus, before the user ever entered the world, before the user heard about the game, or even before the game was placed on the shelf, the virtual world operator expended not only its labor, but original, innovative thought in creating the new cyber-world.²³³

On the other hand, the player, prior to entering the game, created a customized avatar, without which the gaming experience would fail to exist at all.²³⁴ Due to WoW's focus on creating a social network to enhance game-play,²³⁵ the network effects of having many players "laboring" in the virtual world are invaluable.²³⁶ Indeed, the distinguishing and most valuable

²²⁷ JOHN LOCKE, SECOND TREATISE OF CIVIL GOVERNMENT § 28 (1690). "If I own a can of tomato juice and spill it in the sea so that its molecules . . . mingle evenly throughout the sea, do I thereby own the sea. . . ?" ROBERT NOZICK, ANARCHY, STATE AND UTOPIA 175 (1974).

²²⁸ Westbrook, *supra* note 35, at 792; *id.* at 794.

²²⁹ LOCKE, *supra* note 226, at 288.

²³⁰ See Horowitz, *supra* note 27, at 451-56.

²³¹ Thus, it is possible that the operator's Lockean claim in creating and maintaining the virtual world is strong enough to swallow up the user's claim. See *id.*

²³² *Id.* at 451-53.

²³³ *Id.* at 433.

²³⁴ See Westbrook, *supra* note 35, at 792-93.

²³⁵ See *supra* Part II.A.3.

²³⁶ However, there is a good question as to what actually constitutes "labor." See *infra* notes 240-246. Network effects increase "[t]he utility that a subscriber derives from a communications service. . . as others join the system." Jeffrey Rohlfs, *A Theory of Interdependent Demand for a Communications Service*, 5 *The Bell Journal of Economics and Management Science* 1 at 16 (Spring 1974). Historically, network effects have been critical in the development of the telegraph, telephone, broadcast radio,

feature of a MMORPG is the “massive” number of players. Further, all of the in-game assets players acquire, and all of the loot that users create, arose as a result of their time and labor investment. Players spend thousands of hours playing WoW, leveling their avatar, acquiring or crafting rare, high-level items,²³⁷ and may even earn a living in the virtual world.²³⁸ Moreover, the deep virtual world connection causes some players to consider themselves to be dual citizens of their virtual world and the “real world.”²³⁹

Because of these competing Lockean claims, users may not have a strong enough Lockean claim to assert full ownership rights over their virtual items and thus insulate themselves from breach of contract claims against violating the provider’s EULA and TOU terms. But the labor that users expend into their virtual world assets should be a considerable factor in considering a user’s classification as a software “owner” instead of as a mere licensee.

Indeed, there is a fundamental distinction between the user’s claim and the operator’s claim. While Blizzard’s competing Lockean claim may be strong as to the entire virtual world’s framework, a user’s Lockean claim may be stronger as to the particular WoW account and avatar.²⁴⁰ In Lockean terms, the WoW purchaser is responsible for the greatest value of the asset, his avatar, because of his expended time and labor. While the operator created the virtual universe, the user created something unique to the commons that was not present before.²⁴¹

The intense labor investment does not cease once an avatar reaches maximum level. Even after reaching maximum level, a WoW player’s adventure has just begun in terms of the

television, cellular phones, and most recently, the internet. See Robert M. Metcalfe, *It’s All in Your Head*, FORBES, May 7, 2007, available at <http://www.forbes.com/forbes/2007/0507/052.html>.

²³⁷ Westbrook, *supra* note 35, at 792.

²³⁸ See Rob Hof, *Second Life’s First Millionaire*, BUSINESS WEEK, Nov. 26, 2006, available at http://www.businessweek.com/the_thread/techbeat/archives/2006/11/second_lifes_fi.html.

