

Reputation Life Cycle: The Case of Franchising

Uri Benoliel*

Law-and-economics conventional analysis contends that franchise termination laws, prohibiting opportunistic termination by the franchisor, are superfluous. Well-rooted in traditional law-and-economics opposition to such legislation is the belief that the reputation mechanism serves as a sufficient control against opportunistic termination by franchisors. This article questions the idea that reputation concerns can be a substitute for regulation of franchise relationships.

The article argues that the reputation mechanism suffers from an inherent deficiency, which disqualifies it from serving as a substitute for franchise termination legislation. In short, the reputation mechanism often has a limited life cycle consisting of three stages: creation, erosion and collapse. At the early stages of the franchisor organizational life cycle, the reputation mechanism may have some deterring effects on the franchisor. At these stages, the incentive of the franchisor is normally to attract new franchisees and thus the franchisor attaches greater importance to being perceived as fair in the eyes of potential franchisees. As the franchisor's financial resources increase, the effectiveness of the reputation mechanism will be eroded. Particularly, the franchisor's incentive to attract new franchisees will decrease. Similarly, its incentive to maintain its present franchisees will decline. Instead, the franchisor, having greater financial resources, will gradually shift towards self-ownership of the units. Following the continuing decrease in the franchisor's financial constraints and the decrease in its incentive to maintain and attract new franchisees, the reputation mechanism may collapse. This may particularly occur when the reputation-related costs decrease to a minimum level, at which point they will be lower than the franchisor's benefits from opportunistic termination of the franchise contract.

* Assistant Professor, Academic Center of Law & Business. J.S.D., University of California, Berkeley, 2006; LL.M, Columbia University, 2005; LL.B., Hebrew University of Jerusalem, 2003. This Article greatly benefited from comments and criticisms by Gary Castrogiovanni, Richard Epstein, Tamar Frankel, Warren Grimes, Steven Michael, Renana Peres, Yuval Procaccia, James White, and conference participants at the 2009 Israeli Law and Economics annual meeting. This article was presented at the 2009 annual conference of the Canadian Law and Economics Association (October 2009).

INTRODUCTION

Legal economists have long debated whether the reputation mechanism can serve as a substitute for regulating the behavior of legal entities bound to a contractual relationship.¹ If the importance of maintaining a good reputation in the market can effectively regulate or deter opportunistic behavior by the parties, legal regulation of the contractual relationship is a social waste. Conversely, if the market-based reputation mechanism cannot deter opportunism, regulation of the relationship is more justifiable. Traditional law-and-economics analysis, which has found its way into current legislative practices in the United States, tends to suggest that the reputation mechanism can indeed function as an effective substitute for costly regulation of contractual relationships.² One commercial relationship in particular—the franchise relationship—is well-positioned, say legal economists, to be governed by reputation rather than by regulation.

The franchise relationship involves two major legal entities: the franchisor provides its trademark or trade name and a business system, while the franchisee generally pays the franchisor an ongoing royalty and an initial fee.³ To date, the legal ties of franchise relationships are regulated by a minority of fifteen states in the United States. These states enacted termination laws requiring “good cause” for the termination of a franchise contract by a franchisor,⁴ generally defined as the

¹ See Gillian K. Hadfield, *Problematic Relations: Franchising and the Law of Incomplete Contracts*, 42 STAN. L. REV. 927, 978 n.232 (1990).

² See, e.g., Lewis A. Kornhauser, *Reliance, Reputation, and Breach of Contract*, 26 J.L. & ECON. 691, 703 (1983) (presenting a model in which “reputation will substitute perfectly for a damage rule”); DOUGLAS G. BAIRD, ROBERT H. GERTNER & RANDAL C. PICKER, *GAME THEORY AND THE LAW* 57, 159–87 (1994) (suggesting that the reputation mechanism “can bring about long-term cooperation even if there is no enforceable contract”). But see STEVEN SHAVELL, *FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW*, 323–24 (2004) (arguing that the reputation mechanism is an imperfect substitute for courts).

³ The contract underlying the relationship between the two parties is the “franchise contract.” International Franchise Association, *Frequently Asked Questions About Franchising*, <http://www.franchise.org/industrysecondary.aspx?id=10008> (last visited Nov. 1, 2009). See generally Hadfield, *supra* note 1, at 931–32.

⁴ The fifteen “good cause” states are Arkansas, California, Connecticut, Delaware, Hawaii, Illinois, Indiana, Iowa, Michigan, Minnesota, Nebraska, New Jersey, Virginia, Washington, and Wisconsin. See ARK. CODE ANN. § 4-72-204 (2001); CAL. BUS. & PROF. CODE § 20020 (Deering 2007); CONN. GEN. STAT. ANN. § 42-133f (West 2007); DEL. CODE ANN. tit. 6, § 2552(a) (2005); HAW. REV. STAT. ANN. § 482E-6(2)(H) (LexisNexis 2009); 815 ILL. COMP. STAT. ANN. 705/19 (West 2008); IND. CODE ANN. § 23-2-2.7-1(1)(7) (West 2005); IOWA CODE ANN. § 523H.7(1) (West 2007); MICH. COMP. LAWS ANN. § 445.1527(27)(c) (West 2002); MINN. STAT. ANN. § 80C.14(3)(b) (West 2009); NEB. REV. STAT. ANN. § 87-404 (LexisNexis 2007); N.J. STAT. ANN. § 56:10-5 (West 2001); VA. CODE ANN. § 13.1-564 (2006); WASH. REV. CODE ANN. § 19.100.180(2)(j) (West 1999); WIS. STAT. ANN. § 135.03 (West 2009).

franchisee's failure to obey the franchise contract.⁵ Under these laws, a franchisor who concludes the franchise contract without a show of good cause must pay the franchisee damages.⁶ These franchise termination laws aim to protect franchisees against opportunistic treatment by franchisors, who inherently enjoy greater bargaining power.⁷ They are designed to protect franchisees from franchisors, who may terminate the franchise contract and opportunistically appropriate the profitable outlet developed by the franchisee.⁸ More specifically, an opportunistic franchisor may terminate the contract of an efficient franchisee, who fully complies with the franchise contract, in order to sell the latter's profitable unit to a new franchisee for higher franchise fees.⁹ Alternatively, the franchisor can terminate the contract and manage the successful unit himself.¹⁰

Conventional law-and-economics analysis proposes that franchise termination laws, prohibiting opportunistic termination by the franchisor, are superfluous.¹¹ Deeply rooted in traditional law-and-economics opposition to such laws is the belief that the

⁵ See ARK. CODE ANN. § 4-72-202(7)(A) (2001); CAL. BUS. & PROF. CODE § 20020 (Deering 2007); CONN. GEN. STAT. ANN. § 42-133f(a) (West 2007); HAW. REV. STAT. ANN. § 482E-6(2)(H) (LexisNexis 2009); 815 ILL. COMP. STAT. ANN. 705/19(b) (West 2008); IND. CODE ANN. § 23-2-2.7-1(1)(7) (West 2005); IOWA CODE ANN. § 523H.7(1) (West 2007); MICH. COMP. LAWS ANN. § 445.1527(27)(c) (West 2002); MINN. STAT. ANN. § 80C.14(3)(3)(b) (West 2009); NEB. REV. STAT. ANN. § 87-402(8) (LexisNexis 2007); N.J. STAT. ANN. § 56:10-5 (West 2001); WASH. REV. CODE ANN. § 19.100.180(2)(j) (West 1999).

⁶ Recoverable damages may include the following: (1) a fraction of the franchisee's tangible assets used with respect to the terminated franchise, including sales outlets and facilities, offices, warehouses, trucks, furnishings, equipment, and accessories; (2) loss of goodwill; and (3) loss of profits. See generally ARK. CODE ANN. § 4-72-208(b) (2001); CONN. GEN. STAT. ANN. § 42-133g(a) (West 2007); DEL. CODE ANN. tit. 6, § 2553(c) (2005); 815 ILL. COMP. STAT. ANN. 705/26 (West 2008); IOWA CODE ANN. § 523H.13 (West 2007); MINN. STAT. ANN. § 80C.17(1), (3) (West 2009); NEB. REV. STAT. ANN. § 87-409 (LexisNexis 2007); N.J. STAT. ANN. § 56:10-10 (West 2001); VA. CODE ANN. § 13.1-571(a) (2006); WIS. STAT. ANN. § 135.06 (West 2009).

⁷ VA. CODE ANN. § 13.1-558 (2006); Geib v. Amoco Oil Co., 29 F.3d 1050, 1056 (6th Cir. 1994); Wright-Moore Corp. v. Ricoh Corp., 908 F.2d 128, 135 (7th Cir. 1990); Kubis & Perszyk Assocs. v. Sun Microsystems, Inc., 680 A.2d 618, 622 (N.J. 1996); Corp v. Atlantic Richfield Co., 860 P.2d 1015, 1018 (Wash. 1993); Robert W. Emerson, *Franchise Contract Clauses and the Franchisor's Duty of Care Toward Its Franchisees*, 72 N.C. L. REV. 905, 926 (1994); Robert W. Emerson, *Franchising and the Collective Rights of Franchisees*, 43 VAND. L. REV. 1503, 1511 (1990).

⁸ Corp v. Atlantic Richfield, 860 P.2d at 1018; Robert W. Emerson, *Franchise Terminations: Legal Rights and Practical Effects When Franchisees Claim the Franchisor Discriminates*, 35 AM. BUS. L.J. 559, 564-65 (1998); Robert W. Emerson, *Franchise Selection and Retention: Discrimination Claims and Affirmative Action Programs*, 40 ARIZ. L. REV. 511, 519 (1998); Jean Wegman Burns, *Vertical Restraints, Efficiency, and the Real World*, 62 FORDHAM L. REV. 597, 620-21 (1993).

⁹ ROGER D. BLAIR & FRANCINE LAFONTAINE, *THE ECONOMICS OF FRANCHISING* 271 (2005).

¹⁰ *Id.*

¹¹ See Michael J. Lockerby, *Franchise Termination Restrictions: A Guide for Practitioners and Policy Makers*, 30 ANTITRUST BULL. 791, 858-60, 866-71 (1985).

reputation mechanism can serve as a sufficient control against opportunistic termination by franchisors. Since the incentive of franchisors is to maintain and attract franchisees, their motivation is to uphold a reputation of fairness. Accordingly, out of concern for their reputation, franchisors are not likely to opportunistically terminate the franchise contract.¹²

This “conventional law-and-economics opposition to franchise termination laws has been significantly influential in the development of franchise law in general, as witnessed by state and federal policy making.”¹³ Most states to date have refused to regulate general franchise termination laws prohibiting opportunistic termination of the franchise contract by the franchisor.¹⁴ At the federal level, such statutes have also been rejected.¹⁵ Yet legislators may be surprised to discover that the theoretical basis for this law-and-economics opposition to franchise termination laws is not quite as solid as one would hope.

This article, therefore, calls into question the idea that reputation can be a substitute for franchise termination laws. The article argues that the reputation mechanism suffers from an inherent deficiency, which disqualifies it from serving as a substitute for franchise termination laws.

In short, the argument is as follows: The reputation mechanism often has a limited life cycle consisting of three stages: creation, erosion and collapse. At the earliest stage of the franchisor’s organizational life cycle, the reputation mechanism is created. Normally, in this initial phase, the franchisor’s incentive is to maintain its previous franchisees and attract new ones. The franchisor wishes to expand the market of its products as rapidly as possible, but is financially limited and cannot expand on its own. Given the initial need of the franchisor to attract and maintain franchisees, the franchisor will be compelled to act in a manner which appears fair to present and prospective franchisees. Opportunistic termination of the franchise contract is therefore less likely to occur, as legal

¹² See *infra* Part II.A.

¹³ Uri Benoliel, *The Expectation of Continuity Effect and Franchise Termination Laws: A Behavioral Perspective*, 46 AM. BUS. L.J. 139, 141 (2009).

¹⁴ Since 1992, about thirty states have refused to regulate general franchise termination laws. See James A. Brickley, *Royalty Rates and Upfront Fees in Share Contracts: Evidence from Franchising*, 18 J.L. ECON. & ORG. 511, 518 (2002).

¹⁵ For example, in 1998 and 1999 the federal government failed to pass two bills that would have made it unlawful for a franchisor to terminate a franchise contract without good cause. See H.R. 4841, 105th Cong. (2nd Sess. 1998); H.R. 3308, 106th Cong. (1st Sess. 1999).

economists suggest, at the early stages of the franchisor's organizational life cycle.

However, following the creation of the reputation mechanism, its effectiveness is likely to be eroded. Particularly, as the franchisor's financial condition improves, the effectiveness of the reputation mechanism will steadily deteriorate. As the franchisor's financial resources increase, the franchisor will gradually shift away from attracting and maintaining franchisees towards operating, on its own, the present and prospective outlets in its chain. This shift mainly derives from the fact that the franchisor normally gains more profits from a franchisor-owned unit than from a franchisee-owned unit. As a result of this shift, the franchisor will be less concerned about being perceived as fair by present and prospective franchisees, which leads to the conclusion that termination without good cause is more likely to occur at this stage.

Ultimately, the reputation mechanism may collapse. Following the gradual increase in the franchisor's financial resources and the decrease in its incentive to maintain and attract new franchisees, the reputation mechanism will malfunction. The reputation-related costs will reach a minimum level, which will be lower than the franchisor's benefits from the opportunistic termination of the franchise contract. When the franchisor accumulates enough financial resources to operate *the whole* franchise chain on its own, this outcome is to be expected. At this final stage—and herein lies the crux of the argument—the franchisor's reputation costs incurred from being perceived as unfair by present and prospective franchisees are likely to be lower than the benefits from opportunistic termination. Opportunistic termination therefore is likely to occur.

Part I of this article will provide theoretical context by briefly reviewing the debate over the effectiveness of reputation as a substitute for franchise termination laws. First, the arguments in favor of the reputation mechanism, which underlie the traditional law-and-economics disregard for franchise termination laws, will be introduced. Second, I shall review the scant manifestations of skepticism as to the effectiveness of reputation in preventing opportunistic termination by franchisors. Part II will demonstrate the intrinsic deficiency of the reputation mechanism in preventing opportunism by the franchisor, namely the reputation life-cycle. The three major stages of the reputation life cycle will be presented—creation, erosion and collapse—with an eye toward uncovering the Achilles heel of the reputation mechanism. While this article focuses predominantly on the (in)effectiveness of reputation in

preventing opportunism in franchise relationships, I will conclude by suggesting broader implications of my analysis for the legal administration of other long-term commercial relationships.

I. THE DEBATE OVER REPUTATION

A. Law-and-Economics Conventional Analysis

Legal economists oppose franchise laws that require franchisors to demonstrate good cause for termination of a franchise contract.¹⁶ Essential to legal economists' resistance to franchise termination laws is the argument that the market-based reputation mechanism can function as a substitute for regulation.¹⁷ More specifically, legal economists argue that the reputation mechanism can prevent contract termination without good cause, since such termination will impose reputation-related costs on the franchisor.¹⁸ As Charles Goetz and Robert Scott argue, franchisors will "incur substantial 'goodwill' losses if they attempt to exploit a discretionary termination authority."¹⁹

Legal economists furthermore specify the reputation-related costs which will prevent termination without good cause.²⁰ First, a franchisor who terminates the contract without good cause will encounter difficulties in retaining its other franchisees. The franchisor, as Richard Epstein explains, "has incentives to act in

¹⁶ See generally James A. Brickley et al., *The Economic Effects of Franchise Termination Laws*, 34 J.L. & ECON. 101, 103–09, 130 (1991); Benjamin Klein, *The Economics of Franchise Contracts*, 2 J. CORP. FIN. 9, 30 (1995); Lockerby, *supra* note 11, at 858–71. Cf. Richard A. Epstein, *Unconscionability: A Critical Reappraisal*, 18 J.L. & ECON. 293, 314–15 (1975).

¹⁷ Another important argument that underlies legal economists' resistance to franchise termination laws is that a statutory "good cause" requirement will increase franchisor termination costs and therefore tempt the franchisee to cheat or at least to free ride. See Brickley et al., *supra* note 16, at 104; Epstein, *supra* note 16, at 314; Matthew Ellman, *Specificity Revisited: The Role of Cross-Investments*, 22 J.L. ECON. & ORG. 234, 250 n.36 (2006); Richard L. Smith II, *Franchise Regulation: An Economic Analysis of State Restrictions on Automobile Distribution*, 25 J.L. & ECON. 125, 136 (1982). For a critical analysis of this argument, see generally Benoliel, *supra* note 13 (arguing that a statutory good cause requirement will enhance joint action, expression of trust, and fair treatment of both contractual parties).

¹⁸ Lockerby, *supra* note 11, at 860.

¹⁹ Charles J. Goetz & Robert E. Scott, *Principles of Relational Contracts*, 67 VA. L. REV. 1089, 1148 (1981). See also James A. Brickley, Sanjog Misra & R. Lawrence Van Horn, *Contract Duration: Evidence from Franchising*, 49 J. L. & ECON. 173, 178 (2006); William L. Killian, *One Person's Sense of Justice Is Another's Sense of Injustice*, 24 FRANCHISE L.J. 3, 6 (2004); Alan J. Meese, *Antitrust Balancing in a (Near) Coasean World: The Case of Franchise Tying Contracts*, 95 MICH. L. REV. 111, 131 n.96 (1996); Daniel Clough, *Law and Economics of Vertical Restraints in Australia*, 25 MELB U. L. REV. 551, 580 (2001).

²⁰ Epstein, *supra* note 16, at 314–15; Benjamin Klein & Lester F. Saft, *The Law and Economics of Franchise Tying Contracts*, 28 J. L. & ECON. 345, 356 (1985).

a manner which . . . appears to be fair if he wishes to retain his other franchisees.”²¹ Second, the franchisor will find it difficult to sell additional franchise units to new franchisees. As Benjamin Klein & Lester Saft explain, “[t]he franchisor is not likely to terminate franchisees merely to confiscate their sunk investments opportunistically because franchisors must be concerned about their reputations when attempting to sell additional franchise locations.”²²

Law-and-economics scholars, therefore, uniformly base their analysis upon the implicit assumption that franchisors have a *static* incentive: to retain their franchisees and to attract new ones. This assumption, however, is dubious, as will be demonstrated in this article.

B. Skepticism about Reputation Effectiveness

A handful of legal scholars manifest skepticism about the effectiveness of the reputation mechanism in controlling opportunistic termination by franchisors. David Charny posed one of the earlier challenges to the effectiveness of the reputation mechanism, in his innovative article of the early 1990s.²³ Charny argues that the reputation mechanism will not function if the

²¹ Epstein, *supra* note 16, at 315. See also Thomas M. Pitegoff, *Franchise Relationship Laws: A Minefield for Franchisors*, 45 BUS. LAW. 289, 310 (1989); Smith, *supra* note 17, at 130; Timothy J. Muris, *Opportunistic Behavior and the Law of Contracts*, 65 MINN. L. REV. 521, 577 (1981); Boyd Allan Byers, *Making a Case for Federal Regulation of Franchise Terminations—A Return-of-Equity Approach*, 19 J. CORP. L. 607, 651 (1994); Clough, *supra* note 19, at 580.

²² Benjamin Klein & Lester F. Saft, *The Law and Economics of Franchise Tying Contracts*, 28 J.L. & ECON. 345, 356 (1985); Roger D. Blair & Francine Lafontaine, *Understanding the Economics of Franchising and the Laws That Regulate It*, 26 FRANCHISE L.J. 55, 64 (2006) (“There may be lurking concern that franchisors may abuse their ability to terminate franchise contracts in some opportunistic fashion. Although there is some possibility of such opportunism, there is scant empirical support for it. From an economic perspective, this is not surprising because most franchisors want to expand, and it is hard to attract new franchisees while abusing incumbent franchisees.”); See also Epstein, *supra* note 16, at 315; Lockerby, *supra* note 11, at 859–60; Pitegoff, *supra* note 21, at 310; Smith, *supra* note 17, at 130; Robert W. Emerson, *Franchise Terminations*, *supra* note 8, at 586; Alan J. Meese, *Regulation of Franchisor Opportunism and Production of the Institutional Framework: Federal Monopoly or Competition Between the States?*, 23 HARV. J.L. & PUB. POL’Y 61, 68 (1999); Jonathan Klick et al., *Incomplete Contracts and Opportunism in Franchising Arrangements: The Role of Termination Clauses* 16–17 (Am. L. & Econ. Ass’n Ann. Meetings, Working Paper No. 61, 2006), available at http://www.law.georgetown.edu/olin/papers/klick_paper.pdf; Jonathan Klick et al., *The Effect of Contract Regulation: The Case of Franchising* 33 (Geo. Mason L. & Econ. Res. Paper No. 07-03, 2006), available at <http://ssrn.com/abstract=951464>; Henry N. Butler & Barry D. Baysinger, *Vertical Restraints of Trade As Contractual Integration: A Synthesis of Relational Contracting Theory, Transaction-Cost Economics, and Organization Theory*, 32 EMORY L.J. 1009, 1092 (1983); Byers, *supra* note 21, at 651; Clough, *supra* note 19, at 580; Muris, *supra* note 21, at 577; Killion, *supra* note 19, at 6.

²³ David Charny, *Nonlegal Sanctions in Commercial Relationships*, 104 HARV. L. REV. 373 (1990).

franchisor ceases to use franchisees altogether.²⁴ He suggests that this may occur once the franchisor “has established a market niche” or when the acquirer in a takeover puts in place a new management team that decides to terminate existing franchises.²⁵

Skepticism about the potential for reputation to control opportunistic termination by franchisors was also manifested in a seminal article on franchise contracts by Gilliland Hadfield.²⁶ First, Hadfield, similar to Charny, argues that the reputation mechanism will not function if the franchisor stops selling franchisees.²⁷ She furthermore suggests that this may happen if the franchisor reaches a limit on the number of the franchise units it is able to sell.²⁸ This limitation may often exist, she maintains, as franchisors provide their franchisees with territorial exclusivity.²⁹ Hadfield’s second argument highlights the problematic fact that the reputation mechanism depends upon inferences, drawn by potential franchisees, about the franchisor’s past behavior.³⁰ She claims, “the interdependence, the uncertainty, and the length of the relationship, as well as the inexperience of the franchisee all make the identification of franchisor opportunism very difficult.”³¹

Over the next few years, other scholars joined in the dispute over the effectiveness of the reputation mechanism in the context of franchising. Warren Grimes, for instance,³² argues that a franchisor is likely to be less constrained by reputation concerns in several circumstances. First, a franchisor is likely to be less constrained by reputation concerns if a myth of high profitability surrounds franchising or certain types of franchising.³³ Second, the reputation mechanism will be less effective if the franchisor can counter reputation damage by lowering the franchise fee.³⁴ Third, the reputation mechanism is less likely to operate as expected if the franchisor can make less-informed franchisees the target of opportunism.³⁵ Fourth, the reputation mechanism’s effectiveness is dubious if the franchisor has a short-term

²⁴ *Id.* at 434.

²⁵ *Id.* at 435.

²⁶ Hadfield, *supra* note 1.

²⁷ *Id.* at 978 n.232.

²⁸ *Id.*

²⁹ *Id.*

³⁰ *Id.*

³¹ *Id.*

³² Warren S. Grimes, *When Do Franchisors Have Market Power? Antitrust Remedies for Franchisor Opportunism*, 65 ANTITRUST L.J. 105 (1996).

³³ *Id.* at 130–31.

³⁴ *Id.* at 131–33.

³⁵ *Id.* at 133–34.

perspective.³⁶ Finally, the reputation mechanism will be less effective if the franchisor misapprehends the extent of reputation damage.³⁷

While I agree with many of the scholarly objections leveled at the reputation mechanism in the past, the next part of this article contends that the reputation mechanism also suffers from an inherent deficiency: its limited life cycle. The three major stages of the reputation life cycle will be presented in chronological order: creation, erosion and collapse. A detailed analysis of the life cycle will reveal that the reputation mechanism is often, by its very nature, doomed to failure.

II. REPUTATION LIFE CYCLE

A. Stage One: Creation

At the early stages of the franchisor organizational life cycle, the reputation mechanism may have some deterring effects on the franchisor; the franchisor may be particularly discouraged from terminating a franchise contract without good cause. At these early stages, the incentive of the franchisor is normally to attract new franchisees (or other types of intermediaries, such as dealers and distributors) and thus the franchisor attaches greater importance on being perceived as fair in the eyes of potential franchisees.³⁸ The initial incentive to attract new franchisees derives from three cumulative factors, which will be explained in greater detail below. First, franchisors ordinarily wish to rapidly expand the market for their products by opening many pioneering outlets in various locations.³⁹ Second, a young franchisor normally lacks the significant financial resources necessary to promptly expand the market on its own.⁴⁰ Third, new franchisees, attracted by the young franchisor, can effectively support the franchisor in overcoming its initial financial constraints.⁴¹

³⁶ *Id.* at 134–35.

³⁷ *Id.* at 135.

³⁸ See *supra* note 22 and accompanying text.

³⁹ Alfred R. Oxenfeldt & Anthony O. Kelly, *Will Successful Franchise Systems Ultimately Become Wholly-Owned Chains?*, 44 J. RETAILING 69, 74 (1968–69). For a detailed discussion on this first factor see *infra* notes 42–59.

⁴⁰ Rajiv P. Dant et al., *Ownership Redirection in Franchised Channels*, 11 J. PUB. POL'Y. & MKTG. 33, 34 (1992). For a detailed discussion on this second factor see *infra* notes 60–72.

⁴¹ Scott A. Shane, *Hybrid Organizational Arrangements and Their Implications for Firm Growth and Survival: A Study of New Franchisors*, 39 ACAD. MGMT. J. 216, 220 (1996) (describing the types of costs saved by the franchisor). For a detailed discussion on this third factor see *infra* notes 73–105.

Franchisors generally hope to quickly expand the market for their products by opening many pioneering outlets in various locations.⁴² Swift expansion generates several benefits for franchisors. First, a franchisor, entering the market early can snag superior geographic locations before the market becomes saturated by competitors.⁴³ A franchisor who quickly develops its franchise system, on valuable real estate, may preempt its competitors;⁴⁴ followers will face difficulties in gaining possession of the same valuable real estate, already held by the fast franchisor.⁴⁵ McDonald's, for example, effectively preempted precious real estate from its competitors, through favorable leases.⁴⁶ Through its early expansion in the relatively novel sector of fast food, McDonald's benefited from a first-move advantage.⁴⁷ In addition, a franchisor who quickly develops its franchise system, may be able to lease or acquire assets in the early stages of the market at prices below those that will prevail in later stages, after the market has evolved.⁴⁸

Rapid expansion not only gives franchisors an edge over competitors, but also allows the franchisor to shape the market.⁴⁹ A franchisor that promptly expands the market for its products may influence customer preferences.⁵⁰ In the early stages of many markets, consumers may not have a strong and precise opinion about their preferred value of the new product ingredients.⁵¹ They may also not have a strong inclination about their preferred combination of the product features.⁵² For example, when Coca Cola first penetrated the market, not many customers had a predetermined preference for how carbonated or sweet a cola should be.⁵³ A successful early franchisor is likely to have a seminal influence on the customers' preferences regarding

⁴² Oxenfeldt & Kelly, *supra* note 39, at 74.

⁴³ Victoria Bordonaba Juste et al., *Franchise Firm Entry Time Influence on Long-term Survival*, 37 INT'L J. RETAIL & DISTRIB. MGMT. 106, 109 (2009).

⁴⁴ Oxenfeldt & Kelly, *supra* note 39, at 74; Steven C. Michael, *First Mover Advantage Through Franchising*, 18 J. BUS. VENTURING 61, 65 (2003). *See generally* Marvin B. Lieberman & David B. Montgomery, *First-Mover Advantages*, 9 STRAT. MGMT. J. 41, 44 (1988).

⁴⁵ Michael, *supra* note 44, at 65. *See generally* Birger Wernerfelt, *A Resource-based View of the Firm*, 5 STRAT. MGMT. J. 171, 173 (1984).

⁴⁶ JOHN F. LOVE, MCDONALD'S: BEHIND THE ARCHES 164-65 (1986).

⁴⁷ Michael, *supra* note 44, at 65.

⁴⁸ Lieberman & Montgomery, *supra* note 44, at 44; Michael, *supra* note 44, at 65.

⁴⁹ Gregory S. Carpenter & Kent Nakamoto, *Consumer Preference Formation and Pioneering Advantage*, 26 J. MARK. RES. 285, 286 (1989).

⁵⁰ Michael, *supra* note 44, at 66.

⁵¹ Carpenter & Nakamoto, *supra* note 49, at 286; Lieberman & Montgomery, *supra* note 44, at 46.

⁵² Carpenter & Nakamoto, *supra* note 49, at 286.

⁵³ *Id.*

the value of the product features.⁵⁴ A franchisor may also influence customer preferences as to the combination of the product features.⁵⁵ Coca Cola, for example, may have critically impacted the development of customers' preferences for colas.⁵⁶ The influence on customers' preferences may ultimately shift customers to favor the pioneer franchisor over later competitors,⁵⁷ leading to a market share benefit. Rapid expansion, therefore, is quite beneficial for franchisors. Empirical studies confirm that swift expansion is advantageous. Examining 137 U.S. franchisors in the restaurant industry, Steven Michael found that a rapid expansion strategy of franchisors leads to superior outlet share and ultimately to higher profitability.⁵⁸ Similarly, data covering 188 Spanish franchise chains for an eight-year period show that an early entry strategy of franchisors leads to lower failure risk.⁵⁹

Nonetheless, despite the advantages, maneuvering an early entrance into the franchise market is difficult. A young franchisor, wishing to rapidly expand the market for its products on its own, normally lacks the significant financial resources needed to fulfill its desire.⁶⁰ Opening new outlets is often too costly for a financially immature franchisor.⁶¹ Three significant costs must normally be incurred by the young franchisor: information costs, start-up costs, and management costs.⁶²

To begin with, financial resources are required to gather information regarding the local market in which the franchisor may potentially expand.⁶³ When a young franchisor system expands beyond its original geographic location, it generally

⁵⁴ *Id.*; Lieberman & Montgomery, *supra* note 44, at 46.

⁵⁵ Carpenter & Nakamoto, *supra* note 49, at 286.

⁵⁶ *Id.*; Lieberman & Montgomery, *supra* note 44, at 46.

⁵⁷ For empirical support, see Carpenter & Nakamoto, *supra* note 49, at 286, 294; Frank H. Alpert & Michael A. Kamins, *An Empirical Investigation of Consumer Memory, Attitude, and Perceptions toward Pioneer and Follower Brands*, 59 J. MKTG. 34, 36-37, 41 (1995).

⁵⁸ Michael, *supra* note 44, at 66-67, 69-71, 74. See also Regis Coeurderoy & Rodolphe Durand, *Leveraging the Advantage of Early Entry: Proprietary Technologies Versus Cost Leadership*, 57 J. BUS. RES. 583, 584, 587, 589 (2004).

⁵⁹ Juste et al., *supra* note 43, at 109, 112, 115.

⁶⁰ Rajiv P. Dant et al., *Ownership Redirection in Franchised Channels*, 11 J. PUB. POL'Y & MKTG. 33, 34 (1992); Richard E. Caves & William F. Murphy II, *Franchising: Firms, Markets, and Intangible Assets*, 42 S. ECON. J. 572, 581 (1976); Francine Lafontaine & Patrick J. Kaufmann, *The Evolution of Ownership Patterns in Franchise Systems*, 70 J. RETAILING 97, 99 (1994).

⁶¹ Lafontaine & Kaufmann, *supra* note 60, at 99.

⁶² Rajiv P. Dant & Audhesh K. Paswan, *The Effect of Financial Resource Availability and System Size on Ownership Redirection in Franchised Channels of Distribution*, 10 ASIA PAC. J. MKTG & LOGISTICS 85, 86 (1998).

⁶³ Lafontaine & Kaufmann, *supra* note 60, at 99.

lacks information regarding other local markets.⁶⁴ Therefore, the franchisor needs to invest capital in searching for and identifying suitable locations.⁶⁵ It should ideally also evaluate the local market conditions of each potential location.⁶⁶ More specifically, the franchisor will have to invest resources in learning about local marketing strategies, input suppliers, and customer preferences, at each potential location.⁶⁷ These information-gathering costs are likely to augment with increases in the unfamiliarity, diversity, and uncertainty of local markets.⁶⁸

The franchisor, wishing to expand on its own, will furthermore incur considerable start-up costs in establishing each outlet. The total start-up costs of each outlet can range from \$20,000, at a minimum, to over \$1,000,000, depending on the franchise chosen.⁶⁹ Costs may involve renting or building an outlet, stocking it with inventory and equipment, and advertising it.⁷⁰ Finally, a franchisor, ready and willing to promptly expand on its own, will incur significant management costs. These costs include the necessary capital needed to select skilled managers to operate each outlet (*selection costs*).⁷¹ They also include expenditures for monitoring the managers' performance (*monitoring costs*).⁷²

Since expanding rapidly is usually too costly for a young franchisor, its initial incentive will be to attract new franchisees.⁷³ New franchisees will normally be able to

⁶⁴ Thomas Bürkle & Thorsten Posselt, *Franchising as a Plural System: A Risk-Based Explanation*, 84 J. RETAILING 39, 41 (2008).

⁶⁵ James G. Combs et al., *A Strategic Groups Approach to the Franchising-Performance Relationship*, 19 J. BUS. VENTURING 877, 880 (2004).

⁶⁶ *Id.*

⁶⁷ Alanson P. Minkler, *Why Firms Franchise: A Search Cost Theory*, 148 J. INST. THEORETICAL ECON. 240, 242 (1992).

⁶⁸ *Id.* at 242–43.

⁶⁹ International Franchise Association, *Frequently Asked Questions About Franchising*, <http://www.franchise.org/industrysecondary.aspx?id=10008> (last visited Nov. 12, 2009).

⁷⁰ Justin G. Longenecker et al., *SMALL BUSINESS MANAGEMENT: LAUNCHING AND GROWING ENTREPRENEURIAL VENTURES* 99–101 (14th ed. 2007).

⁷¹ See Shane, *supra* note 41, at 219; Seth W. Norton, *An Empirical Look at Franchising as an Organizational Form*, 61 J. BUS. 197, 204 (1988); Combs et al., *supra* note 65, at 880.

⁷² See Shane, *supra* note 41, at 219–20; Seth W. Norton, *Franchising, Brand Name Capital, and the Entrepreneurial Capacity Problem*, 9 STRAT. MGMT. J. 105, 106 (1988); James A. Brickley & Fredrick H. Dark, *The Choice of Organizational Form: The Case of Franchising*, 18 J. FIN. ECON. 401, 403 (1987).

⁷³ Rajiv P. Dant et al., *supra* note 60, at 34; Brent L. Baker & Rajiv P. Dant, *Stable Plural Forms in Franchise Systems: An Examination of the Evolution of Ownership Redirection Research*, in STRATEGY AND GOVERNANCE OF NETWORKS: COOPERATIVES, FRANCHISING, AND STRATEGIC ALLIANCES 87, 92 (George Hendrikse et al. eds. 2008); Shelby D. Hunt, *The Trend Toward Company-Operated Units in Franchise Chains*, 49 J.

effectively assist the financially immature franchisor in overcoming its initial economic constraints. To begin with, franchisees often incur the information costs needed to find suitable locations for expansion in locations which are unfamiliar to the franchisor.⁷⁴ Notably, franchisees are likely to have valuable information regarding potential local markets, which the youthful franchisor lacks.⁷⁵ This is mainly because they usually reside in the community in which the new unit is planned to be established.⁷⁶

In addition, franchisees will commonly incur the start-up costs involved in establishing new outlets, including equipment, inventory building, and real estate costs.⁷⁷ Franchisees furthermore reduce the management costs needed to operate an outlet.⁷⁸ First, they reduce the selection costs needed to recruit qualified managers.⁷⁹ Franchisees typically agree to invest heavily in the outlet and to receive a residual claim on the outlet's uncertain profits.⁸⁰ By doing so, franchisees signal their management capabilities, compared to salaried managers, who are willing to take the job.⁸¹ Second, franchisees reduce the monitoring costs needed to scrutinize the performance of managers.⁸² Since franchisees invest significant resources in the outlet and have a residual claim on the outlet's profit, they are less likely than salaried managers, to put forth a suboptimal effort level.⁸³

Franchisees support the franchisor in overcoming its initial financial constraints, not only by incurring and reducing significant costs, but also by directly providing the franchisor significant capital.⁸⁴ First, if the franchise involves the sale of products by the franchisor to the franchisee, the sale price will usually comprise a mark-up over cost representing the

RETAILING 3, 4 (1973); Oxenfeldt & Kelly, *supra* note 39, at 71; Caves & Murphy, *supra* note 60, at 581.

⁷⁴ Combs et al., *supra* note 65, at 880. Sometimes, franchisors select the location, but then they collect fees from franchisees for this service. Hadfield, *supra* note 1, at 936.

⁷⁵ Bürkle & Posselt, *supra* note 64, at 41; Minkler, *supra* note 67, at 243; Anna Watson et al., *Retail Franchising: An Intellectual Capital Perspective*, 12 J. RETAILING CONSUMER SERVICES 25, 28 (2005).

⁷⁶ Bürkle & Posselt, *supra* note 64, at 41; Combs et al., *supra* note 65, at 880; Lafontaine & Kaufmann, *supra* note 60, at 99.