²³⁹ Kayser, *supra* note 71, at 60. “Participants make sizable investments of social, human, and economic capital in these virtual worlds, often with the questionable expectation that the items they have collected and creations they have developed are their property.” Sheldon, *supra* note 55, at 751; “Virtual environments are now one of the most important forms of entertainment. More South Koreans play in virtual worlds than watch television.” Fairfield, *supra* note 70, at 1061.

²⁴⁰ See Horowitz, *supra* note 27, at 452-53.

²⁴¹ See Sheldon, *supra* note 55, at 761 (comparing WoW with the popular online game “Second Life,” where users invent new objects).

amount of labor required to find the best weapons, armor, and other items that WoW has to offer.²⁴² In other words, while an operator is enabling the avatar's existence, a user is contributing her labor to distinguish her avatar from the rest of the commons (the virtual world) in terms of appearance, items, guild affiliation, and social standing in the online community.²⁴³ From this creation and labor, the user creates not only the avatar, but also greatly increases the value of the avatar from zero to as high as \$9,700.²⁴⁴ Entrepreneurs have created companies whose purpose is to buy and sell virtual items for real money,²⁴⁵ and some make hundreds of thousands of dollars per year selling virtual items on eBay.²⁴⁶ This commodification and increased value would fail to exist without the user's labor.

Further, Locke's limiting "Enough as Good" proviso, where one may only claim property to the extent the claimant leaves "enough and as good" in common for others, is a non-issue in most virtual worlds. In contrast with "real world" rivalrous goods where there is only a finite amount of resources for distribution, in virtual worlds, the supply of goods is limited only by the amount of time that a purchaser invests into the game.²⁴⁷

Additionally, some players have invested so much time into WoW and have become so skilled at the game that WoW supports their career as professional gamers.²⁴⁸ Every year, Blizzard sponsors a gaming event titled "Blizzcon" that players can attend to meet with and compete against other players.²⁴⁹ "Blizzcon" sponsors a WoW player versus player tournament where the winning three-person team takes home \$75,000.²⁵⁰ Among the entrants to the WoW Tournament are high-profile

²⁴² *Id.*

²⁴³ "[W]ithout the inputs of the user, the avatar would not exist at all." Westbrook, *supra* note 35, at 792.

²⁴⁴ *See supra* note 221.

²⁴⁵ Westbrook, *supra* note 35, at 790.

²⁴⁶ Lastowka & Hunter, *supra* note 217, at 39.

²⁴⁷ *Id.* at 47-48; *see also* Fairfield, *supra* note 70, at 1048-50 (discussing the distinction between rivalrous and nonrivalrous goods).

²⁴⁸ Westbrook, *supra* note 35, at 789.

²⁴⁹ Blizzard.com, *What is Blizzcon*, <http://us.blizzard.com/blizzcon/index.xml> (last visited Jan. 6, 2010).

²⁵⁰ Blizzard.com, *Tournaments*, <http://www.blizzard.com/blizzcon/tournaments/> (last visited Mar. 1, 2010).

professional gaming groups that have earned corporate sponsorship.²⁵¹

The fundamental distinction between Blizzard's Lockean claim as an operator of a commons and a WoW purchaser's Lockean claim as a user of an avatar provides another illustration as to why *MDY* was wrongly decided, and gives further support to the proposition that a WoW user should be classified as an "owner" of the software and not a licensee for the purposes of a Section 117 defense under the Copyright Act.

Blizzard, as the operator and greatest Lockean stakeholder of the virtual world, must equitably allocate the rights among the players. It does this by acting as WoW's gatekeeper, enacting a EULA and TOU barring Glider use and other player conduct.²⁵² If a player acquired full virtual property rights to his online commodities, the user would undermine Blizzard's gate-keeping role to the detriment of other users. Notwithstanding any profit-seeking motive, Blizzard must retain its breach of contract claim in order to protect other WoW users' rights. But the ability to file copyright infringement actions against a player makes little sense, because Blizzard is acting outside its Lockean claim as protector of the commons and trespassing into the user's Lockean claim as to the player's own individual virtual avatar.

In *MDY*, however, arguing that a WoW purchaser should be treated as an "owner" instead of a licensee because of the purchaser's labor and time investment may be somewhat paradoxical. Under Lockean labor desert theory, rights should be allocated to players based upon the player's labor investment in the game. However, those rights may not be as strong when a player uses Glider, because Glider reduces the net amount of a player's labor by operating the game for the player. Indeed, Glider users are not physically sitting at the computer investing their time and labor into the virtual commons.²⁵³ Instead, players are simply inputting parameters

²⁵¹ TeamPandemic.net, *Pandemic Partners*, <http://www.teampandemic.net/index.php?page=partners> (last visited Mar. 1, 2010). Outside of WoW, other professional internet gamers describe themselves as "cyber-athletes" and in addition to practicing their games eight to twelve hours a day, exercise to maintain high energy levels, preserve quick reflexes, and improve hand to eye coordination. See Daniel Schorn, *Cyber Athlete 'FatalIty'*, Aug. 6, 2006, CBS NEWS, available at, <http://www.cbsnews.com/stories/2006/01/19/60minutes/main1220146.shtml>.

²⁵² See *supra* notes 63-68.

²⁵³ See *supra* notes 84-85.

into the Glider program and letting it “do the work.”²⁵⁴ It is difficult to argue that WoW purchasers are investing labor into their avatars or into WoW’s social experience while operating Glider. However, because Glider markets to experienced players who have already completed much of the basic WoW game, and does not market to beginning players, this paradox does not mean that the purchasers have not previously invested a great deal of time into the game. It only suggests that they cheat and thereby lessen the value of others’ labor.

Even though Lockean labor desert theory may not be a strong enough argument to afford WoW players a unilateral virtual property right sufficient to overcome Blizzard’s EULA and TOU, it should influence a user’s classification as an “owner” rather than a licensee when determining her eligibility for a Section 117 defense. This is because the user’s actual usage of the game informs the economic realities of the transaction. While Blizzard created the software and the virtual world, the players created their avatars and added value.²⁵⁵ Thus, purchasers should be classified as “owners” and accordingly be free from fear of copyright infringement’s statutory damages. WoW players neither think nor act like licensees. The players maintain exclusive possession of the software, invest a great deal of time and money into the game, and do not expect the virtual world operator to have the right to arbitrarily terminate their account or take their in-game earnings.²⁵⁶ Indeed, as one Second Life player explained:

When a character in the game ‘owns’ something, I feel I ‘own’ it in a similar sense. If the character has the right to destroy it, I feel I have the right to destroy it. If the character has the right to give it away for arbitrary reasons, I feel I have a similar right. Note this isn’t a roleplaying argument, it is quite the opposite. It relies on the avatar and the player being equivalent.²⁵⁷

Looking through a Lockean lens at the software transaction and gamers’ subsequent investment in the virtual world, it is counterintuitive to classify these players as

²⁵⁴ See *supra* notes 89.

²⁵⁵ See *supra* note 246.

²⁵⁶ See Todd David Marcus, *Fostering Creativity in Virtual Worlds: Easing the Restrictiveness of Copyright for User-Created Content*, 52 N.Y.L. SCH. L. REV. 67, 80 (2007) (describing the user’s frustration and lack of recourse against an operator who deletes the user’s in-game goods due to TOU and EULA restrictions).

²⁵⁷ Kurt Hunt, *This Land Is Not Your Land: Second Life, CopyBot, and the Looming Question of Virtual Property Rights*, 9 TEX. REV. ENT. & SPORTS L. 141, 159 (2007).

software licensees instead of “owners.”²⁵⁸ While the software purchasers may not have a strong enough Lockean claim to assert full ownership rights over their avatar, courts should consider the Lockean argument and the user’s *ex post* handling of the software when determining whether the purchaser is an “owner” or a licensee of the software. Further, the underlying principles of copyright law counsel in favor of granting software purchasers more substantial protections for their online commodities.