⁷⁷ Lafontaine & Kaufmann, *supra* note 60, at 99.

⁷⁸ Combs et al., *supra* note 65, at 880.

⁷⁹ Shane, *supra* note 41, at 220.

⁸⁰ *Id.* at 220; Norton, *supra* note 71, at 205; Combs et al., *supra* note 65, at 880.

⁸¹ Shane, *supra* note 41, at 220.

⁸² *Id.* at 219–21.

⁸³ Brickley & Dark, *supra* note 72, at 404–05; Shane, *supra* note 41, at 221; Combs et al., *supra* note 65, at 880; Watson et al., *supra* note 75, at 28.

⁸⁴ Hadfield, *supra* note 1, at 935.

franchisor's return.⁸⁵ In addition, each franchisee ordinarily pays the franchisor an initial lump-sum fee, called a franchise fee⁸⁶—generally ranging between \$10,000 and \$30,000.⁸⁷ Moreover, each franchisee provides a constant and on-going stream of capital to the franchisor, via royalty rates.⁸⁸ Normally, royalty rates are calculated as a percentage of sales revenues.⁸⁹ The average royalty rate paid in 2001 was 5.2%, which in real terms translates to about \$500 per month.⁹⁰ Many franchisees also pay the franchisor advertising fees,⁹¹ to support franchisors in advertising their product or service regionally or nationally.⁹² These fees are most often a steady percentage, usually one to two percent,⁹³ of the franchisee's monthly sales revenue.⁹⁴ Finally, franchisees occasionally pay fees to franchisors for bookkeeping and management consultation services.⁹⁵

Given the significant capital that franchisees provide to the youthful franchisor, and their ability to support him in overcoming its financial constraints,⁹⁶ one would posit that the initial incentive of the franchisor will be to attract franchisees. This theoretical contention is, in fact, empirically supported. Francine Lafontaine conducted a survey of, among other things, franchisors' motivation for using franchisees.⁹⁷ The survey was

⁸⁵ *Id.*

⁸⁶ James G. Combs & David J. Ketchen, Jr., *Can Capital Scarcity Help Agency Theory Explain Franchising? Revisiting the Capital Scarcity Hypothesis*, 42 ACAD. MGMT. J. 196, 199 (1999); Lafontaine & Kaufmann, *supra* note 60, at 99; Hadfield, *supra* note 1, at 935.

⁸⁷ International Franchise Association, Frequently Asked Questions About Franchising, <http://www.franchise.org/industrysecondary.aspx?id=10008> (last visited Nov. 12, 2009). For statistics on the variation in franchise fees, see generally BLAIR & LAFONTAINE, *supra* note 9, at 56–61.

⁸⁸ Hadfield, *supra* note 1, at 935; Watson, *supra* note 75, at 25.

⁸⁹ BLAIR & LAFONTAINE, *supra* note 9, at 62.

⁹⁰ *Id.* at 67, 69.

⁹¹ *Id.* at 69.

⁹² *Id.*

⁹³ Longenecker et al., *supra* note 70, at 100.

⁹⁴ BLAIR & LAFONTAINE, *supra* note 9, at 69.

⁹⁵ Hadfield, *supra* note 1, at 936. Notably, franchisors sometimes also obtain a commission from approved suppliers of their new franchisees. *Id.*

⁹⁶ But see P. H. Rubin, *The Theory of the Firm and the Structure of the Franchise Contract*, 21 J. L. & ECON. 223, 225–26 (1978) (suggesting that passive investors, such as stockholders, are a superior source of capital than franchisees). However, for explanations of why franchisees are in fact a superior source of capital than passive investors, see, for example, Robert E. Martin & Robert T. Justis, *Franchising, Liquidity Constraints and Entry*, 25 APPLIED ECON. 1269, 1271–72 (1993); Lafontaine & Kaufmann, *supra* note 60, at 99; Francine Lafontaine, *Agency Theory and Franchising: Some Empirical Results*, 23 RAND J. ECON. 263, 267 (1992).

⁹⁷ Francine Lafontaine, *How and Why do Franchisors Do What They Do: A Survey Report*, in FRANCHISING: PASSPORT FOR GROWTH & WORLD OF OPPORTUNITY, PROCEEDINGS OF THE SIXTH CONFERENCE OF THE SOCIETY OF FRANCHISING (Patrick Kaufmann ed. 1992).

based on the answers given by 130 franchisors, all members of the International Franchise Association, and involved in a variety of businesses.⁹⁸ The majority of the respondents began franchising only recently.⁹⁹ Franchisors participating in the survey explained that they became involved in franchising mainly because it allowed them to expand rapidly, especially because franchisees provide capital to franchisors.¹⁰⁰ In another survey, conducted by Rajiv Dant, of 79 founders and representatives of franchisors, similar results were found.¹⁰¹ Respondents were primarily drawn from the following franchising sectors: fast food and restaurants, gifts and home decorations, hospitality, convenience stores, and services.¹⁰² The majority of respondents classified their business in the introduction or growth stages of its life cycle,¹⁰³ indicating the relative youthfulness of the franchisors.¹⁰⁴ According to the survey, the most cited motivations for attracting franchisees were “market entry and growth” and “capital access and profits.”¹⁰⁵

B. Stage Two: Erosion

As the franchisor’s financial resources increase, the effectiveness of the reputation mechanism will erode. Particularly, the franchisor’s incentive to attract new franchisees will decrease.¹⁰⁶ Similarly, its incentive to maintain its present franchisees will decline.¹⁰⁷ Instead, the franchisor, having greater financial resources, will gradually shift towards owning

⁹⁸ *Id.* at 2, 4.

⁹⁹ *Id.* at 3.

¹⁰⁰ *Id.* at 11.

¹⁰¹ Rajiv P. Dant, *Motivation for Franchising: Rhetoric Versus Reality*, 14 INT’L SMALL BUS. J. 10 (1995).

¹⁰² *Id.* at 15.

¹⁰³ *Id.*

¹⁰⁴ *Id.* at 16.

¹⁰⁵ *Id.* at 23. For further empirical support to the hypothesis that the initial incentive of franchisors is to attract new franchisees due to financial constraints, see Lorelle Frazer, *Why Franchisors Discontinue Franchising but Continue Operating*, 19 INT’L SMALL BUS. J. 29, 32 (2001); David A. Kirby & Anna Watson, *Overcoming the Financial Constraints on Small Firm Growth: The Case of Franchising*, 6 J. OF PRIVATE EQUITY 61, 63 (2003); Anne Marie Doherty, *The Internationalization of Retailing: Factors Influencing the Choice of Franchising as a Market Entry Strategy*, 18 INT’L J. SERV. INDUS. MGMT. 184, 194 (2007); Scott Weaven & Carmel Herington, *Factors Influencing Governance Choice and Human Resource Management Within Services Franchising Networks*, J. MGMT. & ORG. 126, 132 (2007). For indirect empirical support, see Lafontaine & Kaufmann, *supra* note 60, at 110 (“the fact that franchisors [examined in the study] that were subsidiaries of larger corporations owned more of their own units, point toward the idea that access to resources might constrain some firms to operate less units than they would like.”).

¹⁰⁶ Oxenfeldt & Kelly, *supra* note 39, at 75.

¹⁰⁷ *Id.*

the units himself.¹⁰⁸ The franchisor will specifically want to progressively open new outlets under franchisor-ownership, which would have been opened as a franchisee-owned outlet, when the franchisor lacked financial resources. In addition, the franchisor's motivation will be to gradually transform the existing franchisee-owned outlets in his chain into franchisor-owned outlets.

The shift in the franchisor's incentive stems from the fact that the profits that a franchisor can make, by owning a unit, are normally greater than the profits it can amass through a franchisee-owned unit.¹⁰⁹ True, establishing and maintaining a franchisor-owned unit is costly for the franchisor since it involves information, start-up, and management costs. In addition, a franchisee-owned unit provides the franchisor capital through a lump-sum fee, franchise fees, and other fees.¹¹⁰ However, in the long-run, a franchisor-owned unit is more profitable for the franchisor than a franchisee-owned unit.¹¹¹ With its company-owned and operated units, the franchisor continually keeps all of the profits of the unit, whereas with franchised units its chief source of profit is a lump-sum fee and ongoing fees computed as a small percentage of the gross revenue.¹¹²

The superiority of franchisor-owned units over franchisee-owned units in generating franchisor's profits is empirically supported. For example, Thomas Ehrmann and Georg Spranger examine the financials of seven large public U.S. restaurant retail chains through their annual 10-k filings with the U.S. Securities and Exchange Commission.¹¹³ They examine each franchisor's gross margin ratios as well as the operating income ratios for franchisor-owned to franchisee-owned units.¹¹⁴ For each franchisor analyzed in the study, the single franchisor-

¹⁰⁸ *Id.*; Rajiv P. Dant et al., *What We Know About Ownership Redirection in Franchising: A Meta-Analysis*, 72 *J. Retailing* 429, 434 (1996); Caves & Murphy, *supra* note 60, at 581; Gary J. Castrogiovanni et al., *Shifting Imperatives: An Integrative View of Resource Scarcity and Agency Reasons for Franchising*, 30 *ENTREPRENEURSHIP THEORY & PRAC.* 23, 25 (2006); Baker & Dant, *supra* note 73, at 92.

¹⁰⁹ Baker & Dant, *supra* note 73, at 92.

¹¹⁰ See *supra* Part II.A.

¹¹¹ Hunt, *supra* note 73, at 6–7; Hadfield, *supra* note 1, at 936–37; Paul Tracey & Owen Jarvis, *Toward a Theory of Social Venture Franchising*, 31 *ENTREPRENEURSHIP THEORY & PRAC.* 667, 669 (2007); James G. Combs et al., *Franchising: A Review and Avenues to Greater Theoretical Diversity*, 30 *J. MGMT.* 907, 909–10 (2004).

¹¹² Baker & Dant, *supra* note 73, at 92; James G. Combs & Gary J. Castrogiovanni, *Franchisor Strategy: A Proposed Model and Empirical Test of Franchise Versus Company Ownership*, 32 *J. SMALL BUS. MGMT.* 37, 38–39 (1994).

¹¹³ Thomas Ehrmann and Georg Spranger, *Franchisee Versus Company Ownership—An Empirical Analysis of Franchisor Profit*, in *ECONOMICS AND MANAGEMENT OF NETWORKS* 31, 39–40 (Gérard Cliquet et al., eds. 2007).

¹¹⁴ *Id.* at 40.

owned unit added far more to both the gross margin and to the total operating income than the single franchisee-owned unit.¹¹⁵ Similarly, Urban Ozanne and Shelby Hunt found that sales in franchisor-owned outlets in the fast food industry are on average significantly higher than those in franchisee-owned outlets.¹¹⁶

Anecdotal evidence bolsters the contention that franchisor-owned units are superior to franchisee-owned units in generating franchisor profits, as expressed in several classic statements by franchisors. For example, John Jay Hooker famously stated:

As all of you know, the name of the game is not really franchising. The name of the game is company stores. I was looking at some figures not too long ago and saw where a big company in America has 1,600–1,700 units, and only two hundred of those were company-owned, but the two hundred company-owned units were producing sixty percent of the net after taxes. . . the real name of the game is owning the stores yourself.¹¹⁷

Similarly, Lawrence E. Singer, who was the president of Royal Castle System Inc. has noted: “We make more profit, per unit, than we could possibly make in franchising. This fact has been acknowledged by many of the franchise operators.”¹¹⁸

Given the superiority of franchisor-owned units over franchisee-owned units, the effectiveness of the reputation mechanism is likely to be eroded when the franchisor’s financial constraints decrease. As the resources needed for establishing and operating units become more attainable by the franchisor, forgoing the higher levels of profit attained through franchisor ownership ceases to be necessary.¹¹⁹ Indeed, empirical studies indicate that as the franchisor’s financial resources increase, the reputation mechanism erodes. These studies particularly

¹¹⁵ *Id.* at 41.

¹¹⁶ S. Select Comm. on Small Business, 92d Cong., 1st Sess., Report on the Economic Effects of Franchising 87 (Comm. Print 1971) (written by Urban B. Ozanne & Shelby D. Hunt); Hunt, *supra* note 73, at 7.

¹¹⁷ John Jay Hooker, *The Story of Minnie Pearl—A Case History of One New Company’s Trials, Tribulations and Triumphs*, in FRANCHISING TODAY: REPORT ON THE FIFTH INTERNATIONAL MANAGEMENT CONFERENCE ON FRANCHISING 164, 171 (Charles L. Vaughn ed. 1970).

¹¹⁸ Lawrence E. Singer, *To Franchise or Not—How to Decide: Con—Why Company-Owned Units*, in FRANCHISING TODAY: 1966–1967 23, 24 (Charles L. Vaughn & David B. Slater eds. 1967). Likewise, William Ware of PKI Foods, Inc., stated:

More profits can be gained from company-owned stores. A break-even analysis of operations in the \$225,000 to \$300,000 range shows this fact to be very true. Since our stores average \$270,000 per unit per year and some clear between \$50,000 and \$100,000, it becomes evident that our stores are more profitable than if franchised.

William Ware, *To Franchise or Not—How to Decide: Con—Why Company-Owned Units*, in FRANCHISING TODAY: 1966–1967 27, 29 (Charles L. Vaughn & David B. Slater eds. 1967).

¹¹⁹ Baker & Dant, *supra* note 73, at 92.

confirm that as the financial availability of the franchisor increases, its incentive will gradually shift away from attracting new franchisees and maintaining the present ones, towards owning prospective and present outlets.¹²⁰ For example, Rajiv Dant and Patrick Kaufmann examine the changes in ownership patterns of franchise systems as the franchisor financially matures.¹²¹ They utilize a national mail survey of the franchised fast food restaurant industry franchisors, and analyze a total of 152 questionnaires filled out by franchisors.¹²² The questionnaire asks the franchisor respondent to provide details of ownership patterns, unit reacquisition activity, new units sale activity, and franchisor's resource availability position.¹²³ The questionnaires were most frequently completed by presidents, CEOs, or founders of the franchise.¹²⁴ The typical franchisor commenced its operation in 1966, sold its first franchise in 1974, and had about 200 outlets.¹²⁵ Dant and Kaufmann's study confirms the validity of the hypothesis that the greater the franchisor's internal access to financial, informational, and managerial resources, the more likely the strategic tendency toward company ownership of outlets.¹²⁶ The strategic tendency toward company ownership of outlets is measured by two factors. First, the inclination of franchisors to substitute *existing* franchisee-owned outlets with franchisor-owned outlets (*retrospective substitution*), was considered.¹²⁷ This inclination is captured mainly by comparing (a) the number of franchisee-owned units reported by the franchisor to have been permanently converted into franchisor-owned units and (b) the number of franchisor-owned units

¹²⁰ Notably, other contradicting empirical studies also examine the relationship between the size and age of the franchisor and its ownership patterns. See, for example, Hunt, *supra* note 73, at 9; Scott Shane, *Explaining the Distribution of Franchised and Company-Owned Outlets in Franchise Systems*, 24 J. MGMT. 717, 724–25, 730, 735 (1998); Gary J. Castrogiovanni et al., *Resource Scarcity and Agency Theory Predictions Concerning the Continued Use of Franchising in Multi-Outlet Networks*, 44 J. SMALL BUS. MGMT. 27, 30, 36–37 (2006); Lorelle Frazer, *Causes of Disruption to Franchise Operations*, 54 J. BUS. RES. 227, 229–30, 232 (2001); Lafontaine & Kaufmann, *supra* note 60, at 110; Ilan Alon, *The Use of Franchising by U.S.-Based Retailers*, 39 J. SMALL BUS. MGMT. 111, 117 (2001); Combs & Castrogiovanni, *supra* note 112, at 45.

¹²¹ Rajiv P. Dant & Patrick J. Kaufmann, *Structural and Strategic Dynamics in Franchising*, 79 J. RETAILING 63 (2003).

¹²² *Id.* at 67.

¹²³ *Id.*

¹²⁴ *Id.* at 68.

¹²⁵ *Id.*

¹²⁶ *Id.* at 72. For similar results and dataset, see Patrick J. Kaufmann & Rajiv P. Dant, *Multi-Unit Franchising: Growth and Management Issues*, 11 J. BUS. VENTURING 343, 351, 354 (1996); Rajiv P. Dant & Patrick J. Kaufmann, Robert A. Robicheaux, *Changes in the Mix of Company-Owned and Franchised Outlets: Ownership Redirection Revisited*, in PROCEEDINGS OF THE INTERNATIONAL SOCIETY OF FRANCHISING CONFERENCE 4, 13 (1998).

¹²⁷ Dant & Kaufmann, *supra* note 121, at 67.

reported by the franchisor to have been permanently converted into franchisee-owned units.¹²⁸ The strategic tendency toward company ownership is furthermore measured by the propensity of franchisors to open *new* outlets under franchisor-ownership which would have been opened as a franchisee-owned outlet earlier in the franchisor's history (*prospective substitution*).¹²⁹ This tendency is calculated by comparing (a) the current year's pattern of new units opening activity to (b) the pattern of ownership as established by previous year's composition of franchisor-owned units versus franchisee-owned units.¹³⁰

Empirical support for the erosion of the reputation mechanism in several key industries was also established in a study by Rajiv Dant and Audhesh Paswan. Dant and Paswan culled data from successive volumes of *Franchising in the Economy* to test, among other things, the shifts in ownership patterns of franchise systems as the franchisor financially matures,¹³¹ spanning the period of 1977 through 1986.¹³² They find in the rentals (equipment) industry, as well as in the hotels, motels and campground industry, that as the financial resource availability of franchisor increases there will be greater incidence of ownership redirection favoring the franchisor.¹³³ Ownership redirection, in this context, was captured by the inclination of franchisors to substitute existing franchisee-owned outlets with franchisor-owned outlets. This inclination is measured mainly by comparing the number of franchisee-owned units converted to franchisor-owned units with the number of franchisor-owned units converted to franchisee-owned units.¹³⁴

Another empirical study, conducted by James Combs and David Ketchen, indicates that as the financial resources of the franchisor increase, the reputation mechanism erodes.¹³⁵ Combs and Ketchen's sample is drawn from the restaurant industry during the years 1989 through 1993.¹³⁶ Their study relies mainly on franchisor's annual reports and Compact Disclosure, a comprehensive database of public firms.¹³⁷ Examining data from 91 restaurant chains, Combs and Ketchen empirically confirm the hypothesis that the less a franchisor encounters difficulties

¹²⁸ *Id.*

¹²⁹ *Id.* at 67–68.

¹³⁰ *Id.* at 68.

¹³¹ Dant & Paswan, *supra* note 62, at 90.

¹³² *Id.*

¹³³ *Id.* at 93, 96, 99.

¹³⁴ *Id.* at 93.

¹³⁵ Combs et al., *supra* note 86, at 199.

¹³⁶ *Id.*

¹³⁷ *Id.* at 199, 201.

stemming from a lack of capital, the less the franchisor will rely on expansion through franchising as opposed to franchisor ownership.¹³⁸

C. Stage Three: Collapse

Following the continuing decrease in the franchisor's financial constraints and the decrease in its incentive to maintain and attract new franchisees, the reputation mechanism may collapse. This may particularly occur when the reputation-related costs decrease to a minimum level, at which they will be lower than the franchisor benefits from opportunistic termination of the franchise contract. Such a low level is reached when the franchisor accrues sufficient capital to manage the whole franchise chain on its own. At this stage, the franchisor's reputation costs from being perceived as unjust by present and prospective franchisees are likely to be lower than the benefits from opportunistic termination. Termination without good cause therefore is likely to occur.

An analysis of case law on dealership and distributorship relationships, which are often governed by franchise termination laws, arguably reveals the collapse of the reputation mechanism that follows a gradual improvement in a franchisor's financial position. An early illustrative case is *Kealey Pharmacy & Home Care Servs. Inc. v. Walgreen Co.*¹³⁹ Between 1972 and 1979, Walgreen sold a wide variety of drugs, beauty aids, and household commodities to consumers through a nation-wide system of both Walgreen-owned stores and dealers using the Walgreen brand-name.¹⁴⁰ The dealers independently owned drugstores pursuant to a standard dealership agreement with Walgreen.¹⁴¹ Throughout the contractual relationship between Walgreen and its dealers, the dealers provided capital to Walgreen in two chief ways. First, the dealers paid for merchandise which they purchased from Walgreen.¹⁴² In addition, the dealers provided capital to Walgreen indirectly by

¹³⁸ Combs et al., *supra* note 86, at 199, 202. For similar results and dataset, see Vera L. Hoover, David J. Ketchen, Jr. & James G. Combs, *Why Restaurant Firms Franchise: An Analysis of Two Possible Explanations*, 44 CORNELL HOTEL AND REST. ADMIN. Q. 9, 14 (2003); see also James G. Combs & David J. Ketchen, Jr., *Explaining Interfirm Cooperation and Performance: Toward a Reconciliation of Predictions from the Resource-Based View and Organizational Economics*, 20 STRAT. MGMT. J. 867, 872, 880 (1999).

¹³⁹ *Kealey Pharmacy & Home Care Servs., Inc. v. Walgreen Co.*, 761 F.2d 345, 347 (7th Cir. 1985).

¹⁴⁰ *Kealey Pharmacy & Home Care Serv., Inc. v. Walgreen Co.*, 539 F. Supp. 1357, 1360–1361 (W.D. Wis. 1982), *aff'd*, 761 F.2d 345 (7th Cir. 1985).

¹⁴¹ *Kealey Pharmacy & Home Care Servs., Inc.*, 761 F.2d, at 347.

¹⁴² *Kealey Pharmacy & Home Care Serv., Inc. v. Walgreen Co.*, 607 F. Supp. 155, 158 (W.D. Wis. 1984), *aff'd in part, vacated in part*, 761 F.2d 345 (7th Cir. 1985).

promoting the Walgreen name, image, and reputation in their local communities.¹⁴³ Arguably, Walgreen relied on these dealers to provide the significant financial resources needed to promote Walgreen products, which Walgreen lacked at that stage. During the contractual relationships of Walgreen with its dealers, Walgreen's net sales were almost constantly increasing.¹⁴⁴ For example, Walgreen's net sales in 1972 were \$863,334,000;¹⁴⁵ whereas by 1979, Walgreen's net sales had jumped to \$1,344,542,000.¹⁴⁶ Figure 1 outlines Walgreen's increase in net sales between 1972 and 1979.

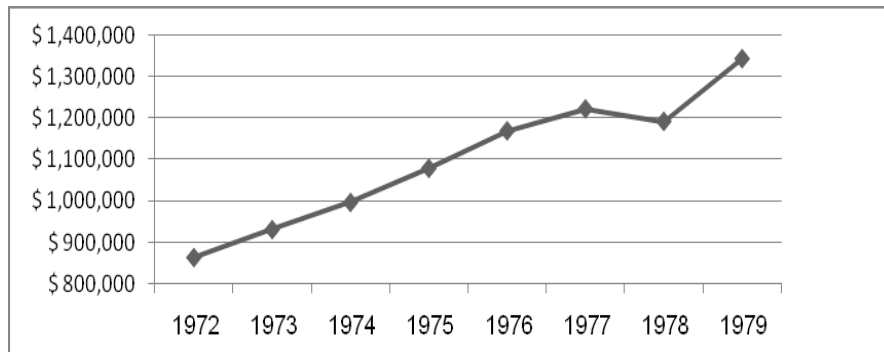


Figure 1: Walgreen's Net Sales: 1972–1979 (in thousands)

In 1980, following the constant increase in Walgreen's annual net sales, it arguably reached a stage in its organizational life cycle in which it had enough financial resources at its disposal to operate all the stores on its own.¹⁴⁷ Indeed, in that year, Walgreen decided to adjust its marketing strategy. In particular, Walgreen made a decision to cease its reliance on dealers.¹⁴⁸ Instead, Walgreen chose to maintain and increase its own stores in the same geographic areas where the dealers had already helped build up Walgreen's reputation and image.¹⁴⁹ At this stage, the reputation mechanism collapsed. Walgreen was no longer interested in relying on the financial resources

¹⁴³ *Kealey Pharmacy & Home Care Serv. Inc.*, 607 F. Supp. at 159. For example, all of the prescription drugs that the dealers sold included the words "Walgreen Agency." *Id.*

¹⁴⁴ These are the annual net sales of Walgreen during 1972 and 1979 (in thousands): \$863,334 (1972), \$930,898 (1973), \$996,616 (1974), \$1,079,143 (1975), \$1,169,779 (1976), \$1,223,249 (1977), \$1,192,855 (1978), \$1,344,542 (1979). See WALGREEN CO., ANNUAL REPORTS (1972–1979).

¹⁴⁵ WALGREEN CO., 1972 ANNUAL REPORT (1972).

¹⁴⁶ WALGREEN CO., 1979 ANNUAL REPORT (1979).

¹⁴⁷ See generally WALGREEN CO., ANNUAL REPORTS (1972–1979); *Kealey Pharmacy & Home Care Servs., Inc. v. Walgreen Co.*, 761 F.2d 345, 347 (7th Cir. 1985).

¹⁴⁸ *Kealey Pharmacy & Home Care Servs.* 761 F.2d at 350.

¹⁴⁹ *Id.*

provided by the dealers, and thus had no reason to fear a blow to its reputation.¹⁵⁰ Termination without good cause was likely to occur. In fact, Walgreen terminated its dealership agreements absent any showing that the dealers failed to comply with the requirements imposed by Walgreen.¹⁵¹

Another important case which reveals the collapse of the reputation mechanism following a gradual improvement in the franchisor's resource availability is *Atlantic City Coin & Slot Serv. Co. v. IGT*.¹⁵² International Game Technology (IGT) designs and manufactures electronic gaming devices.¹⁵³ During 1983 and 1997 IGT expanded its products in New Jersey, Maryland, and the Caribbeans using an exclusive distributor.¹⁵⁴ Throughout the contractual relationship between the IGT and the distributor, the latter provided significant financial assistance to IGT.¹⁵⁵ More specifically, the exclusive distributor invested significant financial resources in promoting the IGT name, expanding a sales force, expanding a marketing facility, expanding a storage area for IGT products, and leasing warehousing devoted to storing, testing, and servicing of IGT gaming devices.¹⁵⁶ Arguably, IGT relied on its exclusive distributor since it could provide the significant resources necessary to promote IGT products—resources which IGT lacked at that early stage. Throughout its contractual relationship with the exclusive distributor, IGT's annual total revenues were almost constantly increasing.¹⁵⁷ For example, in 1983 IGT total revenues were \$60,032,000;¹⁵⁸ whereas by 1997, IGT total revenues had jumped to \$743,970,000.¹⁵⁹ Figure 2 demonstrates graphically IGT's increase in total revenues between 1983 and 1997.

¹⁵⁰ See generally Combs et al., *supra* note 86 (discussing the deterioration of the reputation mechanism that corresponds with an increase in financial resources).

¹⁵¹ *Kealey Pharmacy & Home Care Servs., Inc.* 761 F.2d at 350.

¹⁵² *Atlantic City Coin & Slot Serv. Co. v. IGT*, 14 F. Supp. 2d 644 (D.N.J. 1998).

¹⁵³ *Id.* at 647.

¹⁵⁴ *Id.* at 647, 649.

¹⁵⁵ *Id.* at 654.

¹⁵⁶ *Id.* at 646, 653–54.

¹⁵⁷ These are the annual total revenues of IGT during 1983 and 1997 (in thousands): \$60,032 (1983), \$73,262 (1984), \$56,564 (1985), \$41,574 (1986), \$83,522 (1987), \$98,694 (1988), \$151,152 (1989), \$210,257 (1990), \$237,363 (1991), \$399,443 (1992), \$478,030 (1993), \$674,461 (1994), \$620,786 (1995), \$733,452 (1996), \$743,970 (1997). See INTERNATIONAL GAME TECHNOLOGY, ANNUAL REPORTS (1983–1993); INTERNATIONAL GAME TECHNOLOGY; FORMS 10-K (1994–1997).

¹⁵⁸ INTERNATIONAL GAME TECHNOLOGY, ANNUAL REPORT (1983).

¹⁵⁹ INTERNATIONAL GAME TECHNOLOGY, Form 10-K (1997).

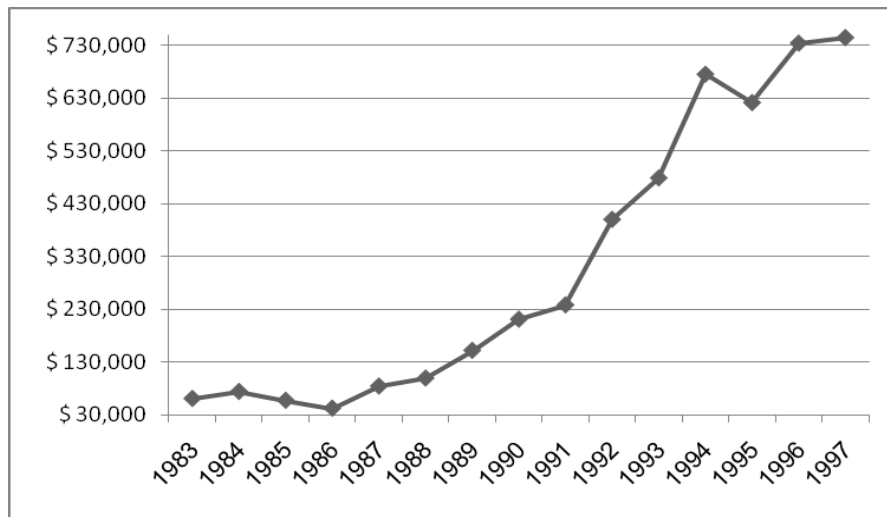


Figure 2: IGT Annual Total Revenue: 1983-1997 (in thousands)

In 1997, following the constant increase in IGT annual total revenues, it arguably reached a stage in its organizational life cycle in which it had access to enough financial resources to replace its exclusive distributor with an IGT-owned distribution system. Predictably in that year, IGT decided to discontinue its relationship with its exclusive distributor.¹⁶⁰ Instead, IGT chose to promote and sell its product on its own.¹⁶¹ At this stage the reputation mechanism collapsed because IGT was no longer interested in relying on distributors.¹⁶² Termination without good cause was likely to occur. Indeed, IGT terminated its distribution agreement with its exclusive distributor in 1998, without any allegations of wrongdoing to justify the termination.¹⁶³ Rather, IGT praised the exclusive distributor's performance just prior to terminating the relationship.¹⁶⁴ Good cause supporting the contract termination was noticeably absent.¹⁶⁵

The collapse of the reputation mechanism is likewise evidenced by the case of *Day Distrib. Co. v. Nantucket Allserve*,

¹⁶⁰ *Atlantic City Coin & Slot Serv. Co. v. IGT*, 14 F. Supp. 2d 644, 658–59 (D.N.J. 1998).

¹⁶¹ *Id.* at 659.

¹⁶² *Id.* at 646.

¹⁶³ *Id.* at 658–59.

¹⁶⁴ *Id.* at 658.

¹⁶⁵ *Id.*

*Inc.*¹⁶⁶ Nantucket Allserve, Inc. (Nantucket) produces a line of upscale sodas known as Stewart's.¹⁶⁷ At the early stages of the organizational life cycle of Nantucket, its resource availability was apparently limited.¹⁶⁸ One of Nantucket's founders, Mr. Scott, was so financially strapped that he slept in his car during the summer of 1991.¹⁶⁹ At that early stage, Nantucket sold its products through distributors¹⁷⁰ who provided financial resources to Nantucket by purchasing high volumes of Nantucket sodas.¹⁷¹ Arguably, Nantucket chose such a marketing strategy due to its financial inability to distribute its products on its own. At this stage in Nantucket's organizational life cycle, the reputation mechanism was relatively effective. Nantucket had to rely on distributors, and therefore was restrained from acting opportunistically towards its distributors.¹⁷²

Nantucket's financial situation improved dramatically over the years. Revenues soared from \$200,000 in 1991 to about \$30 million in 1996.¹⁷³ In addition, the successful Nantucket group was purchased in 2002 by Cadbury Schweppes Plc, a British beverage giant,¹⁷⁴ for an undisclosed sum that was estimated at \$100 million.¹⁷⁵

Following the acquisition of Nantucket by Cadbury, Nantucket presumably reached a stage in its organizational life cycle in which it had enough financial resources to discontinue its reliance on distributors and perform all the distribution functions on its own. In fact, one year following the acquisition of Nantucket by Cadbury, they both began exploring the possibility of discontinuing distributors.¹⁷⁶ Instead, Nantucket and Cadbury

¹⁶⁶ Day Distrib. Co. v. Nantucket Allserve, Inc., 2008 U.S. Dist. LEXIS 57334 (D. Minn. July 25, 2008).

¹⁶⁷ *Id.* at *1.

¹⁶⁸ Robert Halasz, *Company History: Nantucket Allserve, Inc.*, available at <http://www.answers.com/topic/nantucket-allserve-inc> (last visited Nov. 1, 2009).

¹⁶⁹ *Id.*

¹⁷⁰ Day Distrib. Co., 2008 U.S. Dist. at *2. Two of the distributors began as sub-distributors of Nantucket but later became distributors of Nantucket. *Id.*

¹⁷¹ *Id.* at *3.

¹⁷² Halasz, *supra* note 168.

¹⁷³ *Id.* In 1996, Nantucket was ranked 13th in *Inc.*'s fastest-growing U.S. private company list. *Id.*

¹⁷⁴ Corpfin Worldwide, Deal No. 171893 (June 19, 2009) (providing a summary of Snapple Beverage Group's acquisition of Nantucket Allserve Inc.). Notably, Cadbury bought Nantucket through Snapple Beverage Group, a wholly-owned subsidiary of Cadbury. *Id.*

¹⁷⁵ Cadbury Schweppes PLC: Nantucket Nectars Purchase Will Build Snapple Beverages, WALL ST. J., Mar. 26, 2002, at B.6.

¹⁷⁶ Day Distrib. Co. 2008 U.S. Dist. at *5. Two clarifying notes: Firstly, Cadbury Schweppes Americas Beverages, one of the defendants in the Nantucket case, was a subsidiary division of Cadbury in the United States. Secondly, the court referred to

decided to distribute Stewart's directly rather than to rely on distributors.¹⁷⁷ At this stage, the reputation mechanism collapsed and termination without good cause was likely to occur. Unsurprisingly, Nantucket took its distribution operations in-house and terminated the agreements with its distributors.¹⁷⁸ Apparently, Nantucket did not terminate the agreement due to the distributors' failure to comply with their contractual requirements.¹⁷⁹ Good cause supporting a termination was presumably absent.¹⁸⁰

CONCLUSION

This article questioned the effectiveness of the reputation mechanism in deterring opportunistic termination by franchisors. It also cast doubt upon the traditional law-and-economics assumption that reputation can be a substitute for franchise termination laws.

While this article focuses on franchise relationships, its conclusions can have broader implications. The assumption that reputation can serve as a substitute for regulation of long-term commercial relationships must be re-examined. As we have demonstrated, the concern for one's reputation in long-term commercial relations often has a limited life cycle. A firm in the early stages of its organizational life cycle usually lacks internal financial resources needed to achieve its business goals; therefore, it is spurred on to enter into contractual relationships with other firms that can assist it in overcoming its financial constraints.¹⁸¹ To illustrate, a financially immature firm, wishing to erect a factory, may have an initial incentive to enter into a partnership or a joint venture agreement with other firms, in order to recruit necessary capital. At this stage the reputation mechanism has some deterring effects, since the young firm is economically dependent upon its contractual partners. However, as the financial availability of the immature firm improves, it becomes less dependent on its contractual partners. The firm's

Nantucket Allserve and Cadbury Schweppes Americas Beverages as "Cadbury," following the practice of the parties. *Id.* at *1–2.

¹⁷⁷ *Id.* at *5.

¹⁷⁸ *Id.* at *1.

¹⁷⁹ It seems unlikely that all distributors concurrently failed to comply with Nantucket and Cadbury's requirements.

¹⁸⁰ For more cases which potentially reveal the collapse of the reputation mechanism in the context of franchise termination laws see *C. C. Hauff Hardware, Inc. v. Long Mfg. Co.*, 148 N.W.2d 425 (Iowa 1967); *Volvo Constr. Equip. N. Am., Inc. v. CLM Equip. Co.*, 386 F.3d 581 (4th Cir. 2004); *Sims Wholesale Co. v. Brown-Forman Corp.*, 468 S.E.2d 905 (Va. 1996); *Cooper Distrib. Co. v. Amana Refrigeration*, 63 F.3d 262 (3d Cir. 1995).

¹⁸¹ See *supra* Part II.A.

incentive will shift towards owning and operating the business on its own, which will increase its profits. At this inevitable stage, the effectiveness of reputation in preventing opportunism will decline and the reputation mechanism may ultimately collapse.

To Have and to Hold, for Richer or Richer: Prenuptial Agreements in the Comparative Context

Margaret Ryznar* and Anna Stepień-Sporek**

INTRODUCTION

The premarital agreement,¹ which permits prospective spouses to plan for divorce, may well be the world's most unromantic document.² Envisioning the end of a marriage not yet begun, prospective couples must divide property not yet acquired. They must select a legal framework governing their marriage and divorce. Lawyers are often invited to participate in the negotiations, fuelling prospective spouses in their demands. Unsurprisingly, therefore, many people prefer to avoid requesting a premarital agreement, despite judicial and social gains in the acceptance of such agreements.

However, odds do not favor lifelong marriages³ and when divorce ensues, many people resent their divorce settlements.⁴ Prenuptial agreements will therefore always have an important role in many engagements, particularly when one of the partners

* J.D., Notre Dame Law School; M.A., Jagiellonian University; B.A., University of Chicago. The authors would like to thank Meagan Tom and her colleagues on *Chapman Law Review* for phenomenal editorial assistance.

** Associate Professor of Law, University of Gdańsk School of Law, Poland. Doctor and Master of Law, University of Gdańsk.

¹ For a precise definition of the premarital agreement in United States law, also known as a prenuptial or antenuptial agreement, see *infra* Part I.A. In Europe, the term "marital agreement" is used to describe both premarital agreements and marital agreements, but not agreements made by spouses after a divorce. See *infra* Part II.

² Most unromantically, premarital agreements signal divorce to many prospective spouses. See, e.g., Darian M. Ibrahim, *The (Not So) Puzzling Behavior of Angel Investors*, 61 VAND. L. REV. 1405, 1441 (2008) (suggesting premarital agreements signal the possibility of divorce); Saul Levmore, *Norms as Supplements*, 86 VA. L. REV. 1989, 2021 (2000) (suggesting premarital agreements signal distrust); Lior Jacob Strahilevitz, *Reputation Nation: Law in an Era of Ubiquitous Personal Information*, 102 NW. U. L. REV. 1667, 1718 (2008) (identifying, as a classic example of signaling behavior, the lack of a prenuptial agreement as a means of signaling love prior to a marriage).

³ In 2003, there was one divorce for every two marriages. *National Vital Statistics Reports: Births, Marriages, Divorces, and Deaths: Provisional Data for 2003*, CENTERS FOR DISEASE CONTROL AND PREVENTION, (United States Department of Health and Human Services), June 10, 2004, available at http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_22.pdf.

⁴ See *infra* notes 67, 78, and 81 for examples of potential inequalities of bargaining power between spouses.

has noteworthy assets. For example, Hollywood actress Catherine Zeta-Jones reportedly contracted with actor Michael Douglas to receive \$2.8 million per year of marriage upon divorce, and if she proved his infidelity, for an additional \$5 million.⁵ Meanwhile, the premarital agreement between actress Nicole Kidman and singer Keith Urban would purportedly pay Urban about \$640,000 for every year that he spent with Kidman, unless he used illegal drugs during the marriage, in which case he would receive nothing.⁶ Finally, in light of the public speculation on his marital fidelity, golfer Tiger Woods reportedly amended his premarital agreement, contracting to pay his wife significantly more upon divorce if she refrained from filing for divorce in the immediate future.⁷

These illustrations underscore the inordinate power of premarital agreements in shifting wealth between spouses and discouraging undesirable marital behavior. They also symbolize, to people around the world, the typical use of the premarital agreement: to divide property upon divorce. The simplicity of this popular understanding, however, belies the complexity of premarital agreements. In essence, the premarital agreement permits a circumvention of the statutory defaults governing spouses' rights and responsibilities not only during divorce or death, but also during marriage. Furthermore, when legislation or case law alters these rights and responsibilities,⁸ premarital agreements protect spouses from being governed by the unexpected changes in the law.

⁵ Jae-Ha Kim, *Split Ends*, CHICAGO SUN-TIMES, Sept. 29, 2004, <http://jaehakim.com/articles/lifestyles/prenup.htm>.

⁶ *Ten Craziest Prenuptial Agreements*, MYWEDDING.COM, <http://www.mywedding.com/blogs/mywed/2008/02/ten-craziest-prenuptial-agreements.html> (last visited Oct. 14, 2009).

⁷ See, e.g., Bill Zwecker, *Woods, Wife In Intense Marriage Counseling*, CHICAGO SUN-TIMES, Dec. 2, 2009, <http://www.suntimes.com/entertainment/zwecker/1916410,zwecker-tiger-woods-marriage-elin-prenup-120209.article>.

⁸ One example of shifting divorce laws is found in England. See Margaret Ryznar, *All's Fair in Love and War: But What About in Divorce? The Fairness of Property Division in American and English Big Money Divorce Cases*, N.D. LAW REV. (forthcoming 2010). Nonetheless, premarital agreements are still rare in England; currently only 2% of married and divorced people in the United Kingdom have them. *Divorce Lawyers Braced for Busiest Week Ever*, TIMES ONLINE, January 5, 2009, <http://business.timesonline.co.uk/tol/business/law/article5450552.ece>. One reason for the rarity of such agreements might be their unenforceability, although courts often defer to them in determining appropriate ancillary relief. However, the enforceability of premarital agreements may soon be guaranteed in England, particularly if that is the recommendation of the influential Law Commission. See Brian Bix, *Bargaining in the Shadow of Love: The Enforcement of Premarital Agreements and How We Think About Marriage*, 40 WM AND MARY L. REV. 145, 150 n.18 (1998); *Pre-nuptial and Post-nuptial Agreements*, LAW COMMISSION, available at http://www.lawcom.gov.uk/marital_property.htm (last visited Dec. 21, 2009).

However, premarital agreements are not without their problems. Their enforceability in the United States is subject to procedural and substantive review.⁹ They also universally raise public policy issues with regard to the meaning of fairness and the limits on freedom of contract. Such issues heighten in the case of mobile couples, which include those who move both inter-state¹⁰ and internationally.¹¹ Given these issues, it is beneficial to consider the premarital agreement in the comparative context. This is particularly true as state courts and legislatures continue to encounter and address the unresolved issues surrounding premarital agreements.

Although England and the United States have similar understandings of such agreements,¹² the meaning and consequences of premarital agreements in continental Europe markedly differ from the Anglo-American common law tradition, heightening the opportunity for a comparative study. However, while the European approach inherently offers significant insight into the purpose, limits, and effects of premarital agreements, this approach is not as well-known—even to many Europeans—as the American approach, made so famous through Hollywood examples.

This Article therefore endeavors to consider and develop the notion of the premarital agreement in the comparative law context, addressing some of the universal issues surrounding premarital agreements, as well as the particular nuances of certain regulatory frameworks governing this type of agreement. Part I begins by exploring premarital agreements in American law, while Part II reviews the European approach to such agreements, focusing on Poland's representative approach, but also considering the approaches of France, Germany, and Switzerland. Part III draws lessons from a comparison of these various approaches, concluding that much of the distinction between American and European law on premarital agreements stems from the differing limits placed on the prospective spouses' freedom of contract. Part III also considers the desirable level of freedom of contract, the ideal characteristics of the regulatory framework of premarital agreements, and, finally, the popularity of such agreements.

⁹ See *infra* Part I.D.1 for further discussion.

¹⁰ See *infra* note 49.

¹¹ International dating and marriage has been facilitated by online dating and mail order bride programs. Premarital agreements in such cases can be particularly one-sided. See, e.g., *In re Marriage of Shirilla*, 89 P.3d 1, 2–4 (Mont. 2004).

¹² But see *supra* note 8 (noting the unenforceability of premarital agreements in England, which is different from the enforceability of such agreements in the United States).

I. PREMARITAL AGREEMENTS IN THE UNITED STATES

In the United States, family law has traditionally remained in the domain of the states.¹³ Therefore, American law regarding premarital agreements has developed independently in each state, whether by statute or case law. Even with the introduction of the Uniform Premarital Agreement Act (1983) (“UPAA”), the law on premarital agreements is far from uniform.

The premarital agreement in the United States has been rapidly developing since 1970, when courts began abandoning their public policy reasons against enforcing such agreements.¹⁴ *Posner v. Posner*¹⁵ became one of the first cases permitting the enforceability of premarital agreements in the 1970’s, while the UPAA prompted state legislatures to begin drafting statutes on the subject in the 1980’s.

Even today, premarital agreements are subject to certain procedural and substantive limits before a court will uphold their validity. Such agreements also raise important questions of fairness, which state law strives to resolve. Before turning to these questions, however, this Article reviews the brief history and current meaning of premarital agreements in the United States.

A. Definition of Premarital Agreement

At the outset, it is important to define the American premarital agreement, also known as a prenuptial or antenuptial agreement, because its meaning and consequences differ notably from those of the European marital agreement.¹⁶

The UPAA defines a premarital agreement as “an agreement between prospective spouses made in contemplation of marriage and to be effective upon marriage.”¹⁷ This definition, however, does not reflect the inordinate power of the premarital agreement, which permits prospective spouses to regulate their

¹³ See, e.g., *Moore v. Sims*, 442 U.S. 415, 435 (1979) (“Family relations are a traditional area of state concern.”).

¹⁴ However, the *notion* of a premarital agreement has a long history in both England and the United States. For a brief description of this history, see Sarah Ann Smith, *The Unique Agreements: Premarital and Marital Agreements, Their Impact Upon Estate Planning, and Proposed Solutions to Problems Arising at Death*, 28 IDAHO L. REV. 833, 840 (1992) and Judith T. Younger, *Lovers’ Contracts in the Courts: Forsaking the Minimum Decencies*, 13 WM. & MARY J. WOMEN & L. 349, 352–54 (2007).

¹⁵ 233 So. 2d 381 (Fla. 1970).

¹⁶ Continental Europe and the United States significantly differ in their approaches to premarital agreements, while England shares more similarities with the United States on the subject. See *infra* Part II and note 107. But see *supra* note 8.

¹⁷ UNIF. PREMARITAL AGREEMENT ACT § 1 (1983).

rights and responsibilities not only during divorce or death,¹⁸ but also during marriage.¹⁹

In the United States, spouses have significant freedom of contract when it comes to premarital agreements. Spouses may use the agreement to simply assign a piece of property, such as a house, to one of the spouses. Spouses may also completely opt out of the default property distribution regime of their state,²⁰ which would otherwise govern their property distribution upon divorce.

Specifically, each state has a default property distribution regime of either: 1) equitable distribution, which means a fair but not necessarily equal division between the spouses,²¹ or 2) community property, which often results in a roughly equal division of marital property between the spouses.²² Given their

¹⁸ "Since [the mid-nineteenth century] an arrangement in advance regarding each spouse's rights to the other's estate at death has been an acceptable subject for a premarital agreement." Smith, *supra* note 14, at 840. Estate planning should therefore carefully consider the property classifications created by a premarital agreement. See *id.* at 855.

¹⁹ Describing different types of prenuptial agreements:

First, a prenuptial agreement may shield wealth acquired by one spouse before marriage from the other. See, e.g., *Osborne v. Osborne*, 384 Mass. 591, 594, 428 N.E.2d 810, 813 (1981); *DeLorean v. DeLorean*, 211 N.J. Super. 432, 435, 511 A.2d 1257, 1259 (Ch. Div. 1986). Second, a prenuptial agreement may stipulate a division of property that is acquired during marriage. See, e.g., *Ferry v. Ferry*, 586 S.W.2d 782, 783 (Mo. Ct. App. 1979); *Gant v. Gant*, 329 S.E.2d 106, 109 & n.1 (W. Va. 1985). Third, the contract may predetermine the amount and timing of support one spouse will pay to the other after separation or divorce. See, e.g., *Lewis v. Lewis*, 69 Haw. 497, 499, 748 P.2d 1362, 1364 (1988); *Volid v. Volid*, 6 Ill. App. 3d 386, 387-88, 286 N.E.2d 42, 43-44 (1972). Finally, some commentators have advocated the use of prenuptial agreements to structure the terms of the ongoing relationship. See L. WEITZMAN, *THE MARRIAGE CONTRACT* 225-54 (1981); Shultz, *Contractual Ordering of Marriage: A New Model for State Policy*, 70 CALIF. L. REV. 204, 219-23 (1982).

Recent Developments, *Family Law—Prenuptial Agreements—Pennsylvania Supreme Court Rejects Substantive Review of Prenuptial Agreements—Simeone v. Simeone*, 581 A.2d 162 (Pa. 1990), 104 HARV. L. REV. 1399, 1399 n.3 (1991).

²⁰ See, e.g., Christine Davis, Note, *'Til Debt Do Us Part: Premarital Contracting Around Community Property Law—An Evaluation of Schlaefel v. Financial Management Service, Inc.*, 32 ARIZ. ST. L.J. 1051, 1055-57 (2000) (describing the effects on creditors of a prospective spouse's decision to opt out of a default community property regime through a premarital agreement); Deborah H. Bell, *Equitable Distribution: Implementing the Marital Partnership Theory Through the Dual Classification System*, 67 MISS. L.J. 115 (1997) (examining both default property distribution regimes in the United States).

²¹ For further discussion of the equitable distribution principle, see Ryznar, *supra* note 8.

²² In the community property regime, marriage is treated as a partnership in which property and debts acquired during the marriage belong to both spouses in equal, undivided shares. WILLIAM Q. DE FUNIAK & MICHAEL J. VAUGHN, *PRINCIPLES OF COMMUNITY PROPERTY* § 1 (2d ed. 1971). The community property approach to the distribution of property upon divorce is the default approach in only a minority of states, which currently includes Arizona, California, Idaho, Louisiana, Nevada, New Mexico, Texas, Washington, and Wisconsin. See generally Jeffrey G. Sherman, *Prenuptial*

contractual freedom, if prospective spouses reside in an equitable distribution state, they may contract for a community property division. If they reside in a community property state, they are free to write a premarital agreement that would keep their property separate.²³ Prospective spouses may also enter into a premarital agreement that changes the characterization of particular property that would otherwise be characterized as community under the state's default regime.²⁴

The characterization of property is especially important in terms of determining which property one spouse's creditors may collect against.²⁵ This is particularly true in community property states.²⁶ A debtor's marriage in an equitable distribution state has no impact on creditors, unless the debt is incurred to buy household necessities.²⁷ In a community property state, however, creditors' rights expand as a result of the debtor's marriage: debt may be collected from the spouse's resources brought into the marriage.²⁸

Premarital agreements or, if entered into after wedlock, matrimonial agreements may therefore impact how property is

Agreements: A New Reason to Revive an Old Rule, 53 CLEV. ST. L. REV. 359, 370 (2005–06). See also CAL. FAM. CODE §§ 2550–56 (West 2004).

²³ See, e.g., *Elia v. Pifer*, 977 P.2d. 796, 806 (Ariz. Ct. App. 1998) (“We therefore conclude that a valid premarital agreement abrogating community property rights precludes a creditor of one spouse from proceeding against the separate property of the other spouse on a claim arising during marriage.”); *Leasefirst v. Borrelli*, 17 Cal. Rptr. 2d 114, 116 (Cal. 1993) (holding that a “third-party creditor will not be entitled to recover against former community assets transmuted into separate property” by a premarital agreement (citing *Olean Tile Co. v. Schultze*, 169 Cal. App. 3d 359, 365–65 (Cal. 1985))). See also Smith, *supra* note 14, at 836.

²⁴ Andrea B. Carroll, *The Superior Position of the Creditor in the Community Property Regime: Has the Community Become a Mere Creditor Collection Device?*, 47 SANTA CLARA L. REV. 1, 43 (2007).

²⁵ Erik Paul Smith, Comment, *The Uncertainty of Community Property for the Tortious Liabilities of One of the Spouses: Where the Law Is Uncertain, There Is No Law*, 30 IDAHO L. REV. 799 (1994).

²⁶ For background on when and how creditors can reach community property to satisfy the debts of one of the spouses in community property states, see *id.* See also Smith, *supra* note 14 (analyzing the impact of premarital agreements under California law).

²⁷ “In many non-community property states, a nonearning spouse also may incur debts for which the earning spouse is liable. Under the doctrine of necessities, the earning spouse is responsible for payment of expenses incurred by the nonearning spouse for those things that are necessary for the family.” Susan Kalinka, *Taxation of Community Income: It Is Time for Congress to Override Poe v. Seaborn*, 58 LA. L. REV. 73, 94 (1997). “Necessity” is determined by examining factors such as the spouses’ means, social position, and circumstances. *Id.*

²⁸ Carroll, *supra* note 24, at 29. See also Lisa R. Mahle, *A Purse of Her Own: The Case Against Joint Bank Accounts*, 16 TEX. J. WOMEN & L. 45, 78–79 (2006) (“Since creditors can potentially garnish all community property in a joint account, in community property states when a creditor of one spouse wants to garnish a joint account, courts must first determine whether the money in the account is community property, separate property or joint tenancy property.”).

held during the marriage and its effect on third persons, such as creditors. This is particularly important in the nine community property states, and, to give notice to creditors, sometimes these agreements must be recorded in order to be binding on third parties.²⁹ Nonetheless, as the vast majority of states utilize equitable distribution as a default—thereby blocking a creditors' access to the property of the debtor's spouse—the premarital agreement in the United States typically has the most significance upon divorce when the agreement governs its terms.

Premarital agreements may be drafted to either significantly favor or disfavor the more vulnerable spouse upon divorce.³⁰ For example, a homemaker might include a provision that if his or her spouse is unfaithful, and therefore caused the divorce, that spouse must pay a significant portion of the assets.³¹ On the other hand, a significantly lower-income spouse might contract to keep his or her minimal financial marital contributions, leaving the other spouse with the bulk of the assets. Premarital agreements may also be drafted more neutrally toward both parties, so that each maintains some significant assets.

It is important, however, to distinguish the premarital agreement from the separation agreement, which permits already married spouses to contract the terms of their divorce. Cohabiting couples, meanwhile, may not enter into premarital agreements, which become effective only upon marriage.³²

Another important distinction is that between premarital and postmarital agreements—a distinction that does not clearly exist in the European countries.³³ Postmarital, also known as postnuptial, agreements are similar in substance and procedure to premarital agreements, except that they are signed after marriage. They are used to change provisions in the premarital agreement, or if not already covered by a premarital agreement, to make initial provisions, during the marriage, on the rights and responsibilities of the parties upon divorce or death. Therefore, the main difference between premarital and postmarital agreements is their timing in relation to the marriage.

²⁹ Carroll, *supra* note 24, at 32; *see also infra* Part III.B.

³⁰ “Since premarital contracting could be utilized by women to overcome gendered inequalities through marriage, summarily dismissing [premarital agreements] would deny women the possibility of using private ordering for empowerment and advancement.” Leah Guggenheimer, *A Modest Proposal: The Feminomics of Drafting Premarital Agreements*, 17 WOMEN’S RTS. L. REP. 147, 152 (1996).

³¹ *See, e.g., supra* note 5 and accompanying text.

³² For a survey of the law on property distribution following an unsuccessful cohabitation, *see* Ann Laquer Estin, *Ordinary Cohabitation*, 76 NOTRE DAME L. REV. 1381 (2001).

³³ *See infra* Part II.

The premarital agreement is thus an important type of contract with the power to govern a marriage and potential divorce.³⁴ The terms, meaning, and consequences of such agreements in the United States have further been clarified through judicial and legislative law.

B. A Brief History of the Premarital Agreement

Among the most important milestones in the evolution of the American premarital agreement are *Posner v. Posner*,³⁵ the first notable judicial recognition of the enforcement of premarital agreements, and the UPAA,³⁶ an influential draft of statutory law on the subject.

1. Case Law

Posner v. Posner is often cited as the first case upholding the validity of premarital agreements, making Florida the first state to recognize such agreements.³⁷ In its opinion, the *Posner* court noted the artificial distinction in other states' case law that skirted the issue of the validity of prenuptial agreements, but permitted spouses to contract their own property settlements under narrow circumstances.³⁸ The court also took "judicial notice of the fact that the ratio of marriages to divorces has reached a disturbing rate in many states."³⁹ Therefore, the court concluded that premarital agreements may be upheld under certain conditions.

The *Posner* court also noted the differing viewpoints of the appellate judges it overruled in the case, whose views summarized the predominant stances on the premarital agreement at the time. These views included that 1) the trial court need not be bound by premarital agreements, though they are permissible, 2) premarital agreements may be void on public policy grounds, and 3) premarital agreements may be as binding on the trial court as an agreement settling one spouse's property

³⁴ This power of the premarital agreements might be used to achieve fairer results between the spouses following divorce. Currently, many women struggle to keep their households running on a reduced income after a divorce. In 1993, the mean income for divorced American mothers was \$17,859, while for divorced fathers it was \$31,034. Arthur B. LaFrance, *Child Custody and Relocation: A Constitutional Perspective*, 34 U. LOUISVILLE J. FAM. L. 1, 6 (1995). But see generally Kelly Bedard & Olivier Deschênes, *Sex Preferences, Marital Dissolution, and the Economic Status of Women*, 40 THE JOURNAL OF HUMAN RESOURCES 411 (2005) (arguing that divorced women live in households with more income per person than never-divorced women).

³⁵ 233 So. 2d 381, 384 (Fla. 1970).

³⁶ UNIF. PREMARITAL AGREEMENT ACT (1983).

³⁷ 233 So. 2d at 384.

³⁸ *Id.* at 383.

³⁹ *Id.* at 384.

rights upon the death of the other spouse.⁴⁰ In overruling the lower court, *Posner* marked the shift from the judicial practice of voiding premarital agreements to a policy that recognized premarital agreements as binding.

Although most states acknowledged the enforceability of premarital agreements soon after *Posner*,⁴¹ state courts continued to play a significant role in defining the appropriate parameters of premarital contracts.⁴² For example, the Supreme Court of Ohio outlined procedural safeguards in *Gross v. Gross*,⁴³ while a Massachusetts court recently found that pregnancy does not negate a bride's free will to enter into a premarital agreement.⁴⁴ In Rhode Island, both parties need not have counsel in order for a premarital agreement to be valid.⁴⁵ In New Jersey, the doctrine of equitable estoppel cannot be used to validate an otherwise invalid agreement.⁴⁶

However, states have become clearly divided on certain issues, such as the formalities that must attend such agreements⁴⁷ and whether parties could contract on issues such as spousal support.⁴⁸ These interstate inconsistencies were most problematic for mobile couples.⁴⁹ The UPAA, considered next, aimed to remedy some of these inconsistencies.

⁴⁰ *Id.* at 382.

⁴¹ *See, e.g.,* *Simeone v. Simeone*, 581 A.2d 162, 165 (Pa. 1990) (finding that a less deferential approach to the enforcement of premarital agreements would entail "[p]aternalistic presumptions and protections [that sheltered] women from the inferiorities and incapacities which they were perceived as having").

⁴² *See, e.g.,* *Bakos v. Bakos*, 950 So. 2d 1257, 1260 (Fla. Dist. Ct. App. 2007) (deciding that a premarital agreement signed the day before a wedding was voidable, but aggrieved party may ratify it); *Chubbuck v. Lake*, 635 S.E.2d 764 (Ga. 2006) (finding that a premarital agreement was void and unenforceable when it failed to meet the statutory requirement that it be witnessed by two people); *Seherr-Thoss v. Seherr-Thoss*, 141 P.3d 705, 712 (Wyo. 2006) (determining that the laws governing the enforceability of contracts also govern premarital agreements).

⁴³ 464 N.E.2d 500, 506 (Ohio 1984). The *Gross* court stated that:

[Premarital] agreements are valid and enforceable . . . (1) if they have been entered into freely without fraud, duress, coercion, or overreaching; (2) if there was full disclosure, or full knowledge and understanding of the nature, value and extent of the prospective spouse's property; and (3) if the terms do not promote or encourage divorce or profiteering by divorce.

Id.

⁴⁴ *Biliouris v. Biliouris*, 852 N.E.2d 687, 693 (Mass. App. Ct. 2006).

⁴⁵ *Marsocci v. Marsocci*, 911 A.2d 690, 698 (R.I. 2006).

⁴⁶ *In re Estate of Shinn*, 925 A.2d 88, 96 (N.J. Super. Ct. App. Div. 2007).

⁴⁷ UNIF. PREMARITAL AGREEMENT ACT, § 2 cmt (1983).

⁴⁸ *Id.*

⁴⁹ Such movement often triggers conflict-of-law issues. The UPAA, which permits prospective spouses to select "the choice of law governing the construction of the agreement," was specifically drafted to address "[t]he problems . . . exacerbated by the mobility of our population." UNIF. PREMARITAL AGREEMENT ACT Prefatory Note, § 3 cmt (1983). *See also infra* note 51.

2. Statutory Law

The National Conference of Commissioners on Uniform State Laws promulgated the UPAA in 1983 to provide a uniform law on premarital agreements. Approximately half of the states have now adopted some variation of the UPAA.⁵⁰

One of the most important characteristics of the UPAA is its strong support of the freedom of contract. Section 3 of the Act lists several topics a premarital agreement may cover, including property rights, spousal support, and the choice of law governing the agreement.⁵¹ Significantly, this list is not exhaustive and parties may contract on any topic not in violation of either a public policy principle or a criminal statute.⁵² The only topic explicitly forbidden from premarital contracting is child support that adversely affects the child, although as a general rule many child-related provisions are typically considered to be against public policy.⁵³

Enforcement of premarital agreements is considered in section 6 of the UPAA.⁵⁴ This section provides that a premarital agreement is unenforceable against a spouse who did not execute the agreement voluntarily.⁵⁵ The premarital agreement is also unenforceable if it was unconscionable when executed and the spouse: (1) was not provided fair disclosure of the other spouse's financial details; (2) did not waive the right to receive such disclosure; and, (3) did not have adequate knowledge of those financial details.⁵⁶ Therefore, a person with knowledge of a spouse's financial status or reason to know of it, coupled with voluntary execution, cannot contest the premarital agreement.⁵⁷

⁵⁰ Charles W. Willey, *Effect in Montana of Community-Source Property Acquired in Another State (And Its Impact on a Montana Marriage Dissolution, Estate Planning, Property Transfers, and Probate)*, 69 MONT. L. REV. 313, 365 (2008). For examples of various states' UPAA-based laws, see CAL. FAM. CODE § 1615 (West 2004); 750 ILL. COMP. STAT. 10/1 *et seq.* (West 1993); and R.I. GEN. LAWS § 15-17-6 (2003).

⁵¹ UNIF. PREMARITAL AGREEMENT ACT § 3 (1983). *See also* Montoya v. Montoya, 909 A.2d 947 (Conn. 2006); Bradley v. Bradley, 164 P.3d 537 (Wyo. 2007). For an excellent discussion of choice of law issues in premarital agreements, see Julia Halloran McLaughlin, *Premarital Agreements and Choice of Law: "One, Two, Three, Baby, You and Me,"* 72 MO. L. REV. 793 (2007).

⁵² UNIF. PREMARITAL AGREEMENT ACT § 3 cmt (1983).

⁵³ *Id.*

⁵⁴ UNIF. PREMARITAL AGREEMENT ACT § 6 (1983). For constructive criticism of the UPAA's enforcement provision, see Barbara Ann Atwood, *Ten Years Later: Lingering Concerns About the Uniform Premarital Agreement Act*, 19 J. LEGIS. 127 (1993). Meanwhile, section 7 of the UPAA governs enforceability of agreements in marriages that were subsequently determined to be void. *See* UNIF. PREMARITAL AGREEMENT ACT § 7 (1983).

⁵⁵ UNIF. PREMARITAL AGREEMENT ACT at § 6.

⁵⁶ *Id.*

⁵⁷ *Id.* at § 6 cmt.

Similarly, a person who waived knowledge of these financial details and voluntarily executed the agreement is bound by it.⁵⁸

Although the UPAA is a source of some guidance, the law on premarital agreements remains in the realm of the states and any generalization is therefore difficult. Nonetheless, it is a fair observation that American states typically provide prospective spouses significant contractual freedom and premarital agreements are enforceable unless they fail judicial review.⁵⁹

C. Theoretical Underpinnings

The modern premarital agreement is rooted in contract law theories.⁶⁰ Parties to a premarital agreement, viewed as independent negotiators,⁶¹ have almost full discretion over the contents and scope of their agreement, enabling them to dictate the terms of their divorce absent any enforceability issues.⁶² This is particularly important for community property states such as California, where parties may waive their rights to share property.⁶³

Many commentators have noted that marriage itself has evolved from a relationship based on status to one regulated by contract.⁶⁴ This shift from marriage as regulated by the state to marriage as determined by the private ordering between parties has been called “the privatization of family law.”⁶⁵

⁵⁸ *Id.* There is an additional provision that bars enforcement of a premarital agreement to the extent that it would force the lower income spouse onto welfare. *Id.* at § 6(b).

⁵⁹ See *infra* Part I.D.1.

⁶⁰ “[A] premarital agreement is a contract. As required for any other contract, the parties must have the capacity to contract in order to enter into a binding agreement.” UNIF. PREMARITAL AGREEMENT ACT § 2 cmt (1983). The UPAA also draws upon contract and commercial law for the standard of unconscionability. *Id.* at § 6 cmt.

⁶¹ See Elizabeth S. Scott & Robert E. Scott, *A Contract Theory of Marriage*, in *THE FALL AND RISE OF FREEDOM OF CONTRACT* 201 (F.H. Buckley ed., 1999) [hereinafter *FALL AND RISE OF CONTRACT*].

⁶² See *infra* Part I.D.1.

⁶³ See *supra* Part I.A.

⁶⁴ Scott & Scott, *supra* note 61, at 201. Marriage as status means, in essence, that family law automatically bestows a set of rights and obligations upon people who are marrying, which can be altered only by divorce, not by mutual agreement. For an excellent discussion of the interplay between marriage as status and marriage as contract, see Lisa Milot, Note, *Restitching the American Marital Quilt: Untangling Marriage from the Nuclear Family*, 87 VA. L. REV. 701 (2001); Cynthia Starnes, *Divorce and the Displaced Homemaker: A Discourse on Playing with Dolls, Partnership Buyouts and Dissociation under No-Fault*, 60 U. CHI. L. REV. 67 (1993). For a description of freedom of contract in covenant marriage, which is a more binding form of marriage available in a few states, see Margaret F. Brinig, *Contracting Around No-Fault Divorce*, in *FALL AND RISE OF CONTRACT*, *supra* note 61, at 275.

⁶⁵ Scott & Scott, *supra* note 61, at 203. It is important to note here, however, that there is a major distinction between (1) marital behavior being governed by contract and (2) divorce being governed by contract. See, e.g., Marsha Garrison, *Marriage: The Status*

However, there are obvious distinctions between contracts and premarital agreements, casting doubt on whether contract law provides an appropriate framework for premarital agreements.⁶⁶ Most problematically, the bargaining process in the marital context is not at arm's length, but "may be afflicted by unreflective love, even infatuation."⁶⁷ Additionally, the characteristics of marriage, so dependent on life circumstances and children, are sufficiently unique to prevent the blind application of pure contract principles.⁶⁸

Nonetheless, premarital agreements are often defended on partnership principles as well.⁶⁹ There are inconsistencies, however, in the notion that premarital agreements inherently advance the prospective spouses' equality. Specifically, a court's ability to invalidate a premarital agreement suggests that one of the partners is too weak to contract.⁷⁰ Conversely, if a court upholds skewed premarital agreements, then spouses may bargain for unequal treatment. It has been suggested that premarital agreements must move in the direction of dividing property equally, or else they are at odds with the view of marriage as a partnership.⁷¹ This proposition, however, would defeat the entire purpose of a premarital agreement, which is to provide parties a method of contracting around defaults. Therefore, it is not entirely clear whether premarital agreements enhance or undermine the idea of marriage as an equal partnership. They may certainly be used by parties to effectively do either, depending on the terms of the agreement.

of Contract, 131 U. PA. L. REV. 1039 (1983) (reviewing LENORE J. WEITZMAN, *THE MARRIAGE CONTRACT* (1981)).

⁶⁶ Eric A. Posner, *Family Law and Social Norms*, in FALL AND RISE OF CONTRACT, *supra* note 61, at 256.

⁶⁷ Michael J. Trebilcock, *Marriage as a Signal*, in FALL AND RISE OF CONTRACT, *supra* note 61, at 254. For the argument that premarital contracting creates greater equality of bargaining power than either intramarital or postmarital bargaining, see generally *id.*

⁶⁸ *Developments in the Law—The Law of Marriage and Family*, 116 HARV. L. REV. 1996, 2096–97 (2003).

⁶⁹ See, e.g., UNIF. PREMARITAL AGREEMENT ACT § 6 cmt (1983), which prioritizes protecting spouses "against overreaching, concealment of assets, and sharp dealing not consistent with the obligations of marital partners to deal fairly with each other." (emphasis added).

⁷⁰ This recalls the days when women could not legally contract. See, e.g., *Poole v. Perkins*, 101 S.E. 240 (Va. 1919).

⁷¹ *Developments in the Law—The Law of Marriage and Family*, *supra* note 68, at 2096:

[D]eference to freedom of contract in antenuptial agreement law is undesirable. . . . [A]cknowledgment of the partnership conception of marriage demands that parties desiring to execute antenuptial agreements approximate the fifty-fifty division of property implicit in the partnership approach or stand prepared to prove the agreements' substantive fairness at the time of divorce.

Although both the partnership and contractual frameworks have flaws when applied to the marital context, they have underpinned and legitimized premarital agreements. As a result, couples have benefited from the opportunity to contractually circumvent judicial and statutory defaults in the case of divorce.

D. Enforceability

Section six of the UPAA,⁷² or the corresponding state statute, governs the enforceability of premarital agreements. However, it is the courts that are the ultimate arbiters of whether a particular premarital agreement governs the terms of divorce.⁷³ The issue of enforceability arises most frequently following a court's procedural and substantive review of a premarital agreement.⁷⁴

1. Judicial Review

In order to be upheld by the court, a premarital agreement must survive substantive and procedural review.⁷⁵ Occasionally, these separate inquiries are blurred.⁷⁶ In other words, if the substance of the agreement appears fair to the court, defects in the bargaining process may be of lesser importance. However, if the agreement seems particularly unfair to one spouse, courts may examine the procedures surrounding its execution more closely.⁷⁷

In terms of substantive review,⁷⁸ some courts have departed from the standard unconscionability doctrine by which

⁷² UNIF. PREMARITAL AGREEMENT ACT § 6 (1983). *See supra* Part I.B.2.

⁷³ *See* UNIF. PREMARITAL AGREEMENT ACT § 6(c) (1983) ("An issue of unconscionability of a premarital agreement shall be decided by the court as a matter of law.").

⁷⁴ Some commentators have warned that enforcement of premarital agreements must be done carefully so as to not disadvantage women. *See, e.g.,* Atwood, *supra* note 54, at 129.

⁷⁵ For the argument that courts rarely invalidate premarital agreements following procedural and substantive review, see Younger, *supra* note 14, at 358–59, 422–23. One commentator suggests that the procedural and substantive fairness protections placed on premarital agreements reflect each state's view of the appropriate balance between individual autonomy and state oversight of premarital agreements. McLaughlin, *supra* note 51, at 853.

⁷⁶ Younger, *supra* note 14, at 356–57.

⁷⁷ *Id.*

⁷⁸ *See* Karen Servidea, Note, *Reviewing Premarital Agreements to Protect the State's Interests in Marriage*, 91 VA. L. REV. 535, 540–41 (2005). Some critics argue that a substantive review of premarital agreements is paternalistic. *See* Simeone v. Simeone, 581 A.2d 162, 165 (Pa. 1990).

commercial contracts are evaluated.⁷⁹ Instead, judges often examine the fairness of the premarital agreement at the time of divorce.⁸⁰ Furthermore, certain topics fall outside the scope of permissible contracting for public policy reasons,⁸¹ including a child's religion,⁸² child custody, child care payments, or parental visitation.

Meanwhile, the test for procedural fairness focuses on the parties' conduct in obtaining the premarital agreement.⁸³ First, each party must have voluntarily entered into the agreement, absent fraud, overreaching, sharp dealing,⁸⁴ or duress.⁸⁵ Additionally, at the time the parties entered into the agreement, disclosure of each party's financial status is required.⁸⁶ Significant departure from these accepted procedural practices provides courts with the opportunity to circumvent premarital agreements in resolving the issues arising during divorce proceedings.

⁷⁹ *But see supra* notes 60 and 73; *Lane v. Lane*, 202 S.W.3d 577 (Ky. 2006) (finding that although public policy does not render antenuptial agreements per se invalid, such agreements may be analyzed by courts as unconscionable).

⁸⁰ *Servidea, supra* note 78, at 540–41.

⁸¹ I. Glenn Cohen, *The Right Not to be a Genetic Parent?*, 81 S. CAL. L. REV. 1115, 1169 (2008). *See also* Eric A. Posner, *Family Law and Social Norms*, in *FALL AND RISE OF CONTRACT*, *supra* note 61, at 256 (“Aside from the restrictions on termination provisions in prenuptial agreements, potential mates cannot bind themselves legally to marriages in which spouses’ domestic, financial, and sharing obligations are specified by contract. Polygamous and same-sex marriages are prohibited. These laws are not default rules, but restrictions on freedom of marital contract.”). For the argument that courts should lift restrictions on marital contracting to obtain less paternalistic and more efficient results, see Milot, *supra* note 64. Other scholars have similarly argued for even greater contractual freedom. *See, e.g.*, Joan M. Krauskopf & Rhonda C. Thomas, *Partnership Marriage: The Solution to an Ineffective and Inequitable Law of Support*, 35 OHIO ST. L.J. 558 (1974); Marjorie Maguire Shultz, *Contractual Ordering of Marriage: A New Model for State Policy*, 70 CAL. L. REV. 204 (1982). *But see* Sally Burnett Sharp, *Fairness Standards and Separation Agreements: A Word of Caution on Contractual Freedom*, 132 U. PA. L. REV. 1399 (1984) (arguing that absolute freedom of contract may hinder fair results upon divorce).