C. Copyright Policies

Copyright protections that are too favorable to software providers stymie creative development far out of proportion to what Congress intended. Copyright law must evolve in order to foster creativity and innovation in online worlds. WoW provides a perfect example.

In the past, users have developed “illegal” third-party programs, many of which Blizzard bought and incorporated into WoW’s user interface in order to improve the gaming experience.²⁵⁹ This practice allows both Blizzard and the user to benefit from the user’s labor, creativity, and innovation in creating the third-party program. However, the uncertainty of whether the third-party programs will be treated as investment opportunities or copyright infringements expunges any incentive for third-party program developers to innovate on a game. This disincentive is exacerbated if the law allows companies such as Blizzard to irrefutably characterize its sales as “licenses” rather than transfers of ownership.²⁶⁰ The uncertainty of who the software provider will favor and who the software provider will abhor may lead third-party program developers to cease improving upon the software provider’s original work without permission. Essentially, by frustrating and discouraging further innovation, there is a net societal loss.

²⁵⁸ *Id.*

²⁵⁹ Harald Warmelink, *Blizzard-Cosmos. Negotiating Add-On Development*, Mar. 1, 2007, http://sybil.nl/2007/index2.php?option=com_content&do_pdf=1&id=14.

²⁶⁰ See Marcus, *supra* note 256, at 80. However, increased commodification of in-game items may incentivize a shift to allow players to retain copyrights over their virtual property because players will spend most of their time where they will be “best . . . rewarded for their efforts.” *Id.* at 86. “Creating a new virtual platform that allows users to retain copyrights for their creations becomes a safer investment for those seeking new avenues of financial opportunity.” *Id.*

While it may be difficult to argue that Glider improved WoW, for other third-party programs, the line between improvement and harm may be more unclear.²⁶¹

Further, in addition to fostering creativity, Congress intended to protect the incidental copying of software to RAM as a necessary part of everyday software use.²⁶² Indeed, Congress foresaw the exact problem at issue in *MDY* and recognized that due to the software provider's restrictive EULA language, software purchasers that exceed any provision of the provider's terms, no matter how insignificant, might not qualify for a Section 117 defense under the Copyright Act.²⁶³

More significantly, virtual worlds such as WoW may be planting the seeds of a future where people may not just inhabit virtual worlds to level a character, but to meet other people, date, or study.²⁶⁴ There may come a time where the "real" and "virtual" self become so intertwined that there is little distinction between them. Blizzard has the right to make the rules governing how it runs WoW; however, it should not have the right to evade Congressional laws protecting software users from copyright infringement claims under Section 117 defense of the Copyright Act.²⁶⁵

²⁶¹ Consider the popular third-party program "Atlas," which allows players to view the layouts of every WoW dungeon without ever visiting the dungeon. See Curse.com, Atlas, available at <http://wow.curse.com/downloads/wow-addons/details/atlas.aspx> (last visited Jan. 26, 2009).

²⁶² Senator Hatch stated:

Second, I am concerned about the interplay between criminal liability for reproduction in the bill and the commonly-held view that the loading of a computer program into random access memory (RAM) is a reproduction for purposes of the Copyright Act. Because most shrink-wrap licenses purport to make the purchaser of computer software a licensee and not an owner of his or her copy of the software, the ordinary purchaser of software may not be able to take advantage of the exemption provided by sec. 117, allowing the owner of a copy to reproduce the work in order to use it in his or her computer.

143 CONG. REC. S12689 (Nov. 13, 1997) (statement of Sen. Hatch), available at <http://digital-law-online.info/lpd1.0/quotes/fn2-36.htm#q>.