⁸² *See* *Zummo v. Zummo*, 574 A.2d 1130 (Pa. Super. Ct. 1990). *See also* Alexandra Selfridge, *Challenges for Negotiating and Drafting an Antenuptial Agreement for the Religious Upbringing of Future Children*, 16 J. CONTEMP. LEGAL ISSUES 91 (2007).

⁸³ Younger, *supra* note 14, at 357.

⁸⁴ *Id.*

⁸⁵ *See, e.g.*, *In re Estate of Hollett*, 834 A.2d 348 (N.H. 2003).

⁸⁶ *See, e.g.*, *Blige v. Blige*, 656 S.E.2d 822 (Ga. 2008) (holding that both parties entering into an antenuptial agreement must provide a full and fair disclosure of all material facts); *Friezo v. Friezo*, 914 A.2d 533 (Conn. 2007) (determining that disclosure requirements for a premarital agreement are satisfied when the parties disclose a general approximation of their income, assets, and liabilities).

2. Formulaic Premarital Agreements

The question of enforceability plagues not only procedurally and substantively complicated agreements,⁸⁷ but also simple ones. Any internet search reveals premarital agreement packages with boilerplate language, allowing the prospective spouses to sign formulaic contracts without spending money on attorneys' fees.⁸⁸

Ultimately, form premarital agreements are not inherently more or less enforceable than those drafted by lawyers.⁸⁹ They are subject to the same procedural and substantive limitations as any other premarital agreement, thus representing a reliable option for low cost divorce planning.

There are certainly advantages and disadvantages to an increased use of such unsophisticated premarital agreements. On the one hand, without lawyers, prospective spouses may not know the depth and scope of potential negotiations, lessening their bargaining power. On the other hand, inexpensive premarital agreements allow spouses of modest means to plan for a potential divorce.⁹⁰ Furthermore, because one side does not outmatch the other in legal power, perhaps the spouses achieve greater equality in such negotiations.

In many ways, the formulaic premarital agreement parallels the holographic will. Similar public policy reasons permit both, centering on the autonomy of the individual to dispose of his or her own property. Furthermore, the do-it-yourself premarital

⁸⁷ For a good review of the complex financial issues that might be considered by a premarital agreement, see David M. Johnson, *Complex Financial Issues in Family Law Cases*, 37 COLO. LAW. 53 (2008).

⁸⁸ For examples of companies that sell asset protection in marriage, see <http://prenuptialagreementform.com> (last visited Oct. 11, 2009) (form premarital agreements for as low as \$29.99) and <http://www.legalformsbank.biz/premarital.asp> (last visited Oct. 11, 2009) (same for \$9.95). See also Andrew Blair-Stanek, Comment, *Defaults and Choices in the Marriage Contract: How to Increase Autonomy, Encourage Discussion, and Circumvent Constitutional Constraints*, 24 Touro L. REV. 31, 68 n.166 (2008).

⁸⁹ See, e.g., UNI. PREMARITAL AGREEMENT ACT, § 6 cmt. (1983):

Nothing in Section 6 [regarding enforcement] makes the absence of assistance of independent legal counsel a condition for the unenforceability of a premarital agreement. However, lack of that assistance may well be a factor in determining whether the conditions stated in Section 6 may have existed (see, e.g., *Del Vecchio v. Del Vecchio*, 143 So.2d 17 (Fla. 1962)).

This UPAA provision would most likely be applicable to cases where only one of the parties was assisted by independent legal counsel.

⁹⁰ Somewhat counter-intuitively, those of modest means may need premarital agreements the most because marriages are particularly vulnerable to dissolution when spouses encounter financial trouble. See, e.g., HENRY J. SOMMER ET. AL., *COLLIER FAMILY LAW AND THE BANKRUPTCY CODE*, ¶6.05[5], at xiii (2009) ("In some parts of the country as many as half of all marriages end in divorce, often due, at least in part, to financial difficulties.").

agreement symbolizes the preference of the American philosophy for nearly complete freedom of contracting, permitting prospective spouses to enter into an agreement uninfluenced by judicial or legislative preferences.⁹¹ This cornerstone of American philosophy, so favorable to freedom of contract, in fact drives many of the distinctions between American and European law on the subject of premarital agreements.

In sum, Americans may utilize the premarital agreement to avoid judicial and statutory defaults in their states, enjoying significant freedom of contract. The force of the premarital agreement has been effectively developed over the past several decades, resulting in an important role for premarital agreements in many American couples' engagements. Interestingly, however, premarital agreements have a longer history in many European countries, acquiring significantly different consequences and meaning.

II. EUROPEAN LAW ON MARITAL AGREEMENTS AS EXEMPLIFIED BY POLAND

Europe generally does not share the distinction that exists in the United States between premarital and postmarital agreements. Instead, both types of agreements are treated as one type of contract: the marital agreement. The marital agreement may be concluded either before the marriage ceremony or during the marriage.⁹²

This Part analyzes the approaches to marital agreements in Europe, focusing on France, Germany, Poland, and Switzerland.⁹³ The Polish approach is given more in-depth treatment, not only because it is representative of the others, but also because it illustrates notable differences from the American approach.

These European legal systems have several commonalities that are worth mentioning at the outset. Under the French Civil Code,⁹⁴ the German Bürgerliches Gesetzbuch ("BGB"),⁹⁵ the

⁹¹ *But see* sources cited *supra* notes 67 and 73.

⁹² *But see* C. CIV. 1394–97 (France).

⁹³ "Comparisons among [England, France, Germany, and the United States] continue to seem fruitful, not only because of the great influence their legal systems exert in the civil and common law worlds, but also because each has generated a rich assortment of legal and social science materials." MARY ANN GLENDON, *THE TRANSFORMATION OF FAMILY LAW: STATE, LAW, AND FAMILY IN THE UNITED STATES AND WESTERN EUROPE* 3 (1989).

⁹⁴ C. CIV. (France).

⁹⁵ Bürgerliches Gesetzbuch [BGB] [Civil Code] Aug. 18, 1896, Reichsgesetzblatt [RGBL] 195, as amended, §§ 1426, 1432, 1437, 1519, 1549.

Swiss Schweizerisches Zivilgesetzbuch (“ZGB”),⁹⁶ and the Polish Kodeks Rodzinny i Opiekuńczy (“KRO”),⁹⁷ there are a few optional systems of matrimonial property law aside from the statutory regime governing the marriage. As a general rule, spouses may modify the standard statutory regime that would apply to their matrimonial relations by means of a marital agreement.⁹⁸ Importantly, spouses are not obliged to choose any specific contractual system and can avoid at least some of the consequences of the standard statutory regime.⁹⁹ Nonetheless, spouses do not have entirely unrestricted autonomy with regard to their matrimonial property law in any of the countries considered, nor is entering into a marital agreement popular.¹⁰⁰

A. Various European Countries’ Approaches

It should be mentioned that prospective spouses in many European countries resist marital agreements altogether because they think that such documents are only important upon divorce. Therefore, spouses often do not enter into such agreements because they want to underscore that they are not going to divorce.¹⁰¹ Such an opinion of marital agreements derives from American movies and news regarding the divorces of celebrities. Very rarely, however, is the situation of these divorcing celebrities analyzed within a larger context and within the legal circumstances that are specific to the United States. This is understandable because the main aim of movies is not to teach.

Nonetheless, the result is that the opinion of marital agreements in Europe is built upon the false conviction that marital agreements in Europe have the same consequences as those in the United States, when in reality the meaning and consequences of such agreements in Europe differ from those in the United States. In Europe, the regime choice made by the spouses in a marital agreement mainly impacts how property is held during the marriage and which property is available to the creditors of one spouse.¹⁰² While the rules of distribution of property upon divorce are also determined by the chosen regime,

⁹⁶ Schweizerisches Zivilgesetzbuch [ZGB] [Civil Code] Dec. 10, 1907, SR 210, RS 210, art. 177–247 (Switz.).

⁹⁷ Kodeks rodzinny i opiekuńczy – The Family and Guardianship Code of 25th February 1964 - Dz. U. nr 9, poz. 59 with amendments [hereinafter “KRO”].

⁹⁸ See *infra* Parts II.A. & II.B. for further discussion.

⁹⁹ *Id.*

¹⁰⁰ In France, for example, only 10% of spouses decide to enter into a marital agreement, and they do so mostly when important assets or second marriages are involved. See CAROLYN HAMILTON & ALISON PERRY, FAMILY LAW IN EUROPE 261 (2d ed. 2002).

¹⁰¹ This might also be a signaling problem. See *supra* note 2.

¹⁰² See, e.g., *infra* Part II.A.1.

spouses may modify them only in a very narrow way through the marital agreement.¹⁰³

Therefore, among the most important points to initially consider is that the marital agreement has a different meaning in Europe than in the United States, and differing meanings within Europe as well.¹⁰⁴ In certain European countries, all of the contracts between spouses are called marital agreements. The permissible scope of the marital agreement also differs from European country to country. In some countries, a marital agreement concerns only the relations between spouses, while in others, the agreement may regulate the consequences of a spouse's death.¹⁰⁵ Furthermore, in some countries, there are only a few models of property regimes from which spouses may choose. In other countries, spouses are not obliged to follow the statutory models of the regimes and have more freedom with regard to the content of their marital agreements.¹⁰⁶

Nonetheless, two fundamental approaches to marital agreements can be distinguished in Europe.¹⁰⁷ According to the first approach, the marital agreement is a kind of general agreement, constructing the rules of the classification of property and the relations of the spouses, but on the other hand, not regarding any particular property.¹⁰⁸ The second approach is based on the rule that each contract between spouses is a marital agreement, even if it concerns only certain chattels belonging to one spouse.¹⁰⁹

1. France

The meaning of the term "marital agreement" is quite broad in France. It covers not only the agreements in which spouses choose their matrimonial regime, but also the contracts

¹⁰³ See, e.g., *infra* Part II.A.3.

¹⁰⁴ This is a very important point given the current process of European unification and harmonization in the field of family law. See, e.g., Esin Örüçü, *A Family Law for Europe: Necessary, Feasible, Desirable?*, in PERSPECTIVES FOR THE UNIFICATION AND HARMONISATION OF FAMILY LAW IN EUROPE 551 (Katharina Boele-Woelki ed., Intersentia 2003) [hereinafter PERSPECTIVES].

¹⁰⁵ See Allgemeines bürgerliches Gesetzbuch [ABGB] [Civil Code] Justizgesetzsammlung [JGS] No. 946/1816, as amended § 1217 (Austria).

¹⁰⁶ Austria and France serve as examples of such countries. See HAMILTON & PERRY, *supra* note 100, at 12, 260–61.

¹⁰⁷ These remarks concern only continental Europe because the English approach to marital agreements is similar to the American one. See, e.g., Nigel Lowe & Roger Kay, *The Status of Prenuptial Agreements in English Law – Eccentricity or Sensible Pragmatism?*, in FAMILY FINANCES 395–413 (Bea Verschraegen ed., 2009) [hereinafter FAMILY FINANCES]. But see *supra* note 8 (noting the lack of enforceability of premarital agreements in England).

¹⁰⁸ See Lowe & Kay, *supra* note 107, at 395–413.

¹⁰⁹ *Id.*

regarding every chattel belonging to at least one spouse.¹¹⁰ Spouses are free to regulate the rules concerning the management of their property and are entitled, by any kind of agreement, to modify the default statutory community property regime.¹¹¹

There are a few models of property regimes known by the Civil Code in France: separation of property,¹¹² separation in acquisition,¹¹³ universal community,¹¹⁴ and community of movables and acquisitions.¹¹⁵ Spouses may choose among these in their marital agreements, but are prohibited from electing former types of marital regimes, such as, for example, the dotal system.¹¹⁶

If the spouses were married without any provision concerning matrimonial property law, the default statutory system of a limited community property is applied.¹¹⁷ In this system, only property acquired during the marriage is held in common, although gifts and inheritances acquired during the marriage are the separate property of each spouse.¹¹⁸ Community property belongs to both spouses jointly, although each spouse is able to make ordinary acts of administration of community property. However, important transactions relating to this kind of property need the consent of both spouses.¹¹⁹

Importantly, when the record of marriage mentions that a marital agreement has not been made, third parties may assume

¹¹⁰ WALTER CAIRNS & ROBERT MCKEON, INTRODUCTION TO FRENCH LAW 69–70 (1995).

¹¹¹ See C. CIV. 1497 (France).

¹¹² See CAIRNS & MCKEON, *supra* note 110, at 71. In this regime, spouses hold property separately. *Id.*

¹¹³ In this regime, spouses behave as if they were married under the regime of separation of property. At the dissolution of the regime, each spouse receives half of the value of the net acquisitions belonging to the property of the other spouse. *Id.*

¹¹⁴ In this regime, “the community includes not only the acquets and gains of the marriage, but also any property brought into the marriage by either spouse.” Carroll, *supra* note 24, at 27. However, property that is separate by its nature—property delineated as such in C. CIV. 1404 (France)—does not fall into community property, unless otherwise stipulated.

¹¹⁵ This system is similar to the limited community property system. The key characteristic of this system is that immovables are not part of the common property of spouses.

¹¹⁶ The dotal system was in force during medieval times and in some areas later. It was also recognized by Roman law. In this system, the wife was typically given with a dower, which was administrated by her husband during the marriage and was given back to her upon the death of her husband. Being a former property regime, however, it is not currently mentioned by the Civil Code and therefore may not be chosen by spouses today. See CAIRNS & MCKEON, *supra* note 110, at 71.

¹¹⁷ Some authors translate the name of this regime as the “community of ownership of matrimonial property.” See HAMILTON & PERRY, *supra* note 100, at 260.

¹¹⁸ C. CIV. 1402 (France).

¹¹⁹ C. CIV. 1421 (France).

that spouses have been married under the default statutory regime of limited community property.¹²⁰ However, this rule is not applied if the spouses have declared, in the transaction entered into with a third party, that they had made a marital agreement.¹²¹

The freedom of contract in the field of marital agreements is very well-developed in French law, especially when compared to the other European countries.¹²² In their marital agreements, spouses in France may choose one of the property regimes mentioned by the Civil Code, but may also modify the rules of these regimes: spouses may mix different regimes and are even able to establish new regimes that are not recognized by the law.¹²³ Unlike the other European countries examined here, spouses in France are also able to make provisions for what should happen upon their death.¹²⁴ Spouses may also opt for universal community, which is not popular in Europe.¹²⁵

However, there are some limitations on this freedom of contract. For example, the marital agreement cannot be against public order (public morals).¹²⁶ Furthermore, spouses are not able to modify the rules of the so-called primary regime (*régime primaire*)¹²⁷ and the statutory order of successions.¹²⁸ Finally, spouses are not permitted to derogate from the rules regarding parental authority or guardianship,¹²⁹ nor may they derogate from the duties and rights which result from marriage.¹³⁰

In France, marital agreements should be concluded before wedlock. During the marriage, the marital agreement is immutable (*principe d'immutabilité*).¹³¹ However there is an

¹²⁰ The record of marriage is a certificate given to the spouses and serves as proof that they are married.

¹²¹ C. CIV. 1394 (France).

¹²² CAIRNS & MCKEON, *supra* note 110, at 52, 69.

¹²³ See HAMILTON & PERRY, *supra* note 100, at 260–61.

¹²⁴ This kind of provision may not necessarily be implicated in every case, as, for example, when certain property cannot be disposed of freely. *Id.* at 261.

¹²⁵ In this regime, “the community includes not only the acquets and gains of the marriage, but also any property brought into the marriage by either spouse.” Carroll, *supra* note 24, at 27. This prohibition in other countries is explained by the nature of some rights, the subject of which can only be one person.

¹²⁶ See C. CIV. 1387 (France).

¹²⁷ This is a catalogue of rules regarding the rights and duties of spouses from which no derogation may be made. This compulsory regime is laid down by the Civil Code. Its provisions regard, for example, the financial contribution of spouses to household expenses, to the upbringing of children, as well as to the family home.

¹²⁸ C. CIV. 1389 (France).

¹²⁹ C. CIV. 1388 (France).

¹³⁰ *Id.*

¹³¹ C. CIV. 1396 (France).

exception to this rule:¹³² spouses may change their financial relations after two years of having a particular regime.¹³³ In any case, the marital agreement should be agreed upon before a notary and approved by the court.¹³⁴ The marital agreement is enforceable after the decision of the court and has effect on third persons¹³⁵ three months after mention of it has been entered into the margin of both copies of the record of marriage.¹³⁶ Spouses should make a suitable motion to enter information about their marital regime into the record of marriage.

The marital agreement should be concluded by prospective spouses, or by spouses during the marriage, even if one of them has not gained full legal capacity. A minor who obtained consent for contracting into the marriage or an adult in guardianship or curatorship may enter into marital agreements, but must then be assisted by the person who is authorized to give consent for the marriage.¹³⁷ The spouses may also give power of attorney to an agent who will conclude the marital agreement acting on his or her behalf.¹³⁸

It is very important that the parties be simultaneously present and that the marital agreement be made with the consent of both spouses or their agents.¹³⁹ As previously mentioned, the marital agreement needs to be in the form of a notarized deed. The notary public delivers to the spouses a certificate which confirms that the agreement has been

¹³² See C. CIV. 1397 (France).

¹³³ It must be demanded that the change of a matrimonial regime be made "in the interest of the family." *Id.*

¹³⁴ The court referenced here is the court of the spouses' domicile. *Id.*

¹³⁵ This means that spouses are able to rely on the provisions of the marital agreement in limiting their ownership in certain property, particularly when it comes to the creditors of one spouse. These provisions therefore actually determine the scope of property available to creditors. This issue is particularly important in countries where the community of property is the default statutory regime. In this regime, the creditor of one spouse typically has recourse against the community property. But if the spouses limited the community property through a marital agreement, the rights of the creditor would be restricted. On the other hand, if a creditor has recourse only against the separate property of the spouse who is the debtor, and the spouses extend the community property, the creditor's position is weaker. In this case, the creditor has recourse only against the separate property, such as when, for example, the debt regards the separate property of the spouse or the contract was concluded without the consent of the other spouse.

¹³⁶ See C. CIV. 1397 (France).

¹³⁷ See C. CIV. 148 (France).

¹³⁸ Katharina Boele-Woelki, *Matrimonial Property Law from a Comparative Law Perspective*, in C111 SERIES ARS NOTARIATUS 7 (Hans Warendorf trans., 2000) available at <http://www.reading.ac.uk/nmsruntime/saveasdialog.aspx?IID=7018&sID=34870> [hereinafter "Boele-Woelki, *Matrimonial Property Law*"].

¹³⁹ C. CIV. 1394 (France).

completed. This certificate must be lodged with the officer of civil status before the celebration of the marriage.¹⁴⁰

Finally, in their marital agreements, spouses may make provisions for a spouse's death. Specifically, they may decide that the surviving spouse be authorized to receive from the common property "either a specified sum, or a specified property in kind, or a specified quantity of a determined kind of property."¹⁴¹ Such a provision does not affect the rights of the surviving spouse under inheritance law.¹⁴² Expenses arising during the marriage may also be allotted to each spouse by a marital agreement.¹⁴³

2. Germany

According to German law, spouses enter into a marital agreement (Ehevertrag) to regulate their financial relationships.¹⁴⁴ This definition of marital agreement is quite broad and therefore the source of some doubts. For example, it is not clear whether contracts made between spouses in order to transfer ownership of a certain part of their property should be treated as marital agreements.

The marital agreement can introduce a marital regime or change the rules of the regime chosen by the spouses. This means that the marital agreement is a special kind of tool used to decide the financial consequences of marriage and the financial relations between spouses.¹⁴⁵ Through it, spouses may choose from the contractual property regimes in Germany, which include separation of property¹⁴⁶ and community of property.¹⁴⁷

¹⁴⁰ C. CIV. 1395 (France).

¹⁴¹ C. CIV. 1515 (France).

¹⁴² C. CIV. 1516 (France).

¹⁴³ According to article 214 of the French Civil Code, when spouses do not regulate this matter, they shall contribute to the marriage expenses in proportion to their respective means. *See* C. CIV. 214 (France).

¹⁴⁴ Only spouses can be parties to these agreements. If other people are parties to the agreement, it does not qualify as a marital agreement. *See* GERMAN LEGAL SYSTEM & LAWS 467 (Nigel G. Foster & Satish Sule eds., 3d ed. 2002).

¹⁴⁵ *See* INTRODUCTION TO GERMAN LAW 257–58 (Mathias Reimann & Joachim Zekoll eds., 2d ed. 2005). It should be emphasized that different contracts between spouses such as, for example, donation or loan contracts are not marital agreements and are governed by the general rules of BGB (German Civil Code), instead of by the provisions regarding marital agreements. INTRODUCTION TO GERMAN LAW 193 (Werner F. Ebke & Matthew W. Finkin eds., 1996).

¹⁴⁶ Separation of property is a regime in which spouses hold their property separately. *See* Foster & Sule, *supra* note 144, at 468.

¹⁴⁷ *See* BGB §§ 1414–1415. Community of property is a regime in which there are three groups of assets: community property, the property of the wife, and the property of the husband. *See* Foster & Sule, *supra* note 144, at 468. This Part of the Article uses the term "community of property" to describe this specific property regime and the term

If the spouses have not concluded a marital agreement, they remain in the default statutory regime, which is community of surplus (Zugewinnsgemeinschaft).¹⁴⁸ This name is misleading, however, because it is actually a regime based on separation of property during the marriage, with the surplus divided at the end of the marriage.¹⁴⁹ If the spouses decide not to contract around this default statutory regime, they are free to make some changes to this regime.¹⁵⁰ For example, spouses may eliminate some of the restrictions regarding the transfer of certain assets or change the rules concerning the equalization of accruals that determine the division of the surplus.¹⁵¹

Meanwhile, if the spouses contract into the community property regime, they are able to establish the rules on the composition of each spouse's separate capital and community property.¹⁵² They may also change the rules of management of the community property, as well as the rules regarding the division of common property.¹⁵³

Generally, there is some freedom of contract in marital agreements in Germany, but spouses must choose one of the regimes stated in the Civil Code.¹⁵⁴ It is also possible to modify some of the statutory rules, but only within certain limits established by law. Spouses are not allowed to introduce any foreign law's regime that is not recognized by German law.¹⁵⁵ It is also forbidden to mix different regimes.¹⁵⁶ Such limited freedom of contracting in marital agreements is justified by the desire for certainty of business and the guarantee of formality.

In Germany, the marital agreement can be concluded by spouses or prospective spouses. If a spouse does not have full legal capacity, German law is more restrictive than French law. A prospective spouse with limited legal capacity must be assisted by his or her legal representative and, in certain circumstances,

"community property" to describe the property that belongs to both spouses in this regime.

¹⁴⁸ Foster & Sule, *supra* note 144, at 468.

¹⁴⁹ See Reimann & Zekoll, *supra* note 145, at 257. When analyzing this regime and its rules, one can come to the conclusion that it should actually be called the sharing of accruals or the community of increase. See GERHARD ROBBERS, AN INTRODUCTION TO GERMAN LAW 283 (4th ed. 2006).

¹⁵⁰ GERHARD ROBBERS, *supra* note 149, at 283.

¹⁵¹ *Id.*

¹⁵² BGB § 1474.

¹⁵³ See *id.*

¹⁵⁴ See, e.g., Reimann & Zekoll, *supra* note 145, at 257–58.

¹⁵⁵ See 2008 *Martindale-Hubbell Law Digest – European Law Digests* 1 GERMANY: LAW DIGEST OF THE FEDERAL REPUBLIC OF GERMANY 13.03.

¹⁵⁶ See BGB § 1409.

must have the approval of the court (Vormundschaftsgericht).¹⁵⁷ Still, both parties to the agreement must be present in front of the notary public.¹⁵⁸ As in French law, the marital agreement should be made before a notary, but the approval of a court is not necessary if the agreement needs to be changed.¹⁵⁹

The marital agreement should be registered in a special register, kept by a court (Amtsgericht) and called the register of marital regimes (Güterrechtsregister).¹⁶⁰ The registration makes the marital agreement enforceable against third persons from the day of its registration.¹⁶¹ This duty of registration ensures that third persons know the financial situation of the parties, including the property arrangements made in the marital agreement between the spouses that would impact the scope of property available to the creditors of one of the spouses. The motion to register the marital agreement may be made by either spouse.

Due to this registration, third persons are protected. They may rely on the fact that the spouses are in the regime mentioned in the register. If they are not mentioned because they have changed the regime but their marital agreement has not been published in the register, third parties are not affected.¹⁶² Finally, the register is public and each person who is interested may access it without providing a reason.¹⁶³

3. Switzerland

Another example of European regulation of marital agreements is found in Swiss law, which treats the marital agreement as a special kind of contract concluded by spouses or prospective spouses in order to choose or modify their marital regime.¹⁶⁴ In such agreements, spouses are free to introduce general rules regarding the classification of their property and to modify the rules of their marital regime.¹⁶⁵ Spouses are also able to choose their marital regime or change it within the limits introduced by law.¹⁶⁶

¹⁵⁷ BGB § 1411.

¹⁵⁸ BGB § 1318.

¹⁵⁹ BGB § 1410.

¹⁶⁰ BGB § 1558.

¹⁶¹ See ROBBERS, *supra* note 149, at 283.

¹⁶² It does, however, take effect between spouses.

¹⁶³ See BGB § 1558.

¹⁶⁴ See Barbara E. Graham-Siegenthaler, *Principles of Marriage Recognition Applied to Same-Sex Marriage Recognition in Switzerland and Europe*, 32 CREIGHTON L. REV. 121, 123 (1998).

¹⁶⁵ See Jacques-Michel Grossen & Olivier Guillod, *Family Law*, in INTRODUCTION TO SWISS LAW, 59, 64 (F. Dessemontet & T. Ansay eds., 3d ed. 2004).

¹⁶⁶ HAMILTON & PERRY, *supra* note 100, at 670.

The default statutory regime is the deferred community of acquisitions (*Errungenschaftsbeteiligung*).¹⁶⁷ In this regime, each spouse has his or her separate property during the marriage and upon divorce there is a distribution of goods.¹⁶⁸ Spouses who decide to remain in this standard statutory regime are free to change the rules of the classification of property.¹⁶⁹ Importantly, if they are entrepreneurs, spouses may choose which property will receive income from their commercial activities.¹⁷⁰ They may also change each other's share in acquisitions.¹⁷¹

Spouses may also contract around this default and choose a separate property regime or one of a few types of community property.¹⁷² If community of property is the regime chosen by the spouses, they may decide that their community property will consist only of accruals.¹⁷³ They may also establish the rules concerning shares in common property.¹⁷⁴ However, there is a catalog of marital agreements introduced by law and it is forbidden to create regimes containing only certain elements of these different regimes.¹⁷⁵

The marital agreement (*Ehevertrag*) may be concluded by spouses during a marriage or by prospective spouses (*Brautleute*) before wedlock.¹⁷⁶ In order to conclude a valid marital agreement, spouses must possess legal capacity and the contract must be publicly authenticated.¹⁷⁷ Importantly, the marital agreement may be changed at any time during marriage.¹⁷⁸ Contrary to German law, the ZGB does not contain any requirements to publish the marital agreement in a register, however, a marital agreement can be mentioned in a commercial register.¹⁷⁹

B. The Polish Approach

The main subject of the European Part of this Article is Poland's approach to marital agreements. Polish law is interesting for two reasons. First, the Polish approach is different from the American one in many respects. Second,

¹⁶⁷ *Id.*

¹⁶⁸ *Id.*

¹⁶⁹ *Id.*

¹⁷⁰ *Id.*

¹⁷¹ See ZGB art. 216–217.

¹⁷² HAMILTON & PERRY, *supra* note 100, at 670.

¹⁷³ *Id.*

¹⁷⁴ *Id.* at 673.

¹⁷⁵ *Id.*

¹⁷⁶ ZGB art. 179.

¹⁷⁷ See ZGB art. 180–181.

¹⁷⁸ HAMILTON & PERRY, *supra* note 100, at 673.

¹⁷⁹ See generally ZGB.

Poland's present regulation of marital agreements is quite new and was influenced by the above mentioned regulations, which have longer traditions and are well-established.¹⁸⁰ Nonetheless, the concept of the marital agreement has a long history in Poland, especially when compared with the American history of premarital agreements. Even in communist times, Polish law guaranteed spouses the possibility of contractually opting out of the standard statutory regime.¹⁸¹

Marital agreements in Poland are special treaties for spouses.¹⁸² According to Polish law, the marital agreement (*umowa majątkowa małżeńska*, *intercyyza*) is a contract concluded by spouses or prospective spouses in which their property is regulated in a different way than from the default statutory regime.¹⁸³ The marital agreement also organizes the property of the spouses and dictates the ownership of each spouse as a rule in the future.¹⁸⁴ The essence of marital agreements is that they classify property after the agreement comes into force. However, it is also possible to conclude an agreement in which spouses introduce separation of property and divide their common property.¹⁸⁵ In such a case, it is unclear whether this agreement should be treated as a marital agreement because it concerns the previous property as opposed to property which may be purchased in the future.¹⁸⁶

In Poland, freedom of contract in regards to marital agreements is limited. Such limitations are justified by the aim for certainty of transactions, equity for spouses, and protection of family interests.¹⁸⁷ Spouses are free to introduce regimes named in KRO article 47 § 1,¹⁸⁸ but their freedom is confined to the systems provided by law.¹⁸⁹ When the circumstances causing the mandatory regime have ceased, spouses who were in the mandatory regime may decide to reinstate the previous regime or

¹⁸⁰ It is worth noting that from 1918, when Poland regained its independence, to the end of 1946, German law was in force in west and north Poland and the source of civil law in the central part of Poland was the Napoleonic Code, which was replaced by the civil code of the Kingdom of Poland based on the Napoleonic Code. From 1836 to 1918, the matrimonial law in central Poland was regulated by provisions imposed by the tsarist authorities. For further background, see Andrzej Mączyński, *The Influence of European Family Law on the Family Law of Countries Acceding to the EU. The example of Poland*, in PERSPECTIVES, *supra* note 104, at 239.

¹⁸¹ See DOMINIK LASOK, POLISH FAMILY LAW 92 (A.W. Sijthoff-Leyden, 1968).

¹⁸² PERSPECTIVES, *supra* note 104, at 240–41.

¹⁸³ See generally KRO.

¹⁸⁴ PERSPECTIVES, *supra* note 104, at 240–41.

¹⁸⁵ KRO art. 47.

¹⁸⁶ LASOK, *supra* note 181, at 92.

¹⁸⁷ See PERSPECTIVES, *supra* note 104, at 240–41.

¹⁸⁸ KRO art. 47 § 1.

¹⁸⁹ For further discussion, see LASOK *supra* note 181, at 89–96.

choose one of the regimes mentioned in KRO article 47 § 1.¹⁹⁰ The spouses cannot introduce any other regime apart from those regimes authorized by law.¹⁹¹

If spouses do not choose another regime through their marital agreement, the default statutory regime is limited community of property,¹⁹² similar to that of French law.¹⁹³ This means that there are three types of assets in the marriage: the community property, the separate property of the wife, and the separate property of the husband. Both spouses own community property jointly, whether or not the property has been purchased jointly or separately.¹⁹⁴ In this regime, however, spouses cannot change the rules concerning the management of property.¹⁹⁵ This prohibition is in force for both the standard statutory regime of community of property and the contractual extended or limited community of property.¹⁹⁶ Spouses are not allowed to modify the rules of the primary regime and the rules concerning the liability of spouses.¹⁹⁷

This statutory default of limited community of property is problematic, especially when either one or both of the spouses decide to start a commercial activity. The rules of management and liability for debts may make it difficult to be a married entrepreneur or even a shareholder,¹⁹⁸ increasing the importance of having alternate systems into which spouses may contract by means of a marital agreement.

If the spouses decide to extend community property through their marital agreement, however, they are not able to choose universal community property¹⁹⁹ and at least some chattels must belong to the separate property of each spouse, which is a group of chattels that cannot be part of community property. These are enumerated in KRO article 49.²⁰⁰ Spouses are not allowed to

¹⁹⁰ *Id.*

¹⁹¹ *See id.*

¹⁹² For further background on the property system in Poland, as well as on the separation of property, see Elżbieta Skowrońska-Bocian, *Family and Succession Law*, in INTRODUCTION TO POLISH LAW 85, 96–98 (Stanisław Frankowski ed., 2005).

¹⁹³ *Id.*

¹⁹⁴ Skowrońska-Bocian, *supra* note 192, at 97.

¹⁹⁵ *Id.* at 98.

¹⁹⁶ To contractually expand or limit community property, couples simply designate more or less of their property as community property. *See also infra* note 204 and accompanying text.

¹⁹⁷ *See, e.g.*, KRO arts. 41 & 48.

¹⁹⁸ For more details, see ANNA STĘPIEŃ-SPOREK, *DZIAŁALNOŚĆ GOSPODARCZA Z UDZIAŁEM MAŁŻONKÓW* (2009).

¹⁹⁹ In this regime, all property is owned by the wife and husband in common. *See* Carroll, *supra* note 24, at 27.

²⁰⁰ KRO art. 49.

extend the scope of community property to embrace inalienable rights, compensations for personal injury, or material damage or claims for remuneration for work or personal services outstanding at the time of the marriage.²⁰¹ If the spouses decide on a contractual community of property (extended or limited), the rules of administration of community property from the statutory regime are applied.²⁰² The agreement may have a provision for unequal division of common property upon the end of the regime.²⁰³

According to KRO article 47 § 1, spouses may extend or limit community property,²⁰⁴ or choose separation of property²⁰⁵ or the sharing of accruals.²⁰⁶ If spouses choose extended or limited property, they may establish that in the case of the liquidation of the community of property, the fractions of each spouse will differ. By choosing the sharing of accruals, spouses may also change the rules of calculation of accruals.²⁰⁷

Therefore, the marital agreement regulates the spouses' property during the duration of the chosen agreement. Spouses are free to conclude different contracts, but only a few of these are considered marital agreements because they must fulfill certain conditions. For example, marital agreements must take a special form in order to be valid: if spouses wish to change their matrimonial regime (either the statutory standard regime or the contractual regime), a notary must be involved.²⁰⁸

Contrary to American law, the circumstances regarding the formation of the marital agreements in Poland are not as important and are rarely taken into account. For example, Polish law does not pay significant attention to the fair disclosure of each spouse's financial details.²⁰⁹ However, the general rules concerning defects of a will are applied to marital agreements,

²⁰¹ *Id.*

²⁰² KRO art. 47.

²⁰³ The end of the community regime is upon divorce, legal separation, nullity of marriage, the introduction of another property regime, the death of a spouse, or in situations when the compulsory regime is applied. The compulsory regime is the separation of property. See Skowrońska-Bocian, *supra* note 192, at 99.

²⁰⁴ See *id.* at 98–99.

²⁰⁵ In this regime, spouses hold their property separately.

²⁰⁶ This regime is similar to the German and Swiss statutory accrual system, as well as to the French deferred community (separation of acquisitions). The general rule of this regime is that during marriage, each spouse is the sole owner of his or her separate property. Upon termination of the regime, the surplus of separate property of both spouses is divided. For further details, see Anna Stępień-Sporek, *Rozdzielność Majątkowa z Wyrównaniem Dorobków*, 7 PAŃSTWO I PRAWO 73 (2008).

²⁰⁷ Anna Stępień-Sporek, *Sharing of Accruals as the Best Solution for Marriage?*, in FAMILY FINANCES, *supra* note 107 at 371, 375.

²⁰⁸ KRO art. 47.

²⁰⁹ See generally KRO.

i.e., the spouse should not be mistaken, under threat, or in a state of mind excluding the conscious making or expressing of the contract.²¹⁰

In Poland, marital agreements may be concluded by both spouses and prospective spouses. If the marital agreement is concluded by prospective spouses, it will not come into force until they are married. If a prospective spouse or a spouse does not have full legal capacity, the consent of the legal representative is necessary, as is the consent of the court.²¹¹ However, the marital agreement may also be concluded by a proxy.²¹² The form of the power of attorney is essential. If the marital agreement is to be concluded by prospective spouses, the name of the other prospective spouse should not only be mentioned in the document of the power of attorney, but also given in notarial deed.²¹³

The marital agreement—which takes effect either at the moment of conclusion or as established by the spouses—can be changed as often as the spouses desire to change their financial relationships.²¹⁴ A similar rule characterizes German and Swiss law.²¹⁵

The marital agreement in Poland has effects in relation to third parties if they are informed of the agreement and of the regime chosen by the spouses.²¹⁶ This rule is essential. The regime can affect creditors, who are protected by the above mentioned rule. If the spouses have failed to inform third parties, however, the marital agreement does not have any effect on these third parties. In practice, this kind of situation is very common. Third persons who are not aware of the existence of the marriage contract may assume that the spouses are married under the statutory system.

In Poland, there is no special requirement to register a marital agreement,²¹⁷ however, information about marital agreements can be published in commercial registers such as the Krajowy Rejestr Sądowy, the register of companies and stocks, or Ewidencja Działalności Gospodarczej, the register of individuals who are entrepreneurs.

²¹⁰ See Kodeks cywilny – The Law of 23rd April 1964 – art. 82–88 [hereinafter “Kc”].

²¹¹ See KRO art. 101 § 3.

²¹² Skowrońska-Bocian, *supra* note 192, at 92.

²¹³ See Kc art. 99 § 1 and KRO art. 47.

²¹⁴ KRO art. 47.

²¹⁵ See Boele-Woelki, *Matrimonial Property Law*, *supra* note 138, at 7–8.

²¹⁶ KRO art. 47.

²¹⁷ See generally KRO.

III. A COMPARATIVE STUDY

While European countries may differ slightly in their approaches to marital agreements, they all differ markedly from the American approach.²¹⁸ These various contrasts illustrate the range of prospective spouses' possibilities in premarital contracting, as well as the options that legislatures have in terms of regulating premarital agreements. A comparison of these varying approaches to premarital agreements also offers important insights regarding the contractual autonomy of the parties, the possible characteristics of premarital agreements, and the popularity of such agreements.

A. Contractual Autonomy of the Parties

While European matrimonial property law is codified in each country's civil code, the American tradition of freedom of contract provides spouses with the power to contract around state statutory law on the subject.²¹⁹ Americans are therefore not restricted to the property regimes laid out in statutes—whether community property or equitable distribution—and may contract around them subject to few limitations by the court.²²⁰ In fact, spouses may even import into their agreements any of the European property systems, such as a system of accruals.²²¹ Meanwhile, Europeans are often limited to selecting one of the property regimes statutorily permitted in their countries. Although this permits them to avoid the statutory default, they must nonetheless select one of the regimes recognized by law. Only occasionally may spouses alter the rules of those European systems. Americans therefore enjoy more autonomy in premarital contracting relative to Europeans.

These differing levels of contractual autonomy have differing consequences. For example, the general notion espoused by American law is that people should be able to manage their property as they choose,²²² which also justifies the significant

²¹⁸ Only continental Europe, which abides by the civil law system, differs markedly from the United States. British common law is similar on this subject. *See supra* note 107.

²¹⁹ *See, e.g.,* DeMatteo v. DeMatteo 762 N.E.2d 797 (Mass. 2002) (discussing freedom of contract with regard to American premarital agreements).

²²⁰ *See supra* Part I for further discussion.

²²¹ *See supra* Part II.

²²² *See, e.g.,* Shaffer v. Shaffer, 733 P.2d 1013, 1016 (Wash. Ct. App. 1987):

Because of this new freedom for marital partners to divide their property as they see fit, the old rule allowing the court to disregard the property division made by the parties in their [separation] agreement if the division does not conform to the trial court's view of an equitable property division, no longer is appropriate.

freedom provided to people in the contracting of their wills.²²³ Furthermore, limited contractual autonomy is criticized as being paternalistic and less efficient.²²⁴ However, one consequence of such significant contractual autonomy is that American courts must occasionally find a premarital agreement substantively unenforceable, particularly when one party's imaginative contracting significantly disadvantages the other. In the European countries considered, meanwhile, marital agreements will rarely be substantively or procedurally unenforceable because they must adhere to strict statutory requirements in the first place. Furthermore, European spouses cannot introduce their own regime, but instead must select one of the statutory systems recognized by the law. As the Polish approach illustrates, European courts may therefore be slower to find procedural defects in marital agreements.²²⁵ This may also be due to the fact that these marital agreements must be concluded in the presence of the public notary, whose significant role in civil law countries is much greater than in common law countries.²²⁶

Nonetheless, there may be a relationship between contractual autonomy and the risk of an agreement's unenforceability. In other words, the less formulaic the premarital agreement, and the less deferential to statutory defaults, the more opportunity for a judicial declaration of unenforceability. In the quest for the right legislative framework to regulate premarital or marital contracting, then, the task becomes to find the right balance between autonomy and the risk of unenforceability.

However, complete autonomy may nonetheless be prioritized because it would permit prospective spouses to choose their own level of risk regarding the enforceability of their agreements. Of course, permitting spouses unrestricted contractual autonomy, thereby allowing them to choose their agreement's risk of

See also Melvin A. Eisenberg, *The Theory of Contracts*, in *THE THEORY OF CONTRACT LAW* 206, 223 (Peter Benson ed., Cambridge U. Press 2001) ("Autonomy theories of contract are based on the concept that allowing an individual to freely own and dispose of property and freely exercise his will to make choices concerning his person, labor, and property, is a value that is paramount.").

²²³ See, e.g., Trent J. Thornley, Note, *The Caring Influence: Beyond Autonomy as the Foundation of Undue Influence*, 71 IND. L.J. 513, 516 (1996) ("The freedom of testamentary disposition is a basic principle of property law. Though not a constitutional right, many states recognize freedom of testation as a deeply ingrained tradition in our society."). See also *supra* Part I.D.2.

²²⁴ See *supra* notes 78 and 81.

²²⁵ See *supra* Part II.B.

²²⁶ See JAMES G. APPLE & ROBERT P. DEYLING, A PRIMER ON THE CIVIL-LAW SYSTEM 30, [http://www.fjc.gov/public/pdf.nsf/lookup/CivilLaw.pdf/\\$file/CivilLaw.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/CivilLaw.pdf/$file/CivilLaw.pdf) (last visited Dec. 28, 2009).

unenforceability, requires somewhat perfect information. Prospective spouses must not only know that straying from formulaic premarital agreements—as well as from typical legislative defaults—increases their agreements' risk of unenforceability, but they must also be aware of the advantages and disadvantages of choosing particular property regimes and contractual provisions. A lawyer representing each side could help assure this.

Another consequence of unbounded autonomy of premarital contracting, however, is that the courts must continually determine the enforceability of each individual agreement before applying its provisions to a divorce. Scarce judicial resources must therefore be spent, despite the American judiciary's traditional reluctance to meddle in family matters.²²⁷

Therefore, keeping prospective spouses informed of their options and expending judicial resources to monitor premarital agreements are the consequences to a high level of autonomy in premarital contracting, as illustrated by the American approach to premarital agreements and a comparison of it to European approaches. Nonetheless, every jurisdiction, whether an American state or an European country, must choose its own balance among these factors and costs.

B. Potential Characteristics of Premarital Agreements

In searching for an appropriate regulatory framework for premarital contracting, as well as the proper content of such agreements, it is instructive to analyze the desirability of the differing approaches to these agreements, including those that deal with aspects beyond the freedom of contract already considered.²²⁸

For example, some European countries permit, or even require, the registration of marital agreements in order to protect third parties.²²⁹ This registration aims to give creditors notice as to which assets are available for collection. There is a similar requirement in some American community property states for the recording of separation of property agreements.²³⁰ To give

²²⁷ As the courts do not typically become involved in the financial arrangements of intact families, spouses are permitted to determine their own responsibilities during marriage. *See, e.g.*, *Kilgrow v. Kilgrow*, 107 So. 2d. 885 (Ala. 1958); *State v. Rhodes*, 61 N.C. 453 (N.C. 1868).

²²⁸ *See supra* Part III.A.

²²⁹ *See supra* Part II for examples and discussion.

²³⁰ *Carroll, supra* note 24, at 32.

adequate notice to creditors, the agreement binds the spouses once executed, but binds third-party creditors only if recorded.²³¹

Such registration indeed serves an important, albeit lesser-known, purpose of premarital agreements, which is to organize how the spouses hold their assets during the marriage. If spouses hold their property differently from the statutory default, it may be only fair to give notice of this arrangement to third parties who potentially rely on the spouses' property holdings when extending credit.

A third party would be most disadvantaged by a couple secretly opting out of a community property default. In such a case, the third party would ordinarily expect all of the marital assets to be held by each spouse, when in reality, each spouse holds smaller assets separately. However as already mentioned, only a minority of American states utilizes community property as the default system, whereas some European countries use this system as their statutory default. Therefore, public registration of marital agreements is significantly more relevant in the European context.

Americans, meanwhile, maintain the confidentiality of premarital agreements.²³² This effectively reduces third parties' ability to rely on such agreements. Such confidentiality also permits prospective spouses in the United States to contract on intimate details and on their private financial situations. In this way, however, American premarital agreements are often limited to affecting only the married couples who are parties to the agreements.

Finally, in many European countries, prospective spouses must adhere to particular formalities in order to conclude a legally enforceable premarital agreement, such as signing the agreement in front of a notary.²³³ There is no such definitive list of requirements in the United States, although a court may subsequently analyze the procedural fairness of the agreement when called upon to enforce the document.²³⁴ Such procedural review may, however, increase the opportunity for the court to find a premarital agreement unenforceable.²³⁵ The level of required formality will therefore inevitably vary from jurisdiction to jurisdiction as each adopts a particular balance between

²³¹ See, e.g., *id.*

²³² Guggenheimer, *supra* note 30, at 153.

²³³ See, e.g., *supra* note 158 and accompanying text.

²³⁴ See *supra* Part I.D.1.

²³⁵ See *supra* Part III.A.

permitting contractual autonomy and protecting the parties from fraud.

Therefore, choices must be made not only among premarital contracting options, but also among the different regulatory frameworks for such agreements. In choosing the appropriate model, each jurisdiction and couple must therefore weigh the costs of their choices.

C. Popularity of Premarital Agreements

Premarital agreements are more popular in the United States than in Europe, but not particularly popular in either.²³⁶ There may be several explanations for this current lack of popularity in premarital contracting on both continents. Many commentators suspect, however, that the agreements' popularity will increase in the near future.²³⁷

Importantly, premarital agreements in the United States might be rare because they need not necessarily be drafted for a higher income prospective spouse to avoid an unfavorable statutory default. This is because the higher income earner would prefer an equitable distribution regime, which often results in an unequal distribution, rather than a community property one, which instead results in an equal division of assets.²³⁸ However, most American states use equitable distribution as the default, mooted the need to alter this regime through a premarital agreement. Many European countries, on the other hand, have a statutory default of some type of

²³⁶ In the United States, however, the use of premarital agreements tripled between 1978 and 1988 alone. Guggenheimer, *supra* note 30, at 151. "Although it is difficult to get statistics on premarital agreements, it appears that 5% to 10% of couples marrying for the first time and 20% of remarried couples now enter into premarital agreements." Brian McDonald, Presentation to the Western Trial Lawyers Association (June 2005) (transcript available at <http://www.spomcman.com/doc/PREMARITAL%20AGREEMENTS.doc>). Meanwhile, only 2% of British couples marrying seek a premarital agreement. *Divorce Lawyers Braced for Busiest Week Ever*, TIMES ONLINE, January 5, 2009, <http://business.timesonline.co.uk/tol/business/law/article5450552.ece>. In France, only 10% of spouses conclude a marital agreement, and they do so mostly when important assets or second marriages are involved. See HAMILTON & PERRY, *supra* note 100, at 261.

²³⁷ "Premarital agreements are gaining popularity as more people become conscious of the extensive financial rights and obligations arising out of a marriage, and the increasing statistical chance that any marriage will end in divorce." *In re Marriage of Leathers*, 789 P.2d 263, 265 n.5 (Or. 1990) (quoting 12 ABA Family Advocate, No. 3, 54-55 (Winter 1990)). See also Jennifer Kim, *Contesting the Enforceability of a Premarital Agreement*, 11 J. CONTEMP. LEGAL ISSUES 133, 133 (2000); Jennifer L. McCoy, Comment, *Spousal Support Disorder: An Overview of Problems in Current Alimony Law*, 33 FLA. ST. U. L. REV. 501, 523 (2005).

²³⁸ For further background on the relationship between equitable distribution and community property regimes and their European counterparts, see Carroll, *supra* note 24, at 1.

community property, which might sooner prompt a higher income earner to seek a premarital agreement.²³⁹

Nonetheless, Americans may use premarital agreements to regulate many of their rights and responsibilities during marriage and divorce, particularly in regards to specific assets they own. Premarital agreements are especially useful to prospective spouses who fit a particular profile. For example, people with children from previous marriages may choose to protect these children's financial future by virtue of a premarital agreement.²⁴⁰ People may also utilize such agreements when they are skeptical of the institution of marriage because of their own, or their parents', failed marriages.²⁴¹ Premarital agreements may also be more common among prospective spouses with significant income or age disparities.²⁴² Further, prospective spouses may choose to keep their property separate by such agreements so that one can use only his or her portion in paying off debts.²⁴³ Finally, when one partner expects to inherit significant money or a family business, she may decide to request a premarital agreement.²⁴⁴

Still, premarital agreements are not frequently used in either Europe or the United States, with most commentators estimating that less than 10% of any of these populations use such agreements.²⁴⁵ In Europe, this may be due to a potential misunderstanding of the role of the premarital agreement, which most Europeans associate with American celebrity divorces,²⁴⁶ although the meaning and consequences of premarital agreements in Europe differ from those in the United States.²⁴⁷ In the United States, meanwhile, the limited use of the premarital agreement may be due to Americans' rather unrealistic sense of optimism regarding their marriages and their fear of signaling divorce.²⁴⁸ This, as well as the need to

²³⁹ However, it is true that premarital agreements are unpopular in Europe, as well. See *supra* notes 101, 236 and accompanying text.

²⁴⁰ Guggenheimer, *supra* note 30, at 152.

²⁴¹ *Id.* at 151.

²⁴² *Id.* at 152.

²⁴³ *Id.*

²⁴⁴ *Id.*

²⁴⁵ See *supra* note 236.

²⁴⁶ See *supra* Part II.A. This may also be because the average premarital agreement is highly confidential. Guggenheimer, *supra* note 30, at 153.

²⁴⁷ See *supra* Parts I and II.

²⁴⁸ For the argument that most couples are overly optimistic about their marriages, see Margaret F. Brinig, FALL AND RISE OF FREEDOM OF CONTRACT, *supra* note 64, at 276 and Sean Hannon Williams, *Sticky Expectations: Responses to Persistent Over-Optimism in Marriage, Employment Contracts, and Credit Card Use*, 84 NOTRE DAME L. REV. 733, 757–61 (2008). See also *supra* note 2.

protect children from a previous marriage,²⁴⁹ may also explain why many people are more likely to seek premarital agreements upon second and subsequent marriages. Nonetheless, many prospective spouses around the world are currently choosing not to pursue the benefits offered by premarital contracting, although this may change in the near future.

CONCLUSION

In sum, the premarital agreement permits prospective spouses around the world to circumvent their jurisdiction's judicial and statutory defaults in organizing the terms of their marriage and potential divorce. To achieve this force in the United States, the premarital agreement has particularly undergone significant development over the course of the past few decades.

Although the American history of the premarital agreement is relatively short compared to its European counterpart, Americans have quickly achieved unparalleled levels of freedom in marital contracting. This heightened freedom of contract has become one of the most significant differences between the continental European and American approaches to such agreements.

Furthermore, a comparative analysis of these various approaches suggests that the level of autonomy in marital contracts implicates the risk of the agreement's unenforceability by the courts. A comparative study also offers some insights into improving the regulatory frameworks governing these agreements, as well as the reasons behind people's reluctance to use such agreements.

In conclusion, although the premarital agreement has attained significant stability and enforceability in countries around the world, issues surrounding such agreements undoubtedly remain. Specifically, premarital agreements in the United States are subject to procedural and substantive review. They also raise universal public policy issues, particularly in the case of mobile couples, concerning the meaning of fairness and the limits on freedom of contract. As state courts and legislatures continue to encounter and address these issues, they may therefore greatly benefit from a comparative study of such agreements.

²⁴⁹ Gail Frommer Brod, *Premarital Agreements and Gender Justice*, 6 YALE J.L. & FEMINISM 229, 238–39 (1994).

Central Bank Autonomy in Latin America: A Survey and Case Studies

*Harout Jack Samra**

There is no subtler, no surer means of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction, and does it in a manner which not one man in a million is able to diagnose.

— John Maynard Keynes¹

INTRODUCTION

Beads of sweat forming on his furrowed brow, an uncharacteristically anxious Hugo Chavez appeared on television from the Miraflores Palace early in the morning of December 3, 2007.² For a nation accustomed to seeing its bombastic president sing the praises of his “Bolivarian” Revolution, Chavez struck the pose of a leader under siege. Though many experts had speculated that the sweeping constitutional amendments he had proposed a mere two months earlier might fail, amendments which would have wrought far-reaching changes on Venezuela, few were prepared for the political theatre which ensued when they actually did.³ Given Chavez’s seeming grasp on all the levers of Venezuelan power, many observers noted that the proposed amendments, collectively referred to as the *Reforma de la Constitución de la República Bolivariana de Venezuela* (the *Reforma*), would all but make him the latest in a long line of Latin American *caudillos* that litter the region’s history.⁴

* International litigation and arbitration associate at Squire, Sanders & Dempsey L.L.P. J.D., *magna cum laude*, May 2009, University of Miami School of Law; M.B.A., May 2008, University of Miami School of Business Administration; B.A., *cum laude*, University of Miami, December 2004. I want to thank Professor Stanley Langbein for his patience and dedication and R. Scott Nuzum, Jordan Dresnick, and Natalie Rico for their valuable advice.

¹ JOHN MAYNARD KEYNES, *THE ECONOMIC CONSEQUENCES OF PEACE* 236 (1920).

² See Simon Romero, *Venezuela Vote Sets Roadblocks on Chávez Path*, N.Y. TIMES, Dec. 4, 2007, at A1.

³ Editorial, *Not so Fast, Venezuelans Warn Chavez*, ST. PETERSBURG TIMES, Dec. 8, 2007, at 16A.

⁴ See Roger Cohen, Op-Ed, *Democracy In the Americas*, N.Y. TIMES, Dec. 6, 2007, at A41. See also ANDRÉS OPPENHEIMER, *CUENTOS CHINOS* 223–63 (2005); Mario Garcia-Serra, *The “Enabling Law”: The Demise of the Separation of Powers in Hugo Chavez’s Venezuela*, 32 U. MIAMI INTER-AM. L. REV. 265, 266 (2001). A caudillo is defined as:

In its coverage of the referendum, the media largely focused on the abandonment of term limits for the president, which many suggested would result in Chavez becoming a president for life.⁵ In addition to term limits, the *Reforma* also included sweeping changes to several important governmental institutions, such as the military and the central bank, the *Banco Central de Venezuela*.⁶ Indeed, the *Reforma* explicitly provided for the abolition of central bank independence.⁷ Had the *Reforma* passed, proposed Article 318 would have provided the following, in part: The National Executive authority, through the *Banco Central de Venezuela*, acting in strict and obligatory coordination, will determine the monetary policies and will exert the monetary authority of the National Power.⁸

While the scope of this effort is not limited to the “Bolivarian” travails of Mr. Chavez, this episode makes one fact plain: the independence of central banking authorities in the various Latin American states is frequently the subject of great political upheaval. Indeed, the drastic political turmoil that has long been endemic to the region has also led to frequent and radical policy changes in this central state function.⁹ Whether labeled “neoliberal” or “Washington Consensus,” many of the policies advocated for in this paper are not without precedent in the region—and they certainly are capable of arousing significant passion and controversy at the hands of populist leaders. Indeed, Professor Felipe Larrain of the *Pontificia Universidad Catolica*

A Latin American military dictator. In the wake of the Latin American independence movement in the early 19th century, politically unstable conditions and the long experience of armed conflict led to the emergence in many of the new countries of strongmen who were often charismatic and whose hold on power depended on control over armed followers, patronage, and vigilance. Because their power was based on violence and personal relations, the legitimacy of the caudillos' rule was always in doubt, and few could withstand the challenges of new leaders who emerged among their own followers and wealthy patrons.

ENCYCLOPÆDIA BRITANNICA ONLINE (2009), available at <http://www.britannica.com/EBchecked/topic/100372/caudillo>.

⁵ Chris Kraul, *Venezuelan Voters Reject Bid by Chavez to Extend Powers*, L.A. TIMES, Dec. 3, 2007, at A1.

⁶ *Id.*

⁷ It must be noted that the terms central bank “autonomy” and “independence” are used interchangeably throughout this paper.

⁸ La Reforma de la Constitución de la República Bolivariana de Venezuela (2007) [hereinafter “La Reforma”]. Article 318 of *La Reforma* reads as follows: “El Poder Ejecutivo Nacional, a través del Banco Central de Venezuela, en estricta y obligatoria coordinación, fijará las políticas monetarias y ejercerá las competencias monetarias del Poder Nacional.” All translations of Article 318 are by the author.

⁹ See, e.g., Luis Jácome H., *Legal Central Bank Independence and Inflation in Latin America During the 1990s*, 4 (IMF, Working Paper No. 01/212 2001), available at <http://www.imf.org/external/pubs/ft/wp/2001/wp01212.pdf>.

de Chile pointedly noted that while Latin America has made great strides in controlling inflation, the fact that two of the region's largest economies, Venezuela and Argentina, have returned to double-digit inflation rates of around 20% and 10%, respectively, is deeply troubling and has "led some pundits to wonder whether high inflation—and even hyperinflation—may be on its way back to Latin America."¹⁰ Further, even though the renowned Larrain noted the present absence of many of the historical causes of hyperinflation, he nevertheless seemed to strike a dark note, referring particularly to the case of Chavez's Venezuela:

The true risk of high inflation in Latin America lies in populism. It is no coincidence that Venezuela is the country with the highest rate of inflation in the region—though low by historical standards. Hugo Chavez, Venezuela's president, is the face of Latin American populism, though of a different brand; a rare mix of socialism and populism for export which can be sustained without an explosion of inflation as long as oil-related fiscal revenues remain generous. If that situation changes—and it will when the international cycle turns—the risk of high inflation is at the door. Hyperinflation is not dead, it is merely sleeping. Leaders such as Zimbabwe's Robert Mugabe are capable of awakening the monster.¹¹

This ominous warning is perhaps all the more terrifying given the knowledge that Larrain's expectation that energy prices would fall in the short term has indeed come to pass. Are we merely waiting at this stage for the other shoe to drop?

Central bank autonomy, most scholars agree, is a *sine qua non* of monetary or price stability.¹² Recent scholarship has divided this pillar of economic stability into two categories: *de jure* and *de facto* sovereignty.¹³ Clearly, while this study will disproportionately focus on formal legal structures that have been or may be put into place in order to provide for independent monetary judgment, such formalism often fails to truly explain the full extent of political interference in the central bank decision-making processes. A state may have a finely crafted legal structure in place that, while perfect in form, fails utterly in function. It is with this understanding that I shall consider several important case studies focusing on Latin America. Over the last two decades, reforms strengthening central bank

¹⁰ Felipe Larrain, *The Spectre of Latin American Inflation*, THE BANKER, Sept., 2007 at 176.

¹¹ *Id.*

¹² Agustín Carstens & Luis Jácome H., *Latin American Central Bank Reform: Progress and Challenges*, 6 (IMF, Working Paper No. 05/114 2005), available at <http://www.imf.org/external/pubs/ft/wp/2005/wp05114.pdf>.

¹³ See, e.g., *id.*

autonomy have been implemented throughout the region. Between the late 1980s and 1990s, for example, numerous countries in the region made such reforms, including Chile (1989), El Salvador (1991), Argentina (1992), Colombia (1992), Venezuela (1992), Ecuador (1992 and 1998), Peru (1993), Mexico (1993), Bolivia (1995), Costa Rica (1995), Uruguay (1995), Paraguay (1995), Honduras (1996) and Nicaragua (1999).¹⁴ By highlighting several of these reforms, I hope to outline the existing trends in the region.

Following this brief introduction, Section I of this study will provide a comprehensive review of the economic literature relating to the economic benefits of central bank independence, including a description of the various models and methods economists have used to analyze the issue. Section II will consist of three shorter case studies from the region, specifically addressing the central bank structures that have been adopted in Argentina, Chile, and Mexico. In each case, recent central bank policy shifts will be addressed and placed in historical context. Section III will return to the issue of the Venezuelan amendments proposed by Mr. Chavez. In addition to presenting a detailed analysis of the banking amendments that were offered and rejected by the public, this section will also discuss the aftermath of that stunning electoral rejection and the current state of affairs in Venezuela. Finally, the conclusion will place Venezuela in the broader context of the region and discuss the general trends that seem to be forming throughout Latin America.

I. CENTRAL BANK INDEPENDENCE AND PRICE STABILITY

A. The Scourge of Inflation

Prior to discussing the economic literature on the correlation between central bank autonomy and price stability, it is worth stepping back for a moment and discussing inflation as a phenomenon in itself. I will soon discuss the trade-off faced by monetary decision makers between short-term growth and longer-term inflation, and the balance that central bankers or monetary authorities must seek to achieve between these two. However, pointing to the destructive capacity of this economic phenomenon, the eminent economist Milton Friedman noted, "inflation is a disease, a dangerous and sometimes fatal disease, a disease that, if not checked in time, can destroy a society."¹⁵

¹⁴ Jácome, *supra* note 9, at 6.

¹⁵ MILTON FRIEDMAN, MONEY MISCHIEF: EPISODES IN MONETARY HISTORY 191 (1994).

Indeed, referring to Bolivia, Brazil, and Argentina, Friedman noted that Latin America is hardly a stranger to catastrophic hyperinflation in the modern era.¹⁶ Looking over the history of the region, he noted that hyperinflations in Chile, Brazil, and Argentina imposed broad political and social costs, including the establishment of military governments in the mid-twentieth century.¹⁷ More recently, Friedman asserted, “repeated inflations in Brazil . . . in the 1980s have led to the repeated unsuccessful ‘reforms,’ the replacement of governments, flights of capital, and heightened economic instability.”¹⁸

As to the proximate cause of inflation, Friedman noted that “[t]here is probably no other proposition in economics that is as well established as” the principle that “[i]nflation occurs when the quantity of money rises appreciably more rapidly than output, and the more rapid the rise in quantity of money per unit of output, the greater the rate of inflation.”¹⁹ The fundamental challenge embedded in this dilemma, however, is the fact that “[o]utput is limited by the physical and human resources available and by the degree of knowledge and capacity for using those resources,” whereas “modern forms of money—paper and bookkeeping entries—are subject to no such physical limits.”²⁰

Pointing to the Brazilian experience between 1965 and 1989 during which the quantity of money per unit of output correlated quite closely with the high rate of inflation, Friedman noted that “the price level at the end [was] nearly 6 million times the level at the beginning—an average of 86.5 percent a year.”²¹ Considering the Brazilian dilemma, Friedman predicted that the country would not be able to operate at such “hyperinflationary rates without abandoning its national currency and adopting a substitute.”²² Indeed, in the subsequent years, the government would adopt several substitutes in its long and painful journey back to price stability.²³

B. The Centrality of Central Bank Independence

B.W. Fraser, the former Governor of the Reserve Bank of Australia, perhaps summed up the importance of central bank independence most effectively when he noted, “price stability is

¹⁶ *Id.*

¹⁷ *Id.* at 192.

¹⁸ *Id.*

¹⁹ *Id.* at 193.

²⁰ *Id.* at 193–94.

²¹ *Id.* at 199.

²² *Id.* at 200.

²³ *Id.*

generally considered a good thing, and . . . an independent central bank can help to achieve it. I do not need to dwell on the desirability of price stability. Economies work better if investment and wage decisions are not confused and thwarted by high inflation.”²⁴ Further emphasizing the importance of mitigating inflation, Fraser added, “some people see price stability as an anchor not only for the economy, but also for society at large.”²⁵ He emphasized, however, that keeping prices stable is “*not* the natural order of things in a modern economy.”²⁶

Several centuries earlier, the eminent British economist David Ricardo advocated for greater independence for monetary authorities.²⁷ In his *Plan for the Establishment of a National Bank*, Ricardo made plain his lack of confidence in political decision-making in the arena of monetary policy:

It is said that Government could not be safely entrusted with the power of issuing paper money; that it would most certainly abuse it; and that, on any occasion when it was pressed for money to carry on a war, it would cease to pay coin, on demand, for its notes; and from that moment the currency would become a forced Government paper. There would, I confess, be great danger of this, if Government—as that is to say, the Ministers—were themselves to be entrusted with the power of issuing paper money. But I propose to place this trust in the hands of Commissioners, not removable from their official situation but by a vote of one or both Houses of Parliament. I propose also to prevent all intercourse between these Commissioners and Ministers, by forbidding every species of money transaction between them. The Commissioners should never, on any pretence, lend money to Government, nor be in the slightest degree under its control or influence. Over Commissioners so entirely independent of them, the Ministers would have much less power than they now possess over the Bank Directors. Experience shows how little this latter body have been able to withstand the cajolings of Ministers; and how frequently they have been induced to increase their advances on exchequer bills and treasury bills, at the very moment they were themselves declaring that it would be attended with the greatest risk to the stability of their establishment, and to the public interest. . . . I ask, then, whether the country would not possess a greater security against all such influence, over the minds of the issuers of paper, as would induce them to swerve from the strict line of their duty, if the paper money of the country were issued by Commissioners, on the plan I have proposed, rather than by the Bank of England, as at

²⁴ B.W. Fraser, *Central Bank Independence: What Does It Mean?*, RES. BANK OF AUSTRAL. BULL. 1, 2 (1994), available at http://www.rba.gov.au/PublicationsAndResearch/Bulletin/bu_dec94/bu_1294_1.pdf.

²⁵ *Id.*

²⁶ *Id.* (emphasis in original).

²⁷ *Id.* at 1.

present constituted? If Government wanted money, it should be obliged to raise it in the legitimate way; by taxing the people; by the issue and sale of exchequer bills, by funded loans, or by borrowing from any of the numerous banks which might exist in the country; but in no case should it be allowed to borrow from those who have the power of creating money.²⁸

Other economists, including John Maynard Keynes, later adopted similar views to Ricardo's as to the benefits of central bank independence.²⁹ Further, modern empirical analysis has largely affirmed Ricardo's prescience. Such analysis has shed much light on the real, measurable impact of a central bank beholden to politically motivated actors.

Echoing Ricardo, Federal Reserve Board Chairman Ben Bernanke explained "[t]hat central bank independence promotes lower inflation in developed countries is well-established by the economic literature."³⁰ An insufficiently independent central bank, he added, "may be forced to ease policy inappropriately to reduce the financing costs of government debt or to help re-elect incumbents."³¹ Pointing to Latin America in particular, Bernanke noted that "[i]n the late 1980s and early 1990s, several countries took steps to enhance central bank independence, both through constitutional provisions, and through changes in the central bank law."³² It must be noted, however, that he added the following caveat: "Although the evidence for developing countries is more mixed, several studies have found that central bank independence promotes low inflation in those countries as well, especially when *de facto* rather than *de jure* measures of independence are used."³³ Indeed, Bernanke noted that this distinction is particularly important in Latin America and to convey his point more effectively, he gave the example of Brazil:

[A]lthough a number of features of Brazilian law promote the independence of the nation's central bank, the bank's *de facto* independence may be limited by the power of the president to remove members of the Monetary Policy Committee. However, the fact that the left-of-center Lula administration in Brazil has been able to support disciplined monetary and fiscal policies without significant

²⁸ DAVID RICARDO, PLAN FOR THE ESTABLISHMENT OF A NATIONAL BANK (1824), reprinted in THE WORKS OF DAVID RICARDO 506–07 (J.R. McCulloch ed., 1871).

²⁹ Fraser, *supra* note 24 at 1.

³⁰ Ben S. Bernanke, Chairman, Fed. Reserve Sys., *Inflation in Latin America: A New Era?*, Remarks at the Stanford Institute for Economic Policy Research Economic Summit, (Feb. 11, 2005) (transcript available at <http://www.federalreserve.gov/BoardDocs/Speeches/2005/20050211/default.htm>).

³¹ *Id.*

³² *Id.*

³³ *Id.* See *infra* Part I.D for a discussion of the use of *de facto* and *de jure* measures of independence.

loss of popularity is itself testimony to the degree that such policies are becoming more widely accepted in the region.³⁴

Generally, however, Bernanke's overt endorsement of broad central bank independence, coupled with his lauding of the Lula administration's monetary policy, follows to a large degree from his broader conclusion that policies such as this were largely responsible for the achievement of price stability in Latin America during the 1990s.³⁵

Scholars have summarized the theoretical economic literature as essentially asserting three broad arguments to explain why greater independence in monetary decision making yields lower inflation rates.³⁶ First, is the premise that government financial decision makers will tend to exert "pressure on monetary authorities to relax monetary conditions."³⁷ This is largely due to the fact that "[r]estrictive monetary policy worsens the fiscal position of the government through a reduction in seigniorage revenue, increase in the interest burden of the debt, and lower tax receipts because of the temporary economic slowdown."³⁸ The second argument is based on the premise that "a central bank enjoying economic independence will force the government to reduce the deficit" rather than being "forced to finance government deficits by printing money."³⁹ The third and perhaps most influential line of argument is quite different as it relates to time-inconsistency issues.⁴⁰ Governments, it asserts, have short-term objectives which are frequently inconsistent with longer-term objectives.

A timely example follows: In the event that a government attempts to stimulate the economy through a monetary expansion, the likely future cost is higher inflation.⁴¹ The argument continues by noting that a "[l]ack of credibility in the government commitment to the announced inflation objective leads rational economic agents to demand a risk premium to compensate for this inflationary bias."⁴² After all, it was reasonably assumed that "governments facing a trade-off

³⁴ *Id.*

³⁵ *See id.*

³⁶ Eva Gutiérrez, *Inflation Performance and Constitutional Central Bank Independence: Evidence From Latin America and the Caribbean* 1, 4 (IMF, Working Paper No. 03/53, 2003), available at <https://www.imf.org/external/pubs/ft/wp/2003/wp0353.pdf> (discussion drawing from the work of Eijffinger & de Haan).

³⁷ *Id.*

³⁸ *Id.*

³⁹ *Id.* at 5.

⁴⁰ *Id.*

⁴¹ *Id.*

⁴² *Id.*

between inflation and unemployment are tempted to choose higher-than-optimal inflation rates.”⁴³ The policy response to this is clear: by delegating monetary policy to a conservative central bank, independent of more politically sensitive decision makers, the risk premium can be mitigated through a more credible commitment to future inflation objectives.⁴⁴

Others are far more willing to ascribe less benevolent motives to political actors. Fraser noted that “cynics,” whose skepticism he did not equally share, might also blame the inflation bias on the political process itself.⁴⁵ Such “cynics” would assert that “politicians have short time horizons, stretching only as far as the next election.”⁴⁶ Indeed, Fraser outlined the “treadmill of a political business cycle” as follows:

[W]here interest rates are eased prior to an election to generate strong growth around election time, and the inflation does not surface until after the election. After the election, interest rates are raised, thereby perpetuating a boom/bust cycle. In this story, even politicians who understand that there is no long-term trade-off between inflation and economic activity are drawn irresistibly to the short-term trade-off. Independent and longsighted central bankers are needed to rescue politicians from this temptation.⁴⁷

This view of politicians as “single-minded seekers of reelection” has received a significant amount of attention in political science scholarship.⁴⁸ While Fraser disclaims this view, noting that he is “less cynical about politicians generally,” he nevertheless adds, as others have concluded, that “there is an inherent bias in policy making which pushes policy makers towards a trade-off point which is soft on inflation.”⁴⁹ Explaining this phenomenon, Fraser noted that “policy is determined and implemented under conditions of great uncertainty,” and “[i]t is this uncertainty . . . which can inject an inflationary bias into the policy making process.”⁵⁰ He colorfully notes, “[t]hey may be reluctant to take away the punchbowl just when the party gets going—that unpopular task which the Governor of the U.S. Fed from 1951 to

⁴³ Luis I. Jácome & Francisco Vázquez, *Any Link Between Legal Central Bank Independence and Inflation? Evidence from Latin America and the Caribbean* 1, 3 (IMF, Working Paper No. 05/75, 2005), available at <http://www.imf.org/external/pubs/ft/wp/2005/wp0575.pdf>.

⁴⁴ *Id.*

⁴⁵ See Fraser, *supra* note 24 at 2–3.

⁴⁶ *Id.* at 2.

⁴⁷ *Id.* at 3.

⁴⁸ See, e.g., DAVID R. MAYHEW, CONGRESS: THE ELECTORAL CONNECTION 17 (2d ed. 1974).

⁴⁹ Fraser, *supra* note 24 at 3.

⁵⁰ *Id.*

1970, William McChesney Martin, said was the duty of central bankers.”⁵¹

Interestingly, many scholars have also noted that while an independent central bank must, to some degree, be willing to say “no” to political actors in order to maintain its integrity and freedom, it nevertheless derives its independence from a willing delegation by the people’s representatives.⁵² Indeed, Jeffrey Fuhrer, the former Vice President of the Federal Reserve Bank of Boston, noted that the integral question is not whether there should or should not be *any* political influence on central banking decisions, but what constitutes “undue” political pressure.⁵³ Pointing to prior research, Fuhrer cogently added that “a central bank cannot and should not maintain long-term independence from political pressure if it consistently acts contrary to the will of the electorate.”⁵⁴ This balance is a complex one to achieve and not without significant drawbacks. Fuhrer illustrates the time inconsistency problem through an effective illustration: “The situation is analogous to the lenient parent who, observing that the misbehaving child’s misdeed is done and cannot be undone, decides not to punish the child. The result might be that the child misbehaves more often than the parent would like.”⁵⁵ Perhaps surprisingly, and adding to the complexity of this issue, studies have shown that the “median voter” would prefer a “central banker more inflation averse than himself,” while also preferring to remain “time inconsistent.”⁵⁶

Empirical research on the relationship between independence and price stability has yielded a complex but fairly consistent correlation. Lawrence Summers and Alberto Alesina’s research demonstrated a “strong negative relationship between inflation variability and central bank independence.”⁵⁷ Another study, published by the Federal Reserve Bank of Cleveland identified central bank independence as a “key factor” in

⁵¹ *Id.*

⁵² Gutiérrez, *supra* note 36 at 5; see also generally Sylvester C. W. Eijffinger & Jakob de Haan, *The Political Economy of Central-Bank Independence*, 19 PRINCETON INT’L FIN. SEC. (1996) (referring to a pithy quote from *The Economist* which noted that “[t]he only good central bank is one that can say no to politicians”).