²⁶³ *Id.*

²⁶⁴ "In the future, virtual worlds platforms will be adopted for commerce, for education, for professional, military, and vocational training, for medical consultation and psychotherapy, and even for social and economic experimentation to test how social norms develop." Balkin, *supra* note 30, at 2044. "[I]t is possible, if not likely, that many virtual spaces will effectively become shopping malls for both real and virtual goods." *Id.* at 2067. "The United States military uses virtual worlds for training . . . [as t]he [virtual] environment re-creates sections of Baghdad down to street signs and palm trees." Fairfield, *supra* note 70, at 1060.

²⁶⁵ See *supra* note 117.

The rules of the “real world” still apply in the virtual world. Here, Blizzard utilized restrictive EULA language to limit the rights of purchasers, and asserted a copyright infringement claim against MDY when Congress clearly intended to protect the underlying users from copyright liability by enacting Section 117. Courts should reject arguments which advocate that software providers’ rules garner more weight than Congress’ intention of protecting software purchasers.²⁶⁶

V. CONCLUSION

MDY was wrongly decided because courts should afford software purchasers and their labor investments greater protection from the statutory damages of copyright liability. The Ninth Circuit recognized this concern and granted software users such protection when it invoked the First Sale Doctrine in *Wise* and *Vernor*. This enhanced protection is supported by John Locke’s labor desert theory, and the underlying purposes of copyright law to encourage innovation.

Courts should interpret Section 117 of the Copyright Act as the consumer expects to be treated, looking to the practical realities of the sale, unconstrained by the “four corners” of the EULA that the provider forces upon its customers. The incidental copying of the software to RAM, even if the copying is in violation of the software provider’s EULA, is the type of benign copying that Congress intended to shield from liability under Section 117 of the Copyright Act. Such expansive copyright liability for users incentivizes software providers to license everything, and “sell” nothing.

Looking forward, under the *MDY* reasoning, software providers, in boilerplate fashion, will continue to incorporate restrictive EULAs into every agreement. The agreement will provide that the purchasers are only licensed to use the software and own nothing. Once the provider exceeds the EULA license terms, instead of relying on a breach of contract claim, software providers may unleash the brutal statutory damages of copyright infringement upon users that never realized that they were copying anything. To avoid this injustice, courts should invoke the First Sale Doctrine, which

²⁶⁶ See Tobold’s MMORPG Blog, *Virtual Property Rights*, <http://tobolds.blogspot.com/2008/03/virtual-property-rights.html> (Mar. 23, 2008, 14:30 EST).

more equitably allocates the rights between software providers and software purchasers by considering the economic and social realities of the transaction without giving dispositive weight to a software provider's restrictive EULA terms.

Further, because MMORPG users invest substantial amounts of labor into their games, courts should consider this *ex post* activity in determining whether purchasers are "owners" or licensees. MMORPG players neither think nor act as mere licensees. Instead, they act like "owners." Players spend hundreds of hours in virtual worlds and other online communities where they customize an avatar and immensely increase its value. WoW users also form friendships and take part in other social in-game activities. As technology opens doors to new possibilities in virtual worlds, adhering to precedent that ignores many of the similarities between the "real world" and the virtual world will frustrate innovation and over-protect software providers while abrogating basic rights that Congress afforded to software purchasers.

Ross Shikowitz[†]

[†] J.D. Candidate, Brooklyn Law School, 2010; M.M., Indiana University-Bloomington, 2005; B.A., Skidmore College, 2003. I would like to thank the *Brooklyn Law Review* editors and staff for their diligent, professional review as well as their thoughtful comments in the publication of this Note. I would also like to thank Professors Derek Bambauer and Beryl Jones-Woodin for their guidance, and critique. Most importantly, I would like to thank my parents Hillary and Alan Shikowitz, my sister Sara Shikowitz, my aunt Adrienne Kapel, my grandmother Barbara Kapel, and my girlfriend Erica Levy, without whose love, encouragement, and understanding this Note would not have been possible.