⁵³ Jeffrey C. Fuhrer, *Central Bank Independence and Inflation Targeting: Monetary Policy Paradigms for the Next Millennium?*, NEW ENG. ECON. REV. 19, 20 (Jan./Feb. 1997).

⁵⁴ *Id.*

⁵⁵ *Id.*

⁵⁶ Alberto Alesina & Lawrence H. Summers, *Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence*, 25 J. MONEY, CREDIT & BANKING 151, 151–52 (1993).

⁵⁷ *Id.* at 154. The research, however, also showed that there was no demonstrable correlation between central bank independence and the level of variability of economic growth. *Id.*

maintaining price stability.⁵⁸ Pointing to the case of New Zealand's extraordinary success in curbing inflation in the latter half of the twentieth century, the authors asserted that the passage of the Reserve Bank of New Zealand Act in 1989 was the key turning point in accomplishing this because it incorporated two fundamental reforms: "First, it gave the Reserve Bank greater independence from the central government. Second, it established an explicit inflation target, becoming the first of many countries to do so. Once considered the least independent of central banks, New Zealand's now ranks among the more independent ones."⁵⁹ Extrapolating based on the experience of New Zealand, the authors assert that because other nations followed New Zealand's example, "increased independence is, perhaps, responsible for a decline of nearly two percentage points in the average inflation rate for the industrialized nations as a whole."⁶⁰ The empirical results that the study provides are quite striking and clearly argue in favor of greater independence. Attempting to estimate what prior level inflation would have been had the New Zealand central bank been more independent earlier in the century, the authors noted, "holding all else equal, if New Zealand in the earlier time period had an independence score as high as today's, then the annual inflation rate would have been 3.4% instead of the 7.6% that actually occurred" during the period from 1955–1988.⁶¹ During that earlier period, the Reserve Bank of New Zealand was essentially an arm of the financial ministry, and the country's monetary policy was largely at the finance minister's discretion.⁶²

Taking a broader perspective, the authors also draw the sweeping conclusion "that holding everything else equal, the increase in central banks' independence would have lowered the average inflation rate worldwide from 5.6% to 3.8%."⁶³ While the authors concede that independence was not the only factor at play in these reductions of inflation rates and that the impact of central bank independence is more mercurial in the case of developing nations, they nevertheless conclude that "[a]n independent central bank is a government's most effective way to ensure delivery of a low inflation rate. Furthermore, nearly 2 percentage points of developed countries' average decline in

⁵⁸ Charles T. Carlstrom & Timothy S. Fuerst, *Central Bank Independence: The Key to Price Stability?*, ECON. COMMENT. (Fed. Res. Bank of Clev., Cleveland, Ohio), Sept. 1, 2006.

⁵⁹ *Id.* at 1.

⁶⁰ *Id.*

⁶¹ *Id.* at 2.

⁶² *Id.*

⁶³ *Id.*

inflation over time is the direct results [sic] of their central banks' increased independence."⁶⁴

C. The Arguments Against Central Bank Autonomy

While the arguments, both theoretical and empirical, in favor of central bank autonomy have largely won the day among economists, several factions still stand in stark opposition to what they label the "neoliberal" economic agenda, which they assert incorporates central bank autonomy as one of its central tenets.⁶⁵ That the term "neoliberalism" has attained near-pejorative status in certain Latin American quarters is undeniable. Indeed, to a large degree it has become a rallying cry for aspiring despots, self-styled anti-imperialists, and populists throughout the region.⁶⁶ This backlash has, to some extent, been the result of failed or poorly executed economic reforms put in place by various Latin American states over the last half-century. It has also provided a ready basis upon which populist leaders have manipulated the bleak economic disparities endemic in the region. Few Latin American leaders have been quite as successful at this sort of bluster as Hugo Chavez.

In an intense critique of the "neoliberal" economic order, Dean Timothy Canova pointedly asserts that "[c]entral bank autonomy is a particularly sacrosanct foundation of today's global economic order that undermines peace and stability."⁶⁷ In a line of argument familiar to Latin America observers, he goes on to say, "[i]n reality, central bank autonomy is a euphemism for the capture of monetary policy by interested private banking and financial interests—it is a capture intentionally designed by law."⁶⁸ Indeed, it is based on this perceived oligarchic conspiracy that Canova makes the following dire prediction:

The Neoliberal State is destined to fail in providing peace and prosperity for all precisely because central bank autonomy ensures that monetary and economic policies will favor the few over the many. The empirical evidence bears this out. Real interest rates remain at the highest sustained level of the century, and this remains one of the least recognized causes of today's top heavy distribution of wealth and income both directly, by redistributing income from debtors to creditors, and indirectly, by keeping down the real wage rate in many

⁶⁴ *Id.* at 3.

⁶⁵ See, e.g., Timothy A. Canova, *Financial Liberalization, International Monetary Disorder, and the Neoliberal State*, 15 AM. U. INT'L L. REV. 1279 (2000).

⁶⁶ PLINIO APULEYO MENDOZA, CARLOS ALBERTO MONTANER & ALVARO VARGAS LLOSA, *GUIDE TO THE PERFECT LATIN AMERICAN IDIOT* 66–71 (Michael Lajda Ames trans., 2001).

⁶⁷ See Canova, *supra* note 65, at 1294.

⁶⁸ *Id.*

industries, thereby reinforcing the hierarchies of power that dominate the modern corporation The neoliberal state is one that dashes hopes and dreams on a mass scale, and silently impoverishes the public sector, while delivering great wealth to a small minority of global citizens.⁶⁹

As was mentioned above, however, such arguments are hardly unusual in modern era Latin American politics. Their near-ubiquity is underscored in Plinio Apuleyo Mendoza, Carlos Alberto Montaner and Alvaro Vargas Llosa's groundbreaking study of Latin American populism, *The Guide to the Perfect Latin American Idiot*.⁷⁰

According to the authors, the *perfect idiot* "resorts to the old dichotomy of the left against the right, trusting that a purely semantic or emotional factor can tilt the scale in his favor. His is the old scholastic stratagem of vilifying anyone who questions his dogma. Neoliberalism is an anathema the idiot is trying to lodge into the public conscience through frenzied repetition, as was done in the Middle Ages with heresies."⁷¹ Describing the mindset of their *perfect idiot*, the authors add, "in the subsurface of his feeble political training (a mixture of Marxist Vulgate and populism), has remained the idea that a businessman is an exploiter, getting rich off the labor of others."⁷²

Referring specifically to the *perfect idiot*'s view of monetary policy and judgment, Mendoza, Montaner and Llosa note that the *idiot* "believes that [monetary issuance as a way of reactivating demand and making up for the shortage of resources] is also a means not only of development but of what he defines, with grandiloquent pomposity, as social investment. He's a friend of the great money-printing machine. And he considers the policies dedicated to assuring sound money as reactionary, neoliberal, and contrary to popular interests."⁷³ As we shall see, *infra*, these are the very goals that Mr. Chavez outlined for his "reformed" central bank.⁷⁴ Indeed, the authors note, the *idiot*'s policies have been "catastrophic"⁷⁵ for the region. They add, "[i]n this case, the fox in the henhouse is not the businessman but the state, which is plucking the chickens mercilessly."⁷⁶

⁶⁹ *Id.* at 1301–04.

⁷⁰ *See generally* MENDOZA, MONTANER & VARGAS LLOSA, *supra* note 66.

⁷¹ *Id.* at 67.

⁷² *Id.* at 68–69.

⁷³ *Id.* at 69.

⁷⁴ *See infra* Part III.B.

⁷⁵ *See* MENDOZA, MONTANER & VARGAS LLOSA, *supra* note 66, at 69.

⁷⁶ *Id.*

Along these lines, Milton Friedman offered the following scathing critique of the populist view of neoliberal reforms, especially the protection of central bank independence from political interference:

No government willingly accepts the responsibility for producing inflation even in moderate degree, let alone at hyperinflationary rates. Government officials always find some excuse—greedy businessmen, grasping trade unions, spendthrift consumers, Arab sheikhs, bad weather, or anything else that seems remotely plausible. No doubt businessmen are greedy, trade unions are grasping, consumers are spendthrifts, Arab sheikhs have raised the price of oil, and the weather is often bad. Any of these can produce high prices for individual items; they cannot produce rising prices for goods in general. . . . [T]hey cannot produce continuing inflation, for a very simple reason: not one of the alleged culprits possesses a printing press on which it can legally turn out those pieces of paper we carry in our pockets and call money.⁷⁷

While later scholars such as Summers and Alesina asserted that this broad confidence in the power of the monetary was somewhat misplaced,⁷⁸ the essential truth behind Friedman's assertion is undeniable: none of the conspirators that populist demagogues would blame for all their countries' ills can actually print their own money, of which engaging in excessively, whether it be in the context of military conflict or economic upheaval, is most often the true cause of massive, debilitating inflation.

Indeed, given the collective Latin American experience over the last half-century, what is perhaps *most* striking about the sorts of arguments advanced by Professor Canova is their dogged persistence. It is, in part, this challenge that caused Fed Chairman Ben Bernanke to strike a more cautious note when he stated the following about the recent progress in Latin America:

Worldviews tend to evolve slowly, and it is far too soon to claim that ideas and politics in Latin America have changed fundamentally or irrevocably. Nevertheless, during the past fifteen years or so, both the political elites and the broader public in the region have, for the most part, shown an increased willingness to abandon the failed economic models of earlier decades in favor of new approaches that emphasize improved fiscal discipline, the benefits of free trade and free markets, and the vital importance of building strong economic institutions.⁷⁹

There can be little doubt that Bernanke's caution was amply justified and to a large degree the economic policies advanced by Mr. Chavez and his allies in the region (including Bolivian

⁷⁷ See FRIEDMAN, *supra* note 15, at 192.

⁷⁸ See generally Alesina & Summers, *supra* note 56.

⁷⁹ See Bernanke, *supra* note 30.

President Evo Morales, Ecuadorian President Rafael Correa, and Nicaraguan President Daniel Ortega) underscore the lack of consensus around free-market oriented policies, including central bank autonomy.

D. Paradigms of Analysis

In order to more effectively study the issue of central bank independence, scholars have divided the topic into several categories that can be separated from one another to allow for better comparison and analysis. Understanding these categories allows for a more precise definition of specifically what is meant by the term *independence* in the legal context. To a large degree, much of the modern analysis of central bank autonomy has accepted the fact that not all independence, as the expression goes, is equal.⁸⁰ Even the most perfect legal structure can fail utterly in delivering the policy objectives it was crafted to achieve. Perhaps no region has stood in testament to this fact to a greater degree than Latin America. Thus, merely considering formal legal structures in order to determine the extent of a central bank's independent judgment is sorely inadequate. In order to recognize this fundamental distinction, economists have divided the broader concept of central bank independence into two categories: *de jure* and *de facto* independence.⁸¹

Underscoring this distinction, as we shall see in the subsequent section, central banks need not operate within the high walls of a complex legal framework in order to be substantively independent. Beyond the more intuitive distinction between *de jure* and *de facto* independence, scholars have also divided the scope of a central bank's actions into two categories: goal and instrument independence.⁸² This distinction addresses the scope of the bank's responsibilities. Though one can relate it, in a broader sense to the *de jure* and *de facto* distinction, it ultimately rests on the explicit authority vested in the institution itself. As we shall note, the goals of a national central bank or monetary authority are frequently outlined within the authorizing laws of that institution. That, for example, a central bank is guided by elected officials to dedicate itself to the maintenance of a low rate of inflation, or perhaps conversely, a lower rate of unemployment, does not mean that the political actors may affect the actual decision making conducted by the institution in order to achieve that end. The

⁸⁰ See, e.g., Jácome, *supra* note 9.

⁸¹ See *infra* Part I.D.1 for a detailed look at *de jure* and *de facto* independence.

⁸² See *infra* Part I.D.2 for a detailed look at goal and instrument independence.

following sections shall provide a more detailed look at these two sets of distinctions.

1. *De Jure* vs. *De Facto* Independence

The distinction between *de jure* and *de facto* independence essentially rests upon the recognition that “central banks are institutions that operate within government and not apart from it . . . any attempt at legislating autonomy can never entirely eliminate all potential sources of tension between the political and monetary authorities.”⁸³ The former Governor of the Reserve Bank of Australia, B.W. Fraser, notes that in order to account for the full complexities of central bank independence, studies must account for this distinction.⁸⁴ Citing to his experience in Australia, Fraser asserts:

[L]egal independence, which is what the empirical studies seek to measure, can be very different from practical independence. In Australia, the legislation provides that in the event of a dispute over monetary policy, the government can override the Reserve Bank by tabling its objections before both houses of parliament. While this is the legal position (which scores a negative on the index of independence), such a situation has never arisen in practice; and indexes of legal independence cannot capture the role of personalities (such as the Governor and the Treasurer/Finance Minister) who fill in the operational gaps in the relevant legislation. Nor can they capture the changing policy-making environment. For example, the greater reliance on market-based policies (rather than controls) in Australia over the past decade has significantly enhanced the degree of independence of the Reserve Bank, without any change in the Bank’s charter.⁸⁵

Fraser’s two examples underscore an important distinction that has also arisen in Latin America. Very often, as in the case of Brazil, a central bank may be able to act with a greater degree of independence than otherwise provided for by the bank’s founding legislation.⁸⁶ On the other hand, however, there are also clear instances, for example in Venezuela, where a bank lacks the real, political authority to perform its responsibilities as specifically outlined in such legislation.⁸⁷ Luis Jácome provides a more detailed explanation of the Brazil/Venezuela comparison:

[T]he Central Bank of Brazil (CBB) enjoys more effective than legal independence as oppose [sic] to the Central Bank of Venezuela (CBV),

⁸³ PIERRE L. SIKLOS, THE CHANGING FACE OF CENTRAL BANKING: EVOLUTIONARY TRENDS SINCE WORLD WAR II 301 (2002).

⁸⁴ Fraser, *supra* note 24, at 3.

⁸⁵ *Id.* at 5.

⁸⁶ See Jácome, *supra* note 9, at 20.

⁸⁷ See *id.*

which on many aspects is legally rather than effectively independent. Regardless of the lack of legal clarity of the policy objective of the CBB and of having no political independence—the government maintains a majority in the National Monetary Council (presided by the Minister of Finance) and the Executive branch carries legal support to directly remove the President of the CBB—the government has granted the CBB effective independence, which has been articulated through the inflation targeting scheme in practice since 1999. On the other hand, the CBV has been the subject of episodes of political interference that have limited its legally assigned economic independence. In particular, the government imposed capital controls in the midst of the banking crisis and obtained financing for the deposit insurance fund, which adversely affected the conduct of monetary policy. In addition, financial independence established by law has not been followed in practice . . . while simultaneously the CBV has been required to transfer to the government unrealized profits.⁸⁸

To a large degree, Jácome's comparison of the Brazilian and Venezuelan experiences highlights the difficulty in capturing the true extent to which a central bank is independent. Indeed, what is all the more troubling about his example is that the degree to which political actors have interfered with monetary policy-making since his study has only gotten worse in the case of Venezuela and, some would argue, in Brazil as well.⁸⁹

2. Goal Independence vs. Instrument Independence

In their article, *How Independent Should a Central Bank Be?*, DeBelle and Fischer noted that “it is useful to draw a distinction between *goal independence* and *instrument independence*.”⁹⁰ Goal independence, which the authors relate to Grilli, Masciandaro, and Tabellini's earlier concept of political independence, exists when the central bank is free to set the final goals of monetary policy.⁹¹ Along these lines, a central bank would be acting independently in this regard if it could, “for instance, decide that price stability was less important than output stability and act accordingly.”⁹² On the other hand, the authors assert, “[a] bank that has instrument independence is

⁸⁸ See *id.* at 20–22.

⁸⁹ See, e.g., Iuri Dantas, *Banco do Brasil Insists on Independence; Cheap Loans*, BLOOMBERG.COM, Apr. 28, 2009, http://www.bloomberg.com/apps/news?pid=20601086&sid=axCOPPNzm1o8&refer=latin_america; John Fitzpatrick, *Brazil: Lula Starts to Throw His Weight Around*, GRINGOES.COM, Apr. 15, 2009, http://www.gringoes.com/articles.asp?ID_Noticia=2272.

⁹⁰ Guy DeBelle & Stanley Fischer, *How Independent Should a Central Bank Be?*, in GOALS, GUIDELINES, AND CONSTRAINTS FACING MONETARY POLICYMAKERS 197 (Jeffrey C. Fuhrer ed., 1994).

⁹¹ *Id.*

⁹² *Id.*

free to choose the means by which it seeks to achieve its goals.”⁹³ By way of a previous example, which Debelle and Fischer also cite, the Reserve Bank of New Zealand, “whose goals are precisely described in a contract with the government, has no goal independence; however, it has instrument independence since it chooses the method by which it tries to achieve the pre-assigned goals.”⁹⁴

Jeffrey Fuhrer of the Federal Reserve Bank of Boston briefly outlines the two concepts, noting that “[g]oal independence allows the central bank to set its own ultimate goals (price stability, stable employment, and so on) and instrument independence allows the central bank to decide how best to achieve those goals.”⁹⁵ Professor Carl Walsh provides a more expansive summary of this distinction in *The New Palgrave Dictionary of Economics*:

Goal independence refers to the central bank’s ability to determine the goals of policy without the direct influence of the fiscal authority. In the United Kingdom, the Bank of England lacks goal independence since its inflation target is set by the government. In the United States, the Federal Reserve’s goals are set in its legal charter, but these goals are described in vague terms (for example, maximum employment), leaving it to the Fed to translate these into operational goals. Thus, the Fed has a high level of goal independence. Price stability is mandated as the goal of the European Central Bank (ECB), but the ECB can choose how to interpret this goal in terms of a specific price index and definition of price stability.

Instrument independence refers only to the central bank’s ability to freely adjust its policy tools in pursuit of the goals of monetary policy. The Bank of England, while lacking goal independence, has instrument independence; given the inflation goal mandated by the government, it is able to set its instruments without influence from the government. Similarly, the inflation target range for the Reserve Bank of New Zealand is set in its Policy Targets Agreement (PTA) with the government, but, given the PTA, the Reserve Bank has the authority to sets [sic] its instruments without interference. The Federal Reserve and the ECB have complete instrument independence.⁹⁶

Understanding this distinction is necessary to fully grasp the extent of a monetary authority’s independence. Indeed, to a

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ Jeffrey Fuhrer, *Central Bank Independence and Inflation Targeting: Monetary Policy Paradigms for the Next Millennium?*, NEW ENG. ECON. REV. 19, 22 (Jan.–Feb. 1997).

⁹⁶ Carl E. Walsh, *Central Bank Independence*, in 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS 729 (Steven N. Durlauf & Lawrence E. Blume eds., 2d ed. 2008).

large degree it is more susceptible to legal analysis than *de jure* and *de facto* independence to the extent that it can be more readily determined by considering the legal framework and founding documents of the central bank as an institution. The murkiness of determining the extent of *de jure* and *de facto* independence presents a variety of challenges that go far beyond the scope of legal and statutory analysis. However, this is not to say that determining the extent of goal independence, for example, is without significant complexity and ambiguity.

Eggertsson and Le Borgne explored this topic at length in their study, *A Political Agency Theory of Central Bank Independence*,⁹⁷ and though they drew strikingly novel conclusions, they did not alter the basic distinction between these two forms of independence. Specifically, the authors focused on the following question: “Why would an office-motivated politician spontaneously decide to delegate his monetary policy prerogative to an independent but accountable central bank?”⁹⁸ The great novelty of their study, though, is in their refusal to rely on the inflation bias to support their conclusion of lower interest rates, noting that “all agents have the same (imperfect) information.”⁹⁹ Indeed, they propose that a politician’s delegation of monetary policy to an independent central banker with certain minimum assurance of job security, including job tenure, “enables the central banker to commit more effort into the conduct of monetary policy than an elected politician could ever afford to. This extra effort translates in expectations, in better forecasts and fewer policy mistakes, which increases social welfare.”¹⁰⁰ Thus, they assert, it is in a politician’s “private” interest to make such a delegation that is both instrument and goal independent.¹⁰¹

However, the authors conclude that while “instrument independence is welfare enhancing for society, goal independence is not” as it “benefits the current majority of the population at the expense of the future majority.”¹⁰² Considering the case of the United States Federal Reserve, the authors note that the Fed combines both instrument and goal independence. They continue that the Board is clearly instrument independent and, given that its “mandate is cast in very broad terms and covers conflicting

⁹⁷ Gauti Eggertsson & Eric Le Borgne, *A Political Agency Theory of Central Bank Independence* (IMF, Working Paper No. 03/144, 2003), available at <http://www.imf.org/external/pubs/ft/wp/2003/wp03144.pdf>.

⁹⁸ *Id.* at 38.

⁹⁹ *Id.*

¹⁰⁰ *Id.*

¹⁰¹ *Id.*

¹⁰² *Id.* at 39.

tasks (to pursue [sic] both ‘maximum employment’ and ‘stable prices’),” its governing body interprets and assigns relative weight among these objectives: *i.e.*, it is goal independent.¹⁰³

By using the example of the Federal Reserve, Eggertsson and Le Borgne reveal the essence of the distinction. The Federal Reserve is instrument independent in that it has complete discretion over when and to what extent to make open market operations and other decisions. Further, however, the authors note that the malleability of the Federal Reserve’s mandate is so broad that it can also be said to have goal independence.

III. CASE STUDIES

The following case studies are intended to illustrate several of the factors described above. While not intended to be a comprehensive analysis of the central banking structures implemented in the analyzed countries, they are intended to give the reader an overview of the central bank policies that have been adopted in the region in order to get a sense of the larger trends that have been in play in Latin America. Specifically, as noted above, I will focus on specific constitutional or legislative measures that outline the scope of the central banks’ *de jure* sovereignty. In particular, I will focus on the distinction between instrument and goal independence in analyzing the various national central banking frameworks and the historical contexts in which they have operated.¹⁰⁴

A. Argentina

The *Banco Central de la Republica Argentina* (“BCRA”) was founded in 1935.¹⁰⁵ The new institution replaced the existing currency board system, which had been in place in 1899.¹⁰⁶ Article 3 of the bank’s originating Charter Act No. 24,144, provides that the bank “shall primarily and essentially preserve the value of currency,” and continues by noting that “[t]he authorities of the [BCRA] shall involve, to this effect, the regulation of the amount of money and lending within the

¹⁰³ *Id.* at 6.

¹⁰⁴ To this end, Luis Jácome’s breakdown of Central Bank Autonomy in Annex II of his paper, *supra* note 9, was very helpful.

¹⁰⁵ Banco Central de la Republica Argentina, *Former Governors and the Executive Branch*, http://www.bcr.gov.ar/index_i.htm (follow “Institutional” hyperlink; then follow “History and Heritage” hyperlink; then follow “Former Governors and the Executive Branch” hyperlink) (last visited Nov. 1, 2009).

¹⁰⁶ Banco Central de la Republica Argentina, *The BCRA and its History*, http://www.bcr.gov.ar/index_i.htm (follow “Institutional” hyperlink; then follow “History and Heritage” hyperlink; then follow “The BCRA and its history” hyperlink) (last visited Nov. 1, 2009).

economic system and the formulation of monetary, financial and exchange rules under the legislation in force.”¹⁰⁷ Indeed, Article 3 goes on to explicitly announce that the bank shall act independently of the Argentine President, noting, “[a]s regards the preparation and implementation of a monetary and financial policy, the [BCRA] shall not be subject to any order, recommendation or instruction given by the National Executive Power.”¹⁰⁸

Chapter III of Charter Act No. 24,144 specifically addresses issues relating to the tenure and compensation of members of the Board of Directors. Article 7 provides that “[t]he president, vice president and directors shall be appointed by the National Executive Power in agreement with the Senate of the Nation; they shall hold office for six (6) years and may be reappointed.”¹⁰⁹ Further, the President may appoint members of the Board on an interim basis awaiting the consent of the Argentine Senate.¹¹⁰ As to the compensation paid to members of the Board, the Charter Act provides that this shall be left to the Board, noting, “[t]he fees of the president, vice president and directors shall be determined according to the [BCRA] budget.”¹¹¹ Though the budget must be published at regular intervals and members of the Board can be punished for failing to meet this announcement requirement, the Board is left with the discretion of how to control the BCRA’s budget.¹¹²

Finally, Article 9 of the Charter Act provides for the process under which members of the Board may be removed.¹¹³ The process is specifically outlined as follows:

The National Executive Power shall decree, after consultation with a committee set up by the Honourable National Congress, the dismissal of any member of the Board of directors on account of misconduct or non-compliance with his duties as public servant. The Committee shall be chaired by the President of the Senate and shall be composed of both the president of the Budget and Treasury Committee and the president of the Economy Committee of the Senate and of both the president of the Budget and Treasury Committee and the president of the Finance Committee of the Lower House.¹¹⁴

¹⁰⁷ Carta Orgánica del Banco de la República Argentina, Ley 24.144, art. 3 (1992), available at http://www.bcra.gov.ar/pdfs/marco/Carta_Organica.PDF (Spanish version).

¹⁰⁸ *Id.*

¹⁰⁹ *Id.* at art. 7.

¹¹⁰ *Id.*

¹¹¹ *Id.*

¹¹² *Id.* at art. 3.

¹¹³ *Id.* at art. 9.

¹¹⁴ *Id.*

The removal power retained by the Argentine Senate and President would seem to be quite broad, given that it merely takes a decree issued by the executive and approval by the Senate. Further, the specific scope of what constitutes a removable offense was left extremely broad. However, there is not much evidence of abuse in the past. Even during the tumultuous economic collapse experienced by Argentina in 2001, only one member of the Board, Dr. Pedro Pou, was removed.¹¹⁵

B. Chile

The *Banco Central de Chile* (“BCC”) was founded in 1925, eight years prior to the BCRA, on the basis of recommendations made by a commission headed by Edwin Walter Kemmerer, professor of economics at the University of Princeton.¹¹⁶ The projects proposed by the Kemmerer Commission included the monetary law “which sought to stabilize the value of Chile’s currency and the gold standard as the basis of the country’s monetary unit, the creation of the Central Bank of Chile, the passing of the general banking law and the general budget law.”¹¹⁷

Title I, Section 3 of Law 18,840, titled the Basic Constitutional Act of the Central Bank of Chile (“Basic Constitutional Act”), defines the scope of the BCC’s objectives.¹¹⁸ Similar to the BCRA, the Basic Constitutional Act provides that “[t]he Bank shall have as its purposes to look after the stability of the currency.”¹¹⁹ However, in addition to the basic anti-inflationary authority that the BCC shares with the BCRA, the BCRA also has the responsibility of overseeing “the normal functioning of internal and external payments.”¹²⁰ The Basic Constitutional Act also provides that, in striving to achieve these objectives, “[t]he authority of the Bank . . . shall include that of regulating the amount of currency and credit in circulation, the performance of credit transactions and foreign exchange, as well as the issuance of regulatory provisions regarding monetary, credit, financing and foreign exchange matters.”¹²¹

¹¹⁵ David Fuhr & Zachary Klughaupt, *The IMF and AGOA: A Comparative Analysis of Conditionality*, 14 DUKE J. COMP. INT’L L. 125, 130 (2004).

¹¹⁶ Banco Central de Chile, Origin of the Central Bank of Chile, <http://www.bcentral.cl/eng/about/functions/index.htm> (last visited Nov. 1, 2009).

¹¹⁷ *Id.*

¹¹⁸ Ley Orgánica Constitucional del Banco Central de Chile, *as amended*, Diario Oficial, art. 3, sec. 3, 5 de enero de 2006 (Chile), available at <http://www.bcentral.cl/publicaciones/normativas/pdf/leyorganica102006.pdf>.

¹¹⁹ *Id.*

¹²⁰ *Id.*

¹²¹ *Id.*

In terms of the degree of oversight by the Legislative and Executive branches of Chile, the Basic Constitutional Act requires that the BCC “inform” the political branches of a range of decisions.¹²² In particular, Section 4 provides the following:

The Bank shall inform the President of the Republic and the Senate with regard to the policies and regulations of general applicability issued by the Bank in the performance of its duties. Likewise, it shall advise the President of the Republic, upon his request, in all matters within the scope of its duties.¹²³

Interestingly, under Section 6 of the Basic Constitutional Act “the supervision and administration of the Bank shall be vested in the Board of the Central Bank, which shall be responsible for the exercise of the authority and the carrying out of the duties entrusted to the Bank by law.”¹²⁴ However, Section 6 continues with a seemingly important caveat: “When passing its resolutions, the Board shall consider the general orientation of the Government’s economic policies.”¹²⁵ Depending on the scope of its application, this caveat would seemingly limit the extent of the bank’s independence, especially when compared to the more absolute bar on executive interference provided in the Argentine case.

The appointment of members of the Board, governed by Section 7 of the Basic Constitutional Act, occurs with “the prior approval of the Senate, by means of an executive decree issued through the Ministry of Finance.”¹²⁶ Unlike the Argentine case, the Chilean Basic Constitution does not provide for interim appointments, through which the Executive could, though temporarily, expand his or her influence on the Board. Section 8 provides that members’ terms shall last 10 years, and are renewable.¹²⁷ The Chairman of the Bank, whose term as Chair lasts five years (or the remainder of his term on the Board, if shorter) is appointed at the sole discretion of the Chilean President.¹²⁸ Another notable difference with the BCRA’s structure appears in Section 10, and relates to compensation of board members.¹²⁹ Whereas in the Argentine case, the Board of the BCRA establishes its own compensation levels, which it must publish, the Basic Constitutional Act provides the following:

¹²² *Id.* at sec. 4.

¹²³ *Id.*

¹²⁴ *Id.* at sec. 6.

¹²⁵ *Id.*

¹²⁶ *Id.* at sec. 7.

¹²⁷ *Id.* at sec. 8.

¹²⁸ *Id.*

¹²⁹ *Id.* at sec. 10.

Compensation for the Governor, the Vice-Governor and the Board Members shall be set forth by the President of the Republic for periods not to exceed two years. For such purposes, the President of the Republic shall appoint, with the anticipation as may be necessary, a Committee formed by three persons who are former Governors and Vice-Governors of the Bank, who shall present a compensation proposal based upon the salaries being paid by the private sector banking institutions to their top executives.¹³⁰

Though the Basic Constitutional Act leaves a wide degree of discretion in the hands of the Chilean President in terms of the compensation of board members, it simultaneously attempts to make the process somewhat more objective through the requirement that the President institute a commission to make compensation proposals.

As opposed to the Argentine case, however, the BCC's requirements for the dismissal of board members is somewhat, though marginally, more stringent. Section 17 of the Basic Constitutional Act provides that "[t]he President of the Republic by justified reason and the prior approval of the Senate may remove any or all of the Board Members."¹³¹ The requirement that the President's actions be justified is more clearly defined subsequently:

The dismissal may only be based upon the grounds that the concerned Board Member has voted in favor of Bank resolutions representing a material and clear breach of the purposes of the Bank as specified in the first paragraph of Section 3, and provided that said resolution has been the main and direct cause of a material damage to the economy of the country.¹³²

Clearly, the primary distinction between this and the Argentine case is the high bar set for the seriousness of the offenses that the member of the board must be charged with by the President. Additionally, and interestingly not specified in the Argentine case, the Chilean Basic Constitutional Act also requires that "[t]he Board Member or Members concerned shall have the right of a hearing before the Senate."¹³³ The seemingly higher and more specific requirements for the alleged act, in addition to the right of a hearing before the Senate would seem to make the task of removing a member from the Board somewhat more challenging for Chilean political officers.

¹³⁰ *Id.*

¹³¹ *Id.* at sec. 17.

¹³² *Id.*

¹³³ *Id.*

C. Mexico

The *Banco de Mexico*, Mexico's central bank, was also founded in 1925, and finds its underpinnings in the Mexican Constitution.¹³⁴ It was not until 1931 that the *Ley Monetaria de los Estados Unidos de Mexicanos* was passed which provided much greater guidance in the actual operations of the bank, rather than the general terms of the Mexican Constitution.¹³⁵ Nevertheless, Article 28 of the Constitution specifies many, if not most, of the basic governing principles for the Bank, including its central objective and several limitations.¹³⁶ First, Article 28 begins with an explicit call for the creation of the central bank and the requirement that it be independent: "The State shall have a central bank, which shall be autonomous in exercising its function and management."¹³⁷ Later in Article 28, there appears the specific proviso that "No authority shall order the central bank to grant financing."¹³⁸

Article 28 then defines the scope of the *Banco de Mexico's* objective, noting that it "will be to foster the stability of the national currency's purchasing power, therefore strengthening the State's role in guiding the country's development."¹³⁹ Comparing this provision of Article 28 to the previous examples we have considered can be quite revealing. For example, neither the Argentine nor the Chilean central banks are specifically given the additional objective of "guiding the country's development."¹⁴⁰ Indeed, in both of the earlier cases, the authorizing legislation focused on the objective of price stability. Although, based on the text of the Article, the "development" objective would seem to be secondary to the primary objective of price stability, its inclusion is nevertheless notable. One might draw a comparison to the vague terms used to outline the objectives of the United States Federal Reserve. It was noted earlier that these vague terms, and the additional discretion that they imparted on the Federal Reserve, have enhanced its

¹³⁴ Banco de Mexico, *History of the Mexican Banknote*, <http://www.banxico.org.mx/sitioingles/billetesymonedas/didactico/notesManufactFeaturesHistory/historyMexicanNotes.html> (last visited, Nov. 1, 2009).

¹³⁵ *Ley Monetaria de los Estados Unidos de Mexicanos*, Diario Oficial de la Federación [D.O.], 27 de julio de 1931 (Mex.), available at <http://www.diputados.gob.mx/LeyesBiblio/pdf/152.pdf>.

¹³⁶ Constitución Política de los Estados Unidos Mexicanos [Const.], as amended, Diario Oficial de la Federación [D.O.], art. 28, 5 de febrero de 1917 (Mex.) [hereinafter Constitución Política de los Estados Unidos Mexicanos].

¹³⁷ *Id.*

¹³⁸ *Id.*

¹³⁹ *Id.*

¹⁴⁰ For a discussion on the banking policies of Argentina see *supra* Part III. A. For a discussion on the banking policies of Chile see *supra* Part III. B.

independence from political actors by increasing its goal of independence.¹⁴¹ While the terms of the Mexican Constitution would seem not to be subject to as broad an interpretation, they nevertheless would presumably allow for a broader range of action than either the Argentine or Chilean cases.

As is the case in both Chile and Argentina, the members of the bank's board are appointed by the President, subject to confirmation "by the Senate or, should it be the case, by the Permanent Commission."¹⁴² Addressing the tenure of board members, Article 28 provides that "[t]hey shall discharge their office by periods, in line with the autonomous exercise of their functions."¹⁴³ On this point, Article 40 of the *Ley de Banco de Mexico* provides much clarification and specificity:

The Governor will be appointed for a term of six years and Deputy Governors for eight years. The term of the Governor will start on the first of January of the fourth year of the President of the Republic's term. The terms of the Deputy Governors will be staggered, with one starting every other year on the first of January of the first, third and fifth years of the President's term. The individuals occupying these posts may be appointed members of the Board of Governors for more than one term.¹⁴⁴

Thus, although the Governor of the *Banco de Mexico* is appointed from among the existing members of the board at the sole discretion of the Mexican President, the terms of the board's members are staggered to end every other year throughout the President's term. As the President of Mexico is limited to a single 6-year term, he or she would be limited to appointing no more than three members of the 5-member board, barring retirements.

Interestingly, the Mexican Constitution and *Ley de Banco de Mexico* establish a very high bar for the removal of members of the board. The Constitution provides the following:

They shall only be removed by grave causes and they cannot hold any other position or job, except those in which they act on behalf of the bank or those honorary related with teaching, scientific, cultural, or charity associations. The people in charge of the direction of the central bank may be subject to impeachment in accordance with the provisions set forth in Article 110 of this Constitution.¹⁴⁵

¹⁴¹ See *supra* note 103 and accompanying text.

¹⁴² Constitución Política de los Estados Unidos Mexicanos, *supra* note 136, at art. 28.

¹⁴³ *Id.*

¹⁴⁴ Ley de Banco de Mexico, *as amended*, Diario Oficial de la Federación [D.O.], art. 40, 23 de diciembre de 1993 (Mex.) [hereinafter *Ley de Banco de Mexico*].

¹⁴⁵ Constitución Política de los Estados Unidos Mexicanos, *supra* note 136, at art. 28.

Article 43 of the *Ley de Banco de Mexico*, provides the following list of actions which would be sufficient to justify removal:

- I. Mental illness, as well as physical affliction preventing the adequate fulfillment of their functions for more than six months;
- II. Performing any job, position or commission other than those provided for in Article 28 Paragraph Seven of the Constitution;
- III. Stop being a Mexican citizen or fulfilling any of the requirements set forth under Article 39 Section III;
- IV. Failure to observe the agreements sanctioned by the Board of Governors or deliberately acting in excess or in dereliction of his duties;
- V. Using the confidential information available to him or her as a result of his or her position in his or her own benefit or that of third parties, or disclosing said information without the consent of the Board of Governors;
- VI. Knowingly submitting false information for the consideration of the Board of Governors, and
- VII. Leaving his or her post without authorization from the Board of Governors or without justified motive, barring force majeure. The Board of Governors cannot authorize leave of absence for more than six months.¹⁴⁶

It should be noted that this list of possible offenses is far more specific than any of the others, and would seemingly make removal of a board member more difficult. This is especially true, however, when one considers the complexity of the process which must also be undergone in order to facilitate this removal. This process, outlined beginning in Article 44 of the *Ley de Banco de Mexico*, involves several steps before even the impeachment stage may be reached. First, a majority of the Board, excluding the member in question, must come to a verdict that at least one of the conditions outlined in Article 43 have been met.¹⁴⁷ The accused member must be accorded a hearing in this process.¹⁴⁸ Once such a verdict has been issued, it must be sent to the President, along with “the documentary evidence supporting it and the written argumentation the affected party may have presented in his or her own defense.”¹⁴⁹ Article 44 then requires that “[t]he Federal Executive shall in turn send it, together with the aforementioned documentation and the President’s considerations on the admissibility or inadmissibility of the removal, to the Senate or, should the latter not be in session, to

¹⁴⁶ Ley de Banco de Mexico, *supra* note 144, at art. 43.

¹⁴⁷ *Id.* at art. 44.

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

the Permanent Commission, for a final judgment.”¹⁵⁰ This final judgment must follow the ordinary, more rigorous (though arguably not as rigorous as in the United States), process for impeachment provided for in the Constitution.¹⁵¹

III. THE CASE OF VENEZUELA

A. The *Banco Central de Venezuela*

First established in 1939, the last of the central bank institutions that we will discuss in greater detail, the *Banco Central de Venezuela* (“BCV”), was created in the wake of the twenty-seven year dictatorship of Juan Vicente Gomez.¹⁵² As part of the general upheaval of that era in Venezuelan history, the new government adopted a novel national economic plan called the February Program, which implemented a wide range of economic reforms, including the foundation of the Central Bank.¹⁵³ The first major reforms to this new system were put in place in the 1960s and were followed by additional reforms in the 1990s, the latter of which gave the BCV administrative autonomy.¹⁵⁴ The most recent reforms to the BCV were implemented in 2001.¹⁵⁵

The primary legal framework for the BCV is derived from the Venezuelan Constitution and *Ley del Banco Central de Venezuela*.¹⁵⁶ The Venezuelan Constitution, passed in 2001 with the sponsorship of Hugo Chavez, only briefly addresses the establishment of a central bank, though it does note that “the monetary competence of National Power shall necessarily be exercised exclusively by the Venezuelan Central Bank.”¹⁵⁷ Though this would appear to be a fairly comprehensive statement of the scope of the BCV’s power, several caveats are subsequently added. These provisos are largely veiled references to various aspects of Mr. Chavez’s socialist Bolivarian agenda. For example, while it is repeated later in Article 318 that the BCV “is a public-law juridical person with autonomy to formulate and implement policies within its sphere of competence,” the

¹⁵⁰ *Id.*

¹⁵¹ Constitución Política de los Estados Unidos Mexicanos, *supra* note 136, at art. 28.

¹⁵² Banco Central de Venezuela, *History of the BCV*, <http://www.bcv.org.ve/EnglishVersion/c3/index.asp?secc=history> (last visited Nov. 1, 2009).

¹⁵³ *Id.*

¹⁵⁴ *Id.*

¹⁵⁵ *Id.*

¹⁵⁶ Ley del Banco Central de Venezuela [Law on the Central Bank of Venezuela], *as amended*, Publicada en Gaceta Oficial, 3 de octubre de 2001(Ven.) [hereinafter *Ley del Banco Central de Venezuela*].

¹⁵⁷ Constitución de la República Bolivariana de Venezuela, [Const.], *as amended*, Publicada en Gaceta Oficial, art. 318, 30 de Diciembre de 1999 (Ven.).

Constitution includes the additional requirement that the bank “perform its functions in coordination with general economic policy, in the interest of attaining the higher objectives of the State and the Nation.”¹⁵⁸ What exactly would be included among these “higher objectives” is, not surprisingly, left deeply unclear, though one might fairly assume that it is a reference to Mr. Chavez’s “Bolivarian” Revolution. Interestingly, and undoubtedly in concession to the monetary policy implications of his regional agenda, Article 318 includes the following, highly unusual caveat: “In the event a common currency is instituted within the framework of Latin American and Caribbean integration, it shall be permissible to adopt the currency provided for by a treaty signed by the Republic.”¹⁵⁹ Mr. Chavez’s populism frequently incorporates a marked emphasis on regional cooperation against the perceived hegemony of the United States, which he labels “imperial.”¹⁶⁰

As to the responsibilities of the BCV and scope of its purpose, Article 318 provides that “[t]he fundamental objective of the Venezuelan Central Bank is to achieve price stability and preserve the internal and foreign exchange value of the monetary unit.”¹⁶¹ In order to achieve these objectives, Article 318 continues, the functions of the BCV “shall include those of formulating and implementing monetary policy, participating in the design of and implementing foreign exchange policy, currency regulation, credit and interest rate, administering international reserves and any others established by law.”¹⁶²

As to the selection of the board of the BCV, Article 16 of the *Ley del Banco Central de Venezuela* provides that the Executive “shall be responsible for the designation of the President and four (4) Directors of the Bank.”¹⁶³ Only the nomination of the President must be approved by the Venezuelan Senate, not the nomination of the other four directors.¹⁶⁴ Further, one of the President’s nominees must be a “Minister of the economic cabinet,” though it cannot be the Finance Minister.¹⁶⁵ In addition, the National Assembly must designate the two (2) remaining Directors by a majority vote.¹⁶⁶ In order to remove a member from the Board, similar to the process provided for in

¹⁵⁸ *Id.*

¹⁵⁹ *Id.*

¹⁶⁰ *Watch Venezuela*, WASH. POST, Nov. 24, 2008, at A18.

¹⁶¹ *Id.*

¹⁶² *Id.*

¹⁶³ *Ley del Banco Central de Venezuela*, *supra* note 156, at art. 16.

¹⁶⁴ *Id.* at art. 9.

¹⁶⁵ *Id.* at art. 15.

¹⁶⁶ *Id.* at art. 16.

Mexico, either the President or two members of the Board must initiate a removal proceeding.¹⁶⁷ However, in Venezuela, a two-thirds majority of the National Assembly of Representatives is required to remove the Board member.¹⁶⁸

B. The Terms of the Amendment and its Implications

As noted above, the *Reforma* proposed by Hugo Chavez in 2007 would have caused a substantial change in way the nation conducted its monetary policy. While many observers had questioned the *de facto* independence of the BCV, one can only wonder what the complete abandonment of nearly all *de jure* standards of independence would have entailed. Indeed, the proposed amendments did not even seek to maintain the appearance of independence for the BCV; they sought to make the institution utterly subservient to the national executive.

From its very opening, the proposed revision to Article 318 required that the “national monetary system must incline to the benefit of the essential objectives of the Socialist State and the well-being of the people, over any other consideration.”¹⁶⁹ The

¹⁶⁷ *Id.* at art. 26.

¹⁶⁸ *Id.*

¹⁶⁹ *La Reforma*, *supra* note 8. The original text, in Spanish, of Article 318 of *La Reforma* reads as follows:

Artículo 318. El sistema monetario nacional debe propender al logro de los fines esenciales del Estado Socialista y el bienestar del pueblo, por encima de cualquier otra consideración.

El Poder Ejecutivo Nacional, a través del Banco Central de Venezuela, en estricta y obligatoria coordinación, fijará las políticas monetarias y ejercerá las competencias monetarias del Poder Nacional. El objetivo específico del Banco Central de Venezuela, como ente del Poder Ejecutivo Nacional, es lograr las condiciones monetarias, cambiarias y financieras necesarias para promover el crecimiento y el desarrollo económico y social de la Nación. La unidad monetaria de la República Bolivariana de Venezuela es el bolívar. En caso de que se instituya una moneda común en el marco de la integración latinoamericana y caribeña, podrá adoptarse la moneda que sea objeto de los tratados que suscriba la República.

El Banco Central de Venezuela es persona de derecho público sin autonomía para la formulación y el ejercicio de las políticas correspondientes, y sus funciones estarán supeditadas a la política económica general y al Plan de Desarrollo Integral de la Nación para alcanzar los objetivos superiores del Estado Socialista y la mayor suma de felicidad posible para todo el pueblo.

Para el adecuado cumplimiento de su objetivo específico, el Banco Central de Venezuela tendrá entre sus funciones, compartidas con el Poder Ejecutivo Nacional, las de participar en la formulación y ejecución de la política monetaria, en el diseño y ejecución de la política cambiaria, en la regulación de la moneda, el crédito y fijación de las tasas de interés.

Las reservas internacionales de la República serán manejadas por el Banco Central de Venezuela bajo la administración y dirección del Presidente o Presidenta de la República, como administrador o administradora de la Hacienda Pública Nacional.

very next sentence would have put the final nail in the coffin of central bank independence in Venezuela: “The National Executive Power, through *Banco Central de Venezuela*, acting in strict and obligatory coordination, will determine the monetary policies and will exert the monetary authority of the National Power.”¹⁷⁰ This remarkable sentence would have converted the autonomous BCV into nothing more than a mere mechanism of Venezuelan Presidential power. Requiring “strict and obligatory coordination” of the BCV would have likely stripped the long-standing institution of nearly all discretion.

Indeed, Mr. Chavez did not stop his complete restructuring of the BCV at a redefinition of its instrumental independence, he also significantly changed the objectives that the institution was organized to achieve. Recall that the BCV, under Mr. Chavez’s own 2001 Constitution, was charged with the usual Central Bank objective of price stability.¹⁷¹ The Revised Article 318 reflected in *La Reforma* was as follows: “The specific objective of the BCV, just as that of the National Executive Power, is to obtain the monetary, exchange and financial conditions necessary to promote the growth and the economic and social development of the Nation.”¹⁷² That this statement is at once inconsistent with the objectives usually assigned to central banks and deeply populist in tenor does not come as a surprise given Mr. Chavez’s idiosyncratic style. As if to remove any doubt from the equation, the Revised Article 318 continued as follows:

The BCV is person of public right without autonomy for the formulation and the exercise of the corresponding policies, and its functions will be submitted to the general economic policy and the Plan of Integral Development of the Nation to reach the greater objectives of the Socialist State and the greater sum of possible happiness after all the people.¹⁷³

Finally, and perhaps tying all of the various pieces of this “reform” together, the Revised Article 318 would have provided that “[t]he international reserves of the Republic shall be managed by the Central Bank of Venezuela under the administration and direction of the President of the Republic, as administrator of the National Public Treasury.”¹⁷⁴ It is arguably in light of this final provision that one might begin to grasp the full breadth of Mr. Chavez’s intentions. Many observers noted

¹⁷⁰ *Id.*

¹⁷¹ See *supra* note 161 and accompanying text.

¹⁷² *Id.*

¹⁷³ *Id.*

¹⁷⁴ *Id.*

that this amendment would essentially “allow the president to use the country’s reserves as he saw fit.”¹⁷⁵

The stakes were ultimately much higher than the mere question of central bank independence—they related to the very nature of the BCV. Mr. Chavez’s amendments so broadly defined the scope of the nation’s monetary policy, essentially conflating the BCV’s objectives with the advancement of his “Bolivarian” Revolution, such that the BCV would cease to be merely a monetary authority. It would become at once, a slush fund for Mr. Chavez’s regional, political, and military ambitions, not to mention his domestic development goals. A central bank that transforms into a development and relief agency has ceased to be a central bank.

Indeed, there is precedent for this in Venezuela. Such an instance of mixed purposes and the use of a state entity geared towards other purposes can be found in the case of Venezuela’s national oil company, *Petroleos de Venezuela* (“PDVSA”). Since becoming President of Venezuela, Mr. Chavez has used PDVSA’s resources as personal political capital. He has brokered regional energy and pipeline arrangements with allies such as Argentina’s Nestor and Cristina Kirchner.¹⁷⁶ In fact, months before facing his own election, Chavez was embroiled in a scandal when a Venezuelan-American businessman was caught entering Argentina on a private flight with PDVSA officials and \$800,000 hidden in a suitcase, which was allegedly to be used to illegally support the campaign of Cristina Fernandez de Kirchner Wilson.¹⁷⁷ Beyond the regional political motives towards which Mr. Chavez has employed PDVSA’s assets, he has also used billions of dollars of the company’s resources to fund vast social programs.¹⁷⁸ All the while, the once-heralded company has been crumbling due to the fact that basic investments on maintenance have floundered, resulting in inactive oil rigs and falling production.¹⁷⁹

While PDVSA’s example over the last decade is hardly one that merits emulation, the fundamental madness of Mr. Chavez’s proposals to take over the nation’s central bank was in the fact

¹⁷⁵ Editorial, *Saying No to Chávez*, N.Y. TIMES, December 1, 2007, at A14.

¹⁷⁶ Jack Chang, *Chavez Uses Oil Revenue to Win Support in Latin America*, CENTRE DAILY TIMES (State College, PA), March 12, 2006, at A12.

¹⁷⁷ Chris Carlson, *Argentina Seeks Extradition in Case of Briefcase Scandal*, VENEZUELANALYSIS.COM, Aug. 30, 2007, available at <http://www.venezuelanalysis.com/news/2575>.

¹⁷⁸ Cesar J. Alvarez, *Venezuela’s Oil-Based Economy*, COUNCIL ON FOREIGN REL., Feb. 9, 2009, available at <http://www.cfr.org/publication/12089>.

¹⁷⁹ Editorial, *Chavez: A Decade of Disaster*, LATIN BUS. CHRON., Feb. 9, 2009, available at <http://www.latinbusinesschronicle.com/app/article.aspx?id=3138>.

that his intentions seemed to have been to baldly duplicate the PVDSA experiment, writ-large.

C. Aftershocks

As illustrated earlier, although Mr. Chavez's 2007 Amendments ultimately failed to pass, the history of this episode did not end there.¹⁸⁰ While in the euphoria of the moment, one of the leaders of the opposition to Chavez, Manuel Rosales, celebrated "a new path, a democratic path," there can be no doubt that much has changed in the following year.¹⁸¹ Indeed, in his press conference immediately after the results were announced, he noted that the 51 percent to 49 percent defeat was "microscopic."¹⁸² He added, perhaps ominously at the time, though perceptively, "[t]o use a phrase from February 1992, we've fallen short for now."¹⁸³ It was only days after the election before news sources began to report that speculation was beginning to mount about how and when Chavez will attempt to push some of the failed reforms into law, whether unilaterally by decree or through the National Assembly, which he dominates.¹⁸⁴ In the year since, much of this has come to pass. By August 2008, Mr. Chavez had implemented many of the Amendments he had sought to pass the previous year by invoking a sweeping new decree power that had been granted to him by the National Assembly.¹⁸⁵ At the time, he seemed to allude to this new method of implementing his established (and rejected) ideas: "At this point, I'm thinking of presenting a new formula, to expand and launch a new perspective, to raise the speed of the process," Chavez noted.¹⁸⁶ Perhaps most famously, however, in February 2009 Mr. Chavez won a referendum to amend the constitution that lifts the previous two-term limit and ensures he will be able to continue his "socialist revolution" for as long as he retains the support of the electorate.¹⁸⁷ Arguably, given his success in accomplishing much of his rejected reforms, it is only a matter of time before he once again raises the issue of his access to

¹⁸⁰ See *supra* Part III.B.

¹⁸¹ John Otis, *Chavez Power Play Defeated: Opposition Leader Says Nation is on 'a New Path, a Democratic Path'*, HOUSTON CHRON., Dec. 3, 2007, at A1.

¹⁸² Jonathan Gurwitz, *Despot's Loss Doesn't Mean he Won't Try Again Soon*, SAN ANTONIO EXPRESS-NEWS, Dec. 5, 2007, at 11B.

¹⁸³ *Id.*

¹⁸⁴ Daniel Cancel, *Venezuela's Chávez Presses on Undaunted*, CHRISTIAN SCI. MONITOR, Dec. 6, 2007, at 10.

¹⁸⁵ Editorial, *Chavez's Way*, LOS ANGELES TIMES, August 9, 2008, at A18.

¹⁸⁶ *Id.*

¹⁸⁷ Jeremy McDermott, *Chávez Wins Right to Rule Again*, DAILY TELEGRAPH (London), Feb. 17, 2009, at 18.

national reserves and the independence of his nation's central bank.

CONCLUSION

The issue of central bank independence has figured very prominently in Latin America's economic policy debates over the last several decades. While it seemed for some time, especially during the 1990s that the momentum had finally swung in favor of greater autonomy in monetary decision making in this region which had for so long suffered from intermittent hyperinflationary contagion, it now seems that the opposing camp has once again pitched its tents and is preparing for battle. Though the populist forces, most clearly represented by Venezuela's Hugo Chavez, have thus far been unable to do away with central bank autonomy due to resistance on the part of their populace, this has not been for a lack of trying and there is reason to believe they will try again.

Indeed, for a region that has repeatedly suffered the consequences of destructive economic policy, it would seem that once again the barbarians are at the gate. As Professor Larrain warned, "[t]he true risk of high inflation in Latin America lies in populism. . . . Hyperinflation is not dead, it is merely sleeping."¹⁸⁸

The unanswered question before us is whether the region's polities will follow the wisdom of Ulysses and tie themselves to the mast of independent monetary decision making in order to avoid the populist temptations of Latin America's modern-day Sirens or whether the great Latin American ships of state will once again crash into the rocky shoals of hyperinflation.

¹⁸⁸ Larrain, *supra* note 10.

Socrates, Syllogisms, and Sadistic Transactions: Challenges to Mastering U.C.C. Article 9 Through Deductive Reasoning

*Timothy R. Zinnecker**

It is tragic that our law schools do not have an orientation course in logic. We had that great line from Professor Kingsfield in *The Paper Chase*: “You come in here with a head full of mush and you leave thinking like a lawyer.” The Socratic method is the most valuable tool to train students to think like a lawyer. Yet the students—and unfortunately too many of their professors—apparently do not know the elements of deductive and inductive reasoning.

– Hon. Ruggero J. Aldisert¹

INTRODUCTION²

In 2007, an essay entitled “*Logic for Law Students: How to Think Like a Lawyer*” appeared in the *University of Pittsburgh Law Review*.³ The essay, co-authored by federal appellate judge Ruggero J. Aldisert and two of his law clerks, opens with this statement: “Logic is the lifeblood of American law.”⁴ Prompted by Professor Kingsfield’s famous line above,⁵ the co-authors then

* Harry and Helen Hutchens Research Professor, South Texas College of Law (tzinnecker@stcl.edu). Several friends were generous with their time and offered helpful comments on an early draft, including John Blevins, Steve Clowney, John Dolan, Sharon Finegan, Adam Gershowitz, Ken Kettering, Bob Lloyd, Dru Stevenson, and Steve Ware. My employer graciously provided financial support.

¹ See Howard J. Bashman, *20 Questions for Senior Circuit Judge Ruggero J. Aldisert of the U.S. Court of Appeals for the Third Circuit*, http://howappealing.law.com/20q/2003_07_01_20q-appellate-blog_archive.html (last visited Oct. 24, 2009).

² Much of the text and many of the accompanying footnotes in the Introduction and Part I of this article will also appear in the companion article referenced *infra* at note 15 and accompanying text.

³ Ruggero J. Aldisert et al., *Logic for Law Students: How to Think Like a Lawyer*, 69 U. PITT. L. REV. 1 (2007).

⁴ Aldisert, *supra* note 3, at 1. Not every federal appellate judge concurs. See Richard A. Posner, *The Jurisprudence of Skepticism*, 86 MICH. L. REV. 827, 833–34 (1988) (suggesting that “the life of the law really cannot be logic”). Cf. O. W. HOLMES, JR., *THE COMMON LAW* 1 (1881) (“The life of the law has not been logic; it has been experience.”).

⁵ The oft-quoted line is from the following soliloquy delivered by Professor Charles Kingsfield to his first-year contracts class at Harvard Law School:

The study of law is something new and unfamiliar to most of you—unlike any schooling you have ever been through before. We use the Socratic method here: I call on you, ask you a question, and you answer it. Why don’t I just give you a lecture? Because through my questions you learn to teach yourselves.

ask, “What is thinking like a lawyer?” They offer this response: “It means employing logic to construct arguments.”⁶ The authors contend that “our law schools do not give students an orientation in the principles of logic,”⁷ which, in their opinion, “does violence to the essence of the law.”⁸ They then express the essay’s purpose:

[W]e endeavor to explain, in broad strokes, the core principles of logic and how they apply in the law school classroom. Our modest claim is that a person familiar with the basics of logical thinking is more likely to argue effectively than one who is not. We believe that students who master the logical tenets laid out in the following pages will be better lawyers and feel more comfortable when they find themselves caught in the spotlight of a law professor on a Socratic binge.⁹

The authors then expound on these “core principles of logic,” discussing for several pages both deductive reasoning through syllogisms¹⁰ and inductive reasoning by analogy.¹¹

I teach Secured Transactions¹² once (and sometimes two or three times) each year. The course introduces students to many of the legal rules that govern transactions in which personal property (rather than real estate) secures payment of a financial obligation. Particular attention is given to U.C.C. Article 9¹³ and selected provisions of the United States Bankruptcy Code.¹⁴

Through this method of questioning, answering, questioning, answering, we seek to develop in you the ability to analyze that vast complex of facts that constitute the relationships of members within a given society. Questioning and answering. At times, you may feel that you have found the correct answer. I assure you that this is a total delusion on your part. You will never find the correct, absolute, and final answer. In my classroom, there is always another question, another question to follow your answer. Yes, you’re on a treadmill. My little questions spin the tumblers of your mind. You’re on an operating table; my little questions are the fingers probing your brain. We do brain surgery here. You teach yourselves the law, but I train your mind. You come in here with a skull full of mush, and you leave thinking like a lawyer.

The Paper Chase (Twentieth Century Fox 1973). For his portrayal of the “stuffy, crusty, brilliant, and despotic” law professor, the 71-year-old John Houseman received the Oscar for Best Supporting Actor. See 1973 Academy Awards Winners and History, <http://www.filmsite.org/aa73.html> (last visited Oct. 24, 2009).

The movie was based on the book of the same title, published in 1970, and authored by John Jay Osborn, Jr., who, coincidentally, clerked on the United States Court of Appeals for the Third Circuit during the 1971–72 term, during Judge Aldisert’s tenure. THE AALS DIRECTORY OF LAW TEACHERS 872 (2007–08).

⁶ Aldisert, *supra* note 3, at 1.

⁷ *Id.* at 2.

⁸ *Id.*

⁹ *Id.* (footnote omitted).

¹⁰ *Id.* at 3–12.

¹¹ *Id.* at 16–20.

¹² Some of my customers prefer the moniker “Sadistic Transactions” (hence the title).

¹³ One commercial law professor describes Article 9 as “the crowning achievement of the UCC project, and perhaps of the entire uniform law enterprise.” See Edward J.

This article, together with a companion article (*Syllogisms, Enthymemes, and Fallacies: Mastering Secured Transactions Through Deductive Reasoning*¹⁵), respond to Judge Aldisert's co-authored plea for a renewed emphasis on logic and deductive reasoning in the classroom, from my perspective as a Secured Transactions professor. Both articles begin with an introduction to the syllogism, the classic form of deductive reasoning. But the two articles then travel different paths.

Following Part I's introduction to syllogistic analysis, Part II of this article offers seven reasons why Article 9 itself may hinder the effective use of syllogisms in a Secured Transactions course.¹⁶ First, Article 9 occasionally redefines ordinary terms in an unconventional and counter-intuitive manner, which may lead students to reach logical, but incorrect, conclusions when crafting a syllogistic argument. Second, Article 9 sometimes fails to warn the reader that compliance with a straightforward statutory provision may dictate a conclusion that is logical, yet yields disastrous consequences. Third, Article 9 can lead to incorrect conclusions because selected rules cannot be read literally. Fourth, Article 9 occasionally adopts rules that are inconsistent with policy-based analysis, which may lead students to craft incorrect major premises and, accordingly, to reach erroneous conclusions. Fifth, sometimes a rule of Article 9 triggers a result so unexpected and inconsistent with the norm that a logical approach to its understanding is undermined. Sixth, Article 9 occasionally adopts a rule that is so illogical as to render futile any attempt to understand it logically. And seventh, many of the

Janger, *Predicting When the Uniform Law Process Will Fail: Article 9, Capture, and the Race to the Bottom*, 83 IOWA L. REV. 569, 571 (1998). See also Lawrence Ponoroff & F. Stephen Knippenberg, *Having One's Property and Eating it Too: When the Article 9 Security Interest Becomes a Nuisance*, 82 NOTRE DAME L. REV. 373, 374 (2006) ("Powerful, creative, comprehensive yet elegant, Article 9 is generally regarded as the most innovative of the Code's articles.") (footnote omitted). My students are not so effusive in their praise, often asking if Article 9 was drafted in a foreign language and then translated. I hesitate to disclose to them that Grant Gilmore, the principal architect of Article 9, taught French at Yale University for a few years before enrolling in its law school at the age of 29. See Robert M. Jarvis, *Gilmore and Black at 50*, 38 J. MAR. L. & COM. 135, 139-40 (2007).

¹⁴ 11 U.S.C. §§ 101 *et seq.* (2006).

¹⁵ The manuscript of the companion article will be circulated to law journals in spring 2010. A pre-publication draft is available by contacting the author.

¹⁶ Notwithstanding these statutory roadblocks, the companion article will contend that Secured Transactions is an ideal course in which legal principles can be introduced, analyzed, and mastered through deductive reasoning. The article will illustrate—through ten examples covering each of the five major topics of the Secured Transactions course (attachment, perfection, priority, default, and bankruptcy)—how narrative analysis can be expressed as deductive syllogisms, and how deductive syllogisms can be the foundation for enhanced narrative analysis. The article will also examine enthymemes (deductive reasoning which requires the audience to infer part of the argument) and review the formal fallacies which can undermine the logic of the syllogistic argument.

statutory rules of Article 9 are riddled with exceptions, frustrating the ability to easily craft the major premise of the syllogism. The article then offers a brief conclusion (and a glimpse of hope).

I. DEDUCTIVE REASONING AND THE SYLLOGISM

Judge Aldisert and his co-authors open their discussion of deductive reasoning by noting its widespread application, and thus its importance. They suggest that “90 percent of legal issues can be resolved by deduction”¹⁷ and contend that “deductive reasoning . . . is the driving force behind most judicial opinions.”¹⁸ In their opinion, the form in which deductive reasoning most often appears is the syllogism—“a label logicians attach to any argument in which a conclusion is inferred from two premises.”¹⁹

The authors then introduce the reader to syllogisms by offering “the immortal example of logicians everywhere”²⁰:

All men are mortal.

Socrates is a man.

Therefore, Socrates is mortal.²¹

The first statement (“All men are mortal.”) is “the major premise,” which “states a broad and generally applicable truth.”²² The second statement (“Socrates is a man.”) is “[t]he minor premise” and “states a specific and usually more narrowly applicable fact.”²³ The third statement (“Therefore, Socrates is

¹⁷ Aldisert, *supra* note 3, at 2.

¹⁸ *Id.* at 3.

¹⁹ *Id.* See also LEWIS H. LARUE, GUIDE TO THE STUDY OF LAW: AN INTRODUCTION 172 (2nd ed. 2001) (“In the kingdom of law, the syllogism reigns supreme.”); Daniel S. Goldberg, *I Do Not Think It Means What You Think It Means: How Kripke and Wittgenstein’s Analysis on Rule Following Undermines Justice Scalia’s Textualism and Originalism*, 54 CLEV. ST. L. REV. 273, 276 (2006) (“The syllogism is perhaps the touchstone of modern legal method.”); Amy E. Sloan, *Erasing Lines: Integrating the Law School Curriculum*, 1 J. ASS’N LEGAL WRITING DIRECTORS 3, 4 n.4 (2002) (“Whatever its shortcomings, however, syllogistic reasoning remains at the core of much legal analysis and is, in many respects, the starting point of a first-year law student’s education in law and argumentation.”).

²⁰ Vincent A. Wellman, *Practical Reasoning and Judicial Justification: Toward an Adequate Theory*, 57 U. COLO. L. REV. 45, 91 n.206 (1985).

²¹ Aldisert, *supra* note 3, at 3. This example (sometimes in a slightly different format) makes an occasional appearance in judicial opinions. See, e.g., Tyler v. Cain, 533 U.S. 656, 672–73 (2001) (Breyer, J., dissenting); *Dukes v. Wal-Mart, Inc.*, 474 F.3d 1214, 1245 (9th Cir. 2007) (Kleinfeld, J., dissenting); *Bird, Marella, Boxer & Wolpert, et al. v. Superior Court*, 106 Cal. App. 4th 419, 430 (2003).

²² Aldisert, *supra* note 3, at 4. Cf. PATRICK J. HURLEY, A CONCISE INTRODUCTION TO LOGIC 260 (6th ed. 1997) (defining the “major term” as “the predicate of the conclusion” and the “minor term” as “the subject of the conclusion”).

²³ Aldisert, *supra* note 3, at 4.

mortal.”) is “the conclusion,” which “draws upon these premises and offers a new insight that is known to be true based on the premises.”²⁴ The authors then reveal this important observation:

[T]he three parts of a syllogism—the two premises and the conclusion—are themselves built from three units. Logicians call these units “terms.” Two terms appear in each statement: the “major term” in the major premise and conclusion, the “minor term” in the minor premise and conclusion, and the “middle term” in the major and minor premises but not in the conclusion. Notice that the middle term covers a broad range of facts, and that if the conclusion is to be valid, the minor term must be a fact that is included within the middle term. Although the jargon can get confusing, the basic idea isn’t hard to grasp: Each statement in a syllogism must relate to the other two.²⁵

The syllogism strikes me as an ideal instruction tool in Secured Transactions, a course built on a foundation of statutory rules, most of which are codified in U.C.C. Article 9. Article 9 reveals how a debtor may create an enforceable security interest in its property.²⁶ It provides rules that address perfection of a security interest.²⁷ The statute devotes twenty-three provisions to priority matters,²⁸ twenty-seven to filing concerns,²⁹ and twenty-eight to default issues.³⁰ Additionally, Article 9 includes a list of eighty (EIGHTY!) defined terms!³¹ Given Article 9’s attention to detail and exhaustive coverage, Secured Transactions seems like an ideal course in which to emphasize syllogistic analysis. State the rule (major premise). State the relevant facts (minor premise). Draw a conclusion.³² How hard can this be?³³

²⁴ *Id.*

²⁵ *Id.* at 6.

²⁶ See generally U.C.C. § 9-203(a), (b) (2009). Statutory citations are to the American Law Institute’s 2009–2010 edition of the Uniform Commercial Code Article 9.

²⁷ See generally §§ 9-301 through 9-316.

²⁸ See §§ 9-317 through 9-342.

²⁹ See §§ 9-501 through 9-527.

³⁰ See §§ 9-601 through 9-628.

³¹ See § 9-102(a)(1)–(80). Article 9 also incorporates definitions found elsewhere in the U.C.C. See § 9-102(b), (c).

³² See James M. Boland, *Legal Writing Programs and Professionalism: Legal Writing Professors Can Join the Academic Club*, 18 ST. THOMAS L. REV. 711, 722 (2006) (suggesting that, in describing the syllogism, “the most important concept to grasp is that the major premise states the rule of law, the minor premise states the facts relevant to the major premise, and the conclusion flows logically from the premises”).

³³ The syllogism bears a striking resemblance to the traditional “IRAC” formula on which all law students are weaned, an observation shared by others. See, e.g., Adam Todd, *Neither Dead Nor Dangerous: Postmodernism and the Teaching of Legal Writing*, 58 BAYLOR L. REV. 893, 938 (2006) (observing that “the deductive syllogism [is] commonly known by the acronym IRAC”); Jill J. Ramsfield, *Is “Logic” Culturally Based? A Contrastive, International Approach to the U.S. Law Classroom*, 47 J. LEGAL EDUC. 157, 186 n.117 (1997) (“IRAC is the acronym for Issue, Rule, Application, Conclusion that is often taught to first-year students as an adaptation of the syllogism’s major premise,

From my side of the lectern, the approach holds quite an attraction. But Article 9 has been my traveling companion for over fifteen years, so maybe my perspective is not shared by the student who journeys at my side for only fifteen weeks. My students and I can readily conclude that Socrates must be mortal if we know that all men are mortal and Socrates is a man. But using that simple illustration as a template to master a complex and technical collection of statutory provisions may appear impossible to the typical student. The task is not impossible, but it may be daunting. Some of the challenges are created by Article 9 itself and are examined in Part II.

II. STATUTORY ROADBLOCKS TO THE EFFECTIVE USE OF THE SYLLOGISM

Given its emphasis on statutory analysis, Secured Transactions is a course where deductive reasoning can be an effective, and oft-used, teaching tool. But Article 9 itself creates roadblocks which may frustrate learning through syllogistic analysis. This part of the article highlights seven areas of concern.

A. "What we've got here is failure to communicate."³⁴

The major premise of the syllogism—"All men are mortal"—is the rule of law. In Secured Transactions, the rule of law is most often drawn from Article 9 (e.g., "A secured party may perfect a security interest in the debtor's equipment by filing a financing statement."). The minor premise of the syllogism—"Socrates is a man"—is a factual statement. In Secured Transactions, the factual statement is extracted from a case or a problem (e.g., "Alamo Bank filed a financing statement against BizCorp's equipment in which Alamo Bank has an enforceable security interest."). The conclusion of the syllogism—"Therefore,

minor premise, and conclusion."); Brooks R. Fudenberg, *Unconstitutional Conditions and Greater Powers: A Separability Approach*, 43 UCLA L. REV. 371, 445 n.284 (1995) ("The well-known IRAC (Issue, Rule, Application, Conclusion), would take the form of syllogisms. For example, after one identifies the issue, the next steps are to find the rule (major premise), do the application (minor premise), and reach a conclusion.").

³⁴ Most film buffs will recognize this quotation from *Cool Hand Luke* (Warner Bros. 1967). The line is spoken by "the Captain" (Strother Martin) and addressed to Luke (Paul Newman), who is having difficulty conforming to rural prison life after being sentenced to a chain gang for destroying several parking meters. The line is #11 on the American Film Institute's list of "AFI's 100 Years . . . 100 Movie Quotes." See <http://connect.afi.com/site/DocServer/quotes100.pdf?docID=242> (last visited Nov. 8, 2009).

Initially, Telly Savalas was cast as the title character, but he was in Europe and refused to fly to the United States. The producers did not wait for Savalas to return by boat and instead chose to cast Paul Newman. See Telly Savalas News and Trivia, http://www.celebritywonder.com/html/Tellysavalas_trivia1.html (last visited Oct. 24, 2009).

Socrates is mortal”—is a truth drawn from the major and minor premises. In Secured Transactions, the student reviews the rule of law and the factual statement and reaches a conclusion (*e.g.*, “Alamo Bank has perfected its security interest in BizCorp’s equipment.”).

Observe, however, that a party may reach an incorrect conclusion if terms common to both premises have different meanings. Perhaps the author of the major premise—the rule maker—defines “man” as “all males over 21 years of age.” Suppose the author of the minor premise—the fact finder—defines “man” as “all males who have reached puberty.” Because the authors of both premises are not singing from the same hymnal, the syllogism will yield an incorrect conclusion if Socrates has reached puberty but is not over 21 years of age.³⁵ And this problem can arise in a Secured Transactions course because Article 9 occasionally redefines ordinary terms in an unconventional and counter-intuitive manner, which may lead students to reach logical, but incorrect, conclusions.

Consider, then, the following three examples which examine some basic Article 9 terminology.

Example #1: Roger, a lawyer, is an avid collector of scarce baseball cards. Last week he purchased the “Mona Lisa” of baseball cards—a rare Honus Wagner³⁶ card released in 1909 by the American Tobacco Company—for \$3.25 million.³⁷ He keeps the card in a temperature-

³⁵ Cf. Posner, *supra* note 4, at 834:

Even the syllogism about Socrates would be unimpressive if the speaker were a lexicographical dictator (O’Brien in 1984?) who could redefine ‘man’ and ‘mortal’ at a stroke of the pen; or if Socrates were a boy, a god, a character in fiction, or a hermaphrodite; or if listeners had imperfect command of the speaker’s language.

This point hit home for me during a recent discussion with a colleague. We were talking about his clerkship and whether he enjoyed life in Charleston. He made some comments that suggested to me that we had very different opinions of Charleston. Of course we did. He was talking about Charleston, West Virginia, but I assumed we were talking about Charleston, South Carolina.

³⁶ Honus Wagner, a longtime shortstop for the Pittsburgh Pirates, “was a sensational hitter, a brilliant base-runner and a flawless fielder” and “is considered by many to be baseball’s greatest all-around player.” See Honus Wagner Biography, <http://www.honuswagner.com/biography.htm> (last visited Oct. 24, 2009). He is one of five members of the inaugural class inducted into baseball’s hall of fame in 1936 (along with Ty Cobb, Walter Johnson, Christy Mathewson, and Babe Ruth). See Dan Holmes, *The First Five: A Look at the Original Members of the Baseball Hall of Fame*, Feb. 18, 2007, http://www.baseballhalloffame.org/news/article.jsp?ymd=20070218&content_id=571&vkey=hof_new.

³⁷ The reader who believes Roger overpaid should note that a Wagner “T206” card once owned by hockey legend Wayne Gretzky sold for \$2.8 million in September 2007, only six months after selling for \$2.35 million. See NBCSports.com, *1909 Wagner card sells for record \$2.8 million*, Sept. 6, 2007, <http://nbcports.msnbc.com/id/20629113/>. A recently published book recounts the history of this card, what the dust jacket describes as the “holy grail” of baseball card collecting. See MICHAEL O’KEEFFE & TERI THOMPSON,

controlled safety deposit box at his bank.

Under Article 9, the Honus Wagner card is:

- A. equipment.
- B. a consumer good.
- C. investment property.

Example #2: A retail store sells stand-alone software disks, music and book CDs, and movie DVDs. Most customers buy these products for personal, family, or household use.

Under Article 9, the items in the store are:

- A. general intangibles.
- B. inventory.
- C. consumer goods.

Example #3: Jennifer, a recent college graduate, received a \$15,000 business loan last week. The bank required Jennifer's parents to provide adequate collateral.³⁸ Jennifer and both parents executed the promissory note, obligating themselves jointly and severally.

Under Article 9:

- A. both parents are debtors.
- B. both parents and Jennifer are debtors.
- C. Jennifer is the only debtor.

Most students select an answer based on the following reasoning. In Example #1, Roger is a natural person, rather than a dealer, so the baseball card is a consumer good (or perhaps investment property if Roger was motivated to purchase the card because its market value is likely to increase with time). Equipment? How silly. In Example #2, the retail store is in the business of selling the software, CDs, and DVDs, so the items are inventory (or perhaps consumer goods if a customer's intended use is relevant). They cannot be general intangibles because the disks have a tangible quality. In Example #3, Jennifer is a debtor because she applied for and received the loan proceeds and executed the promissory note. Her parents, too, are debtors because they executed the promissory note. In each instance, the

THE CARD: COLLECTORS, CON MEN, AND THE TRUE STORY OF HISTORY'S MOST DESIRED BASEBALL CARD (2007).

³⁸ If Jennifer intended to use the loan proceeds for a personal, family, or household purpose (instead of a business purpose), then the bank's insistence on collateral might constitute an "unfair act or practice" under the Federal Trade Commission Act. See 16 C.F.R. §§ 444.2(a)(4) (2009) (indicating that a lender commits an unfair act or practice if it takes or receives from a consumer a non-possessory, non-purchase money, security interest in household goods); 444.1(d) (defining a "consumer" as "a natural person who seeks or acquires goods, services, or money for personal, family, or household use"); 441.1(i) (defining "household goods").

student uses reasoning to reach a conclusion that is logical—yet possibly wrong.

In Example #1, the card is movable, so it is a “good”³⁹ and cannot be “investment property”⁴⁰ (even if Roger may have been partially motivated to purchase the card for its probable appreciation in value⁴¹). The four mutually-exclusive categories of goods are consumer goods, inventory, farm products, and equipment.⁴² The baseball card might not be a “consumer good” because it is stored in a vault and not being “used” for a “personal, family, or household” manner.⁴³ Roger, an attorney, is

³⁹ See § 9-102(a)(44) (defining “goods” as “all things that are movable when a security interest attaches” and expressly excluding, among other types of collateral, “investment property”).

⁴⁰ *Id.* (expressly excluding from the definition of “goods” several types of collateral, including “investment property”). See also § 9-102(a)(49) (limiting “investment property” to securities [defined at § 8-102(a)(15)], security entitlements [§ 8-102(a)(17)], securities accounts [§ 8-501(a)], commodity contracts [§ 9-102(a)(15)], and commodity accounts [§ 9-102(a)(14)]).

⁴¹ I use a casebook in which the authors pose a problem involving a Ming Dynasty horse statue kept by a consumer debtor in a temperature-controlled bank vault. In classifying the statue, the authors write: “Is the statue ‘investment property?’ NO. Investment property is a term of art defined in § 9-102(a)(49). It includes stocks, bonds, etc. Investment property is not goods. The statue is goods (moveable, tangible, etc.). Goods held for investment purposes are either equipment or inventory.” RAYMOND T. NIMMER, ET AL., COMMERCIAL TRANSACTIONS: SECURED FINANCING—CASES, MATERIALS, PROBLEMS, 155, 159 (3rd ed. 2003) (Teacher’s Manual, Problem 3.11). I believe that goods held for investment purposes also may be “consumer goods.” But that point does not detract from the authors’ conclusion (with which I agree) that an acquisition motivated by investment sentiment does not convert a “good” into “investment property.”

⁴² See § 9-102 cmt. 4.a. (first paragraph).

⁴³ See § 9-102(a)(23) (defining “consumer goods”). With respect to the Ming Dynasty horse statue, one of the authors “thinks the statue is equipment . . . the residual category of goods . . . because [the owner] was not using the statue for personal, household or family purposes. Indeed, was he ‘using’ the statue at all?” This author acknowledges, however, that a court “might conclude the statue is consumer goods. Students do so every year.” See NIMMER, *supra* note 41, at 159. Professor Ken Kettering, however, believes that a purchase “for the sheer pleasure of ownership” is “personal use” and contends that an “argument that [the owner] isn’t ‘using’ it because it’s tucked away in a vault is just a bad pun. He can’t enjoy it unless he hangs it on a wall, rubs it over his body? Nonsense.” See Memorandum from Ken Kettering to Tim Zinnecker, dated August 20, 2008 (copy on file with author). The fact that prominent commercial law scholars can disagree on the classification further illustrates my point.

The distinction between a “consumer good” and “equipment” is important for several reasons. First, if the security agreement or financing statement describes the collateral by type, choosing the incorrect label may dictate whether the secured party has a security interest or the security interest is perfected. See §§ 9-203(b)(3)(A) (requiring the debtor to authenticate a security agreement “that provides a description of the collateral”); 9-502(a)(3) (indicating that a financing statement is sufficient if it “indicates the collateral”); 9-108(b)(3) (permitting collateral descriptions by “type of collateral defined” by Article 9). Second, a purchase-money security interest in many consumer goods, but not equipment, may be perfected automatically upon attachment. See § 9-309(1). Third, some of the default provisions in Part 6 of Article 9 apply only to consumer goods. See, e.g., §§ 9-620(e) (imposing mandatory dispositions of consumer goods in some instances); 9-624(c) (prohibiting waiver of redemption rights in consumer

not holding the card for sale or lease in the ordinary course of his business (at least absent evidence that he sells his cards on a regular basis), and the card does not otherwise fit within the definition of “inventory.”⁴⁴ And even if Wagner spent time on a “farm club” roster,⁴⁵ neither the card (nor Wagner himself) would be a “farm product.”⁴⁶ So it can be plausibly argued that the card may fall within the residual category: equipment.⁴⁷

In Example #2, the various software disks, music CDs, and movie DVDs are not “consumer goods” in the hands of the retail store. The fact that most customers purchase the items for personal, family, or household use is irrelevant (until the customer is the debtor and becomes the party whose primary use dictates the proper label). Nor are the items necessarily “inventory,” even though the retail store is in the business of selling these tangible products. “Inventory” applies only to “goods,”⁴⁸ but the definition of “goods” excludes “a computer program embedded in goods that consist solely of the medium in which the program is embedded.”⁴⁹ The stand-alone software

goods); 9-625(c) (permitting recovery of minimum statutory damages for secured party’s noncompliance if collateral is consumer goods).

⁴⁴ See § 9-102(a)(48) (defining “inventory”) and cmt. 4.a (“Implicit in the definition [of “inventory”] is the criterion that the sales or leases are or will be in the ordinary course of business.”).

⁴⁵ A “farm club” is a minor league baseball team. The term is credited to baseball executive Branch Rickey, who jokingly observed that minor league teams in small rural towns were “growing players down on the farm like corn.” See Minor League Baseball: Information From Answers.com, <http://www.answers.com/topic/minor-league-baseball> (last visited Oct. 19, 2009). Rickey, who earned a law degree from the University of Michigan, was instrumental in integrating major league baseball. His accomplishments included signing African-American Jackie Robinson and later drafting the first Latin-American superstar, Roberto Clemente (who reached the 3,000-hit milestone on the last day of the 1972 season and then tragically died on December 31, 1972, in a plane crash while on a humanitarian mission to Nicaragua). See John Stiglich, *Branch Rickey: Forgotten American Hero*, THE MICHIGAN DAILY, May 7, 2006, available at <http://www.michigandaily.com/content/john-stiglich-branch-rickey-forgotten-american-hero>; ‘Clemente’ Tells Story of a True Baseball Hero, <http://www.npr.org/templates/story/story.php?storyId=5369849> (last visited Oct. 19, 2009).

⁴⁶ See § 9-102(a)(34) (defining “farm products” to include crops and livestock that are the subject of a “farming operation”).

⁴⁷ See § 9-102(a)(33) (defining “equipment” as “goods other than inventory, farm products, or consumer goods”).

⁴⁸ See § 9-102(a)(48) (“Inventory” means goods . . .”).

⁴⁹ See § 9-102(a)(44) (penultimate sentence of definition). See also 9A WILLIAM D. HAWKLAND ET AL., UNIFORM COMMERCIAL CODE SERIES [REV] § 9-102.10 at 9-135 (2000) (“[A] program embedded in goods that consist solely of the medium in which the program is embedded is not a good, such as a diskette.”); Ralph J. Rohner, *Leasing Consumer Goods: The Spotlight Shifts to the Uniform Consumer Leases Act*, 35 CONN. L. REV. 647, 693 (2003) (observing that “a computer program that the consumer downloads into a home computer is likely to be treated as ‘software’ and not goods, even though the software is embedded in a tangible plastic disk”) (footnote omitted).

disks fall within the exclusionary language,⁵⁰ as might the music CDs and movie DVDs if the audio and video data are accompanied by a computer program. But even if these audio and video disks consist solely of data and are not accompanied by a computer program, the value of the disks is not the tangible disks themselves but the data on the tangible disks.⁵¹ These disks, then, could be “general intangibles.”⁵²

Finally, in Example #3, Jennifer is not a debtor, even though she has joined her parents in executing the promissory note and incurring liability for repayment. Jennifer’s parents are offering the collateral, so they are “debtors.”⁵³ But Jennifer does not have any property interest in the collateral, so she is not a “debtor” even though she is the direct beneficiary of the loan proceeds and is contractually obligated to repay the loan.⁵⁴

⁵⁰ The software disks probably are “software,” defined at § 9-102(a)(75) and, therefore, are expressly included within the definition of “general intangibles” at § 9-102(a)(42).

⁵¹ See *In re Information Exch., Inc.*, 98 B.R. 603, 604–05 (Bankr. N.D. Ga. 1989) (concluding nonfiling creditor’s possession of computer tape did not perfect a security interest because the collateral—the information and programming on the tape, rather than the tape itself—was a “general intangible” and not a “good”); *Page v. Hotchkiss*, 52 U.C.C. Rep. Serv. 2d 365, 368–71 (Conn. Super. Ct. 2003) (initially concluding under U.C.C. Article 2 that software developed by Hotchkiss for Page was a programming “service” rather than a “good,” and then observing under U.C.C. Article 9 “that what Hotchkiss provided was not goods since what was provided was a modified computer program embedded in a convenient medium for Page to access” and that “[t]he real object of Page’s purchase was the intellectual property which had been loaded and stored on a transferable medium”). Cf. *In re C Tek Software, Inc.*, 117 B.R. 762, 767–68 (Bankr. D.N.H. 1990) (relying on *In re Bedford Computer Corp.*, 62 B.R. 555, 567 (Bankr. D.N.H. 1986), to conclude that “the source code is more tangible than intangible because ‘the source code . . . cannot exist independent from the actual hardware components to which it gives operational life’”).

⁵² Determining whether the CDs and the DVDs are “inventory” or “general intangibles” is important for several reasons under Article 9. For example, using one—but not both—terms in the collateral description in the security agreement or financing statement could leave the secured party at best with an unperfected security interest in the items and at worst no security interest in the items. See §§ 9-203(b)(3)(A) (requiring the debtor to authenticate a security agreement “that provides a description of the collateral”); 9-502(a)(3) (indicating that a financing statement is sufficient if it “indicates the collateral”); 9-108(b)(3) (permitting collateral descriptions by “type of collateral defined” by the U.C.C.). Also, a secured party can perfect a security interest in inventory, but not general intangibles, by taking possession. See § 9-313(a) (listing “goods” but not “general intangibles”).

⁵³ See § 9-102(a)(28) (defining “debtor” as “(A) a person having an interest . . . in the collateral . . .”).

⁵⁴ See § 9-102 cmt. 2.a. (exs. 2 and 3). Because Jennifer has incurred a payment obligation on a secured debt, she is an “obligor” under § 9-102(a)(59)(i). Correctly identifying the “debtor” is important for a variety of reasons. For example, most security interests are memorialized in a written security agreement, which must be authenticated by the “debtor.” See § 9-203(b)(3)(A). Also, if the secured party disposes of the collateral following default, the “debtor” usually is entitled to receive an asset disposition notice under § 9-611(c)(1) and any surplus proceeds under § 9-615(d)(2).

The preceding discussion, then, illustrates one possible roadblock to the effective use of syllogisms in a Secured Transactions course. Article 9 occasionally redefines ordinary terms in an unconventional and counter-intuitive manner, which may lead students to reach logical, but incorrect, conclusions.

- B. “[B]ecause thou hast not given him warning, he shall die in his sin.”⁵⁵

Socrates decides to attend law school (he’s curious to know more about this “Socratic method” that seems to be so popular with the students). The application form states: “Your application is complete only if it includes an undergraduate transcript, two letters of reference, a recent LSAT score, and \$50.” So Socrates completes and personally delivers the application form, together with the other requested materials, to the admissions committee. Several months come and go, and Socrates does not hear from the admissions committee. He contacts the committee and receives a short letter, the body of which states: “The admissions committee has not yet reviewed your application because your file is missing a ‘personal essay.’ We have filled all slots for the incoming class, but we encourage you to apply again next year.” Alas (for Socrates), the application form failed to direct Socrates to a paragraph found elsewhere in the law school’s literature which gives the admissions committee the discretion to decline to review an application file that fails to include a “personal essay.” According to the form, Socrates submitted a complete application. He can logically conclude from the statement on the form, and his compliance with the form’s requests, that his application will not be ignored by the admissions committee because his file fails to provide additional information not mentioned on the application form. But his logical conclusion has led to an unwelcome result.

And Article 9 sets a similar trap for the unwary by occasionally failing to warn the reader that compliance with a straightforward rule may render a logical conclusion, but yield disastrous consequences.

Consider the following hypothetical. To secure repayment of a \$25,000 loan, BizCorp grants to SmallBank an enforceable security interest in all of BizCorp’s current and future accounts, equipment, and inventory. SmallBank will perfect its security interest by filing a financing statement with the appropriate

⁵⁵ *Ezekiel* 3:20 (King James). It gets worse: “. . . and his righteousness which he hath done shall not be remembered : but his blood will I require at thine hand.” *Id.* Yikes!

filing office.⁵⁶ Section 9-502 states that a financing statement “is sufficient only if it: (1) provides the name of the debtor; (2) provides the name of the secured party . . . ; and (3) indicates the collateral covered by the financing statement.”⁵⁷ SmallBank’s lawyer prepares the following form for filing:

FINANCING STATEMENT

NAME OF DEBTOR:	<i>BizCorp</i>
NAME OF SECURED PARTY:	<i>SmallBank</i>
COLLATERAL:	<i>BizCorp’s current and future accounts, equipment, and inventory</i>

She submits the form, along with the correct fee and an instruction letter, to the appropriate filing officer. The filing officer timely informs SmallBank’s lawyer that it has rejected the filing.⁵⁸

TRUE or FALSE: Because SmallBank’s filing provided all of the information required by § 9-502(a), the filing officer improperly rejected the filing.

Most students will compare the information on the form (minor premise) against the information required by § 9-502 (major premise) and conclude that the filing is “sufficient.” These students then conclude that because the filing is “sufficient” (e.g., adequate, proper, correct, etc.),⁵⁹ the filing officer improperly rejected the filing. But this logical rationale leads to an incorrect conclusion. In fact, the rejection was proper!

Concerned that filing officers might exercise too much discretion when deciding whether to accept a financing statement, the drafters provided an exclusive list of grounds for rejection.⁶⁰ Those grounds include the omission of specific pieces

⁵⁶ See § 9-310(a) and cmt. 2 (stating the general rule that a financing statement will perfect an enforceable security interest in most types of collateral).

⁵⁷ § 9-502(a).

⁵⁸ See § 9-520(b) (requiring the filing officer to “communicate to the person that presented the [financing statement] the fact of and reason for the refusal . . . in no event more than two business days after the filing office receives the [financing statement]”).

⁵⁹ Cf. §§ 9-520(c) (noting that a “filed” financing statement which satisfies § 9-502(a) is “effective”); 9-520 cmt. 3 (referring to a financing statement which complies with § 9-502 as “legally sufficient to perfect a security interest”).

⁶⁰ See Steven L. Harris & Charles W. Mooney, Jr., *Revised Article 9 Meets the Bankruptcy Code: Policy and Impact*, 9 AM. BANKR. INST. L. REV. 85, 100 n.83 (2001) (noting the inclusion of provisions for the purpose of “limiting the grounds upon which a filing office may reject a financing statement”); Christopher S. Bose, *A Trap for the Unwary: Revised UCC Article 9’s Deceptive Technical Guillotine for Financing Statements*, 55 CONSUMER FIN. L.Q. REP. 152, 152 (2001) (noting the drafters’ “competing consideration . . . to restrain instances where filing offices reject attempted filings on improper grounds”).

of information *not required by § 9-502*, including: (i) a mailing address for the secured party,⁶¹ (ii) the mailing address for the debtor,⁶² and (iii) an indication of whether the debtor is an individual or an organization⁶³ (and, if the latter, the type of organization, the jurisdiction of organization, and an organizational identification number⁶⁴). In effect, then, Article 9 bifurcates information into two classes: information necessary for an effective filing (§ 9-502(a)), and information necessary to avoid rejection by the filing office (§ 9-516(b)).⁶⁵ Because the filing submitted by SmallBank's lawyer failed to include these pieces of information, the filing officer had several legitimate grounds for rejecting the filing.⁶⁶

Section 9-502(a) specifically describes SmallBank's filing as "sufficient" because it includes the information mandated by the subsection. But surprisingly (at least to students who retain an amateur status), a "sufficient" filing can be rejected, an act which effectively leaves the secured party unperfected. Certainly the reader deserves some warning that § 9-502(a) is potentially misleading and may trap the unwary.⁶⁷ So where is this warning found? The reader will find no cautionary language in subsection (a)⁶⁸ or any of the other three statutory subsections of § 9-502.⁶⁹

⁶¹ See § 9-516(b)(4).

⁶² See § 9-516(b)(5)(A).

⁶³ See § 9-516(b)(5)(B). See also §§ 1-201(b)(25) (defining "organization"); 9-102(c) (incorporating definitions in Article 1).

⁶⁴ See § 9-516(b)(5)(C)(i)–(iii).

⁶⁵ One author describes the "analysis of this bifurcation" of information as "increasingly complex." Lawrence R. Ahern, III, *Workouts Under Revised Article 9: A Review of Changes and Proposal for Study*, 9 AM. BANKR. INST. L. REV. 115, 148 (2001). See also Richard H. Nowka, *Minor Errors in 'In-Lieu-Of' Statements Under U.C.C. Section 9-706: Did the Drafters of Revised Article 9 Forget the Safety Net?*, 42 BRANDEIS L.J. 721, 729–30 (2004) (observing the "interesting drafting peculiarity" of §§ 9-502 and 9-516). Cf. Ingrid Michelsen Hillinger & Michael G. Hillinger, 2001: *A Code Odyssey (New Dawn for the Article 9 Secured Creditor)*, 106 COM. L.J. 105, 144 (2001) ("If [the information stated in § 9-502(a)] is all that Revised Article 9 requires by way of valid financing statement, you might legitimately wonder why the initial financing statement form included in revised section 9-521 asks for considerably more information.").

⁶⁶ The information required by § 9-516(b) "assists searchers in weeding out 'false positives,' i.e., records that a search reveals but which do not pertain to the debtor in question." § 9-520 cmt. 3.

⁶⁷ Bose, *supra* note 60, at 157 ("Section 9-502(a) may lead the less experienced Article 9 reader to believe that the requirements for a financing statement are simple, easy, and nontechnical, thereby trapping the unwary into a failure to achieve perfection of the security interest.").

⁶⁸ Subsection (a) does direct the reader to subsection (b) if the filing covers as-extracted collateral, timber to be cut, or goods which are or may become fixtures. § 9-502(a) ("Subject to subsection (b) . . .").

⁶⁹ The reader who reviews the accompanying "Official Comments" may stumble across this inconspicuous and rather innocuous statement: "In addition, the filing office must reject a financing statement lacking certain other information formerly required as a condition of perfection (e.g., an address for the debtor or secured party). See Sections 9-516(b), 9-520(a)." § 9-502 cmt. 4. Most of us who have been anointed to teach Secured

Nor will the reader observe any red flags in either of the two neighboring statutes, §§ 9-501 and 9-503.⁷⁰ Not until several sections (and pages) later are the grounds for rejecting a “sufficient” filing tucked away—in § 9-516(b).⁷¹

The foregoing analysis reveals another potential difficulty in presenting material in a Secured Transactions course in a syllogistic format. Article 9 occasionally fails to warn the reader that compliance with a straightforward rule may yield both a logical conclusion and disastrous consequences.

C. “How clever you are, my dear! You never mean a single word you say.”⁷²

Socrates is flying to Chicago to present a paper on the musings of noted philosopher L. P. Berra.⁷³ He arrives at the airport and notices several signs that state: “Each passenger is limited to one carry-on article. No exceptions.” Socrates observes a passenger in the boarding line with a laptop

Transactions, however, will attest that few of our students read the Official Comments on a consistent basis. *Cf.* Dale Beck Furnish, *The Creation and Notice of Security Interests in Movable Property*, 36 U.C.C. L.J. 99, 107 n.17 (2003) (concluding that “[o]nly the rigorously compulsive and insightful reader of the statute is likely to make the connection” between §§ 9-502(a) and 9-516(b) because the “only clue” is “hidden” in § 9-502 cmt. 4).

⁷⁰ These two sections address filing location and sufficiency of party names, respectively.

⁷¹ Section § 9-520(a) states that a filing officer “shall refuse to accept” a financing statement for any reason stated in § 9-516(b), which might prompt the reader to conclude that a rejectable financing statement cannot be effective. That conclusion, although logical, is incorrect. *See* § 9-520(c) (indicating that the filing is effective, “even if the filing office is required to refuse to accept it for filing under subsection (a),” but affording protection under § 9-338 to specific parties who may have relied on incorrect information in the filing). *See also* Robert M. Lloyd, *The New Article 9: Its Impact on Tennessee Law (Part I)*, 67 TENN. L. REV. 125, 173 (1999) (referring to the “convolution” that exists between §§ 9-520(a) and 9-516(b)).

⁷² Penned by playwright Oscar Wilde and spoken by Lady Hunstanton to Mrs. Allonby in *A Woman of No Importance* (act 2). *See* OSCAR WILDE, *A WOMAN OF NO IMPORTANCE*, in 6 THE COMPLETE WORKS OF OSCAR WILDE 65 (The Aldine Publishing Co., 1910). Readers may not know that Wilde was christened with three middle names (Fingal O’Flahertie Wills), and, despite his magic as a wordsmith, authored only one novel: *THE PICTURE OF DORIAN GRAY*. *See* H. MONTGOMERY HYDE, *OSCAR WILDE: A BIOGRAPHY* 5, 148 (1975).

⁷³ Known as “Yogi” to his friends. His most profound thoughts include the following: “This is like *deja vu* all over again.” “You can observe a lot just by watching.” “Nobody goes there anymore; it’s too crowded.” “If you come to a fork in the road, take it.” “Texas has a lot of electrical votes.” “You should always go to other people’s funerals; otherwise, they won’t come to yours.” *See* Things People Said: Yogi Berra Quotes, <http://www.rinkworks.com/said/yogiberra.shtml> (last visited Oct. 19, 2009). Yogi Berra also was one of baseball’s greatest catchers. During his 19-year career with the New York Yankees, he won the American League M.V.P. award three times and played on 14 pennant-winning teams and 10 World Series championships. He was elected to the Hall of Fame in 1972. *See* National Baseball Hall of Fame and Museum, *Yogi Berra*, <http://www.baseballhalloffame.org/hofers/detail.jsp?playerId=110925> (last visited Oct. 19, 2009).

computer, a briefcase, and a small garment bag. Socrates, who has never flown before this trip, logically concludes, from the signs, that this passenger will be asked to check two of the items.⁷⁴ But the passenger is permitted to board the airplane with all three items! Socrates discovers that he is sitting next to this passenger. Socrates expresses amazement that the passenger could carry the laptop computer, the briefcase, and the small garment bag onto the plane. “Doesn’t the sign tell all passengers to limit themselves to one carry-on article?” Socrates asks. “Yes,” says the passenger, “but you can’t read the sign literally.”

And a student who reads Article 9 literally may confront the same predicament of reaching logical, but incorrect, conclusions.⁷⁵

⁷⁴ He also might logically conclude that the passenger carrying a six-inch screwdriver and long knitting needles will have those items confiscated. Maybe not! See TSA: Prohibited Items, <http://www.tsa.gov/travelers/airtravel/prohibited/permitted-prohibited-items.shtm#4> (last visited Oct. 19, 2009); TSA: Transporting Knitting Needles & Needlepoint, http://www.tsa.gov/travelers/airtravel/assistant/editorial_1252.shtm (last visited Oct. 19, 2009). Snow globes (how deadly!), of course, are a no-no. See TSA: Prohibited Items, <http://www.tsa.gov/travelers/airtravel/prohibited/permitted-prohibited-items.shtm#11> (last visited Oct. 19, 2009).

⁷⁵ Judge Learned Hand once wrote: “There is no surer way to misread any document than to read it literally” *Guisseppi v. Walling*, 144 F.2d 608, 624 (2nd Cir. 1944) (Hand, J., concurring). Perhaps rigid adherence to the text should be discouraged if a literal reading will render an absurd result or the reader is confronting a matter unforeseen by the author. But I do not raise either of those situations.

One of Judge Hand’s best-known law clerks was constitutional scholar Gerald Gunther, who took more than twenty years to research and write what one reviewer for *The New York Times* called “the fullest, most sensitive, most penetrating of judicial biographies.” See Ari L. Goldman, *Gerald Gunther, Legal Scholar, Dies at 75*, N.Y. TIMES, Aug. 1, 2002, available at <http://www.nytimes.com/2002/08/01/us/gerald-gunther-legal-scholar-dies-at-75.html>.

In the biography of his mentor, Professor Gunther recounts this amusing story from his clerkship, concerning Judge Hand’s request for Gunther’s critique of a draft of a dissenting opinion in *United States v. Remington*, 208 F.2d 567 (1953):

After seven weeks, he handed his most recent effort to me. “Now look at this one; see if this one holds water any better.” I studied the new draft for several hours and returned to his desk. Hand looked up eagerly: “Well, will it wash?” I responded that portions of the opinion now seemed reasonably airtight, but there were still weaknesses in other sections. Hand looked at me darkly, pain and annoyance clouding his face. He heaved a deep sigh, then picked up a small paperweight on his desk and threw it in my general direction, missing me by only a narrow margin. “Damn,” he shouted, “I can’t go on forever like this! Thirteen drafts and it’s still not satisfactory? Son, I get paid to decide cases. At some point, I have to get off the fence and turn out an opinion. Enough!” I had never heard Hand speak in such anger. I turned pale and retreated, shaken, to my desk in the adjacent office. I flung my head down on the desk and tried to regain my composure. After a minute or so, I felt a hand gently tapping the back of my head. Judge Hand, in his stocking feet, had silently left his desk, come into my office and hoisted himself to a sitting position on my desk. I raised my face and looked up into his bemused countenance. “Now, now,” he gently consoled, “you can’t take it that way! It’s all part of the job! Don’t take it so hard—you did your job; I have to do mine.”

A party that wins a lawsuit against a debtor may become a “lien creditor,” defined as “a creditor that has acquired a lien on the property involved by attachment, levy, or the like.”⁷⁶ For example, Jane might acquire a lien on Sarah’s jewelry after Jane wins her defamation lawsuit against Sarah, or SmithCorp might obtain a lien on the inventory of JonesCo following SmithCorp’s successful litigation against JonesCo for breach of contract.⁷⁷

As a general rule, a priority dispute between a lien creditor and the Article 9 secured party turns on whether the secured party holds a perfected security interest in the debtor’s property when the competing creditor first becomes a “lien creditor” (*i.e.*, when its lien initially encumbers the debtor’s property).⁷⁸ So, Lender will enjoy priority in Sarah’s jewelry if Lender holds a perfected security interest in the jewelry when Jane acquires her lien, but SmithCorp will enjoy priority if its lien on the JonesCo inventory arises before Lender obtains a perfected security interest in the inventory.

That evening, my wife came to pick me up at the office so that we could go together to a social engagement. She told me that she had encountered Hand on her way in and that he had chided her, “Come on, now, you’ve been married three or four years. Don’t you ever yell at him? You must yell at him once in a while. He’s not used to it. He’s got to get used to people yelling at him!”

GERALD GUNTHER, *LEARNED HAND: THE MAN AND THE JUDGE* 620–21 (1994).

⁷⁶ § 9-102(a)(52)(A). The term also includes certain assignees and receivers, and the dreaded trustee in bankruptcy (also known as “the prince of darkness”). *Id.* at (B)–(D).

⁷⁷ The procedures necessary for the lien to attach to the debtor’s personal property vary among the states:

but ordinarily a creditor first must obtain a judgment against the debtor and then secure a writ of execution ordering the sheriff to levy on the debtor’s property. Some states, of course, allow the creditor to obtain a lien on personal property without first obtaining a judgment as in a prejudgment attachment[.]

HAWKLAND, *supra* note 49, § 9-317:2 at 9-271 (footnote omitted). *See also* Thomas M. Ward, *Ordering the Judicial Process Lien and the Security Interest under Article Nine: Meshing Two Different Worlds (Part I—Secured Parties and Post-Judgment Process Creditors)*, 31 ME. L. REV. 223, 234–64 (1980) (discussing creation, characteristics, and timing of “liens” and resulting “lien creditor” status under different state laws).

⁷⁸ *See* § 9-317(a)(2)(A). The statute states that the Article 9 security interest is subordinate, or junior, to the rights of a person who becomes a lien creditor “before” the security interest is perfected. The statute’s use of “before” effectively gives the Article 9 security interest priority in the case of a tie, which may arise when the lien attaches simultaneously with the moment of perfection. This may occur when the secured party has taken all steps necessary to perfect its security interest in the debtor’s current and future assets, but the debtor has not yet acquired the particular asset subject to the lien.

The statute also awards priority to the secured party who, when the lien attaches, has a security agreement in place that covers the asset and has already filed a financing statement. *See* § 9-317(a)(2)(B). Unlike the rule in (A), the rule in (B) protects the secured party who lacks perfection only because it has yet to give value. *See* §§ 9-203(b)(1) (requiring value as a condition to enforceability); 9-203(a) (indicating that an enforceable security interest “attaches to collateral”); 9-308(a) (requiring attachment as a prerequisite to perfection).

Unless its collateral is a consumer good, a purchase-money creditor (e.g., the dealer who sells on credit, or the third-party financier whose funds are used by the debtor to purchase the item from the dealer) must file a financing statement to perfect its purchase-money security interest.⁷⁹ Whether the purchase-money creditor is the dealer or the third-party financier, the creditor most often will not file its financing statement until sometime after the actual sale occurs and the buyer/debtor takes possession of the asset. Under the general priority rule, then, the purchase-money creditor takes the risk that someone may become a “lien creditor” against the asset before the financing statement is filed. In an effort to accommodate this risk and acknowledge this typical business practice (“sell first, file later”), the general priority rule is expressly subject to the following purchase-money carve out:

[I]f a person files a financing statement with respect to a purchase-money security interest before or within 20 days after the debtor receives delivery of the collateral, the security interest takes priority over the rights of a . . . lien creditor which arise between the time the security interest attaches and the time of filing.⁸⁰

This rule protects the purchase-money creditor from losing priority to a party who becomes a lien creditor against the asset after the security interest attaches to the asset and before the security interest is perfected, if the secured party perfects its security interest by filing a financing statement against the asset no later than the twentieth day after the debtor receives delivery of the asset. On a timeline, then, the important events occur in this order: (1) the purchase-money security interest attaches; (2) a party becomes a “lien creditor” against the asset; and (3) the purchase-money creditor files its financing statement. The only other important event is the date on which the buyer/debtor takes delivery of the asset.

⁷⁹ The technical requirements of a purchase-money security interest are found in § 9-103. Section 9-309(1) states the general rule that a purchase-money security interest in most consumer goods is perfected at attachment (often referred to as “automatic” perfection). The negative implication is that a purchase-money security interest in collateral other than consumer goods—such as inventory and equipment—is not automatically perfected at attachment. Instead, perfection of the security interest requires some additional step, usually the filing of a financing statement.

⁸⁰ § 9-317(e). See also § 9-317(a)(2) (“except as otherwise provided in subsection (e) . . .”); Grant Gilmore, *The Purchase Money Priority*, 76 HARV. L. REV. 1333, 1387 (1963) (referring to “the business practice of filing after delivery in cases of purchase money security interests in collateral other than inventory” as justification for the grace period, and then observing that “if the debtor insists that he must have the goods today, the purchase money financier can deliver them, without sacrificing his . . . priority, provided he perfects” within the prescribed grace period).

All of this seems rather straightforward. So consider the following hypothetical. ABC Florist Company buys, and takes possession of, a new delivery van on May 1. Dealer retains a purchase-money security interest in the van that attaches on that date. Smith, a successful litigant against ABC in a recent lawsuit, becomes a “lien creditor” against the van on May 3. Mindful of applicable state laws governing encumbrances on non-inventory motor vehicles, Dealer does not file a financing statement against the van. Instead, Dealer complies with applicable motor vehicle laws and submits paperwork to the applicable government office, which then issues a certificate of title marked to show Dealer’s security interest. Dealer’s date of perfection is May 18. A few weeks later, Smith and Dealer squabble over the van. *Which creditor has priority in the vehicle?*

Under the general priority rule of § 9-317(a), Smith enjoys priority because she became a “lien creditor” on May 3, several days before Dealer became perfected on May 18. A student may address the relevance of the purchase-money exception with the following syllogism:

A purchase-money creditor must timely file a financing statement to enjoy priority over a lien creditor whose interest arose between attachment and filing.

Dealer, a purchase-money creditor (stipulated), did not file a financing statement.

Dealer does not enjoy priority over a lien creditor whose interest arose between attachment and filing.⁸¹

The syllogism states the rule, analyzes whether facts are present which satisfy the requirements of the rule, and draws a logical conclusion from that analysis. So why is the conclusion wrong? Because the rule cannot be read literally!

The rule protects a purchase-money creditor who timely “files a financing statement.”⁸² Read literally, then, a person who timely complies with certificate-of-title laws cannot invoke the rule. But the drafters of Article 9 did not intend for the rule to be read literally. Instead, the rule is to be read as if it referred to a person who timely “files a financing statement *or complies with the requirements of a statute, regulation, or treaty described in Section 9-311(a).*” An example of a statute mentioned in § 9-311(a) is a “certificate-of-title statute . . . which provides for a

⁸¹ Smith is such a lien creditor, a point hopefully addressed in one or more separate syllogisms.

⁸² § 9-317(e).

security interest to be indicated on the certificate as a condition or result of perfection.”⁸³

How is the student supposed to know that the meaning of “files a financing statement” is something other than its plain meaning? One might assume that the statutory language of § 9-317(e) should be the first place to look for some clue. But no clue exists in subsection (e), or in any of the other subsections of § 9-317. The critical statutory language is found several sections (and pages) earlier. Section 9-311(b) states in relevant part: “Compliance with the requirements of a [certificate-of-title] statute . . . described in subsection (a) for obtaining priority over the rights of a lien creditor is equivalent to the filing of a financing statement under this article.”⁸⁴ But § 9-311 is never cross-referenced in the body of § 9-317! Perhaps the student will stumble across the reference to § 9-311 in the concluding paragraph of the eighth (and final) Official Comment to § 9-317.⁸⁵ But why should the reader not find such an important cross-reference in the statutory language itself?⁸⁶

Article 9 strives to use straightforward, unambiguous, statutory language. But as noted in the preceding analysis, sometimes that language fails to tell the reader the full story. Such incomplete language, when read literally, can lead to incorrect conclusions. This inability to apply the statutory language at face value is yet another reason why mastering the legal principles in a Secured Transactions course through syllogisms can be challenging.

⁸³ § 9-311(a)(2).

⁸⁴ § 9-311(b). *See also* §§ 9-311(a)(2) (requesting states to “list any certificate-of-title statute covering automobiles, trailers, mobile homes, boats, farm tractors, or the like, which provides for a security interest to be indicated on the certificate as a condition or result of perfection”); 9-102(a)(10) (defining “certificate of title”).

⁸⁵ Official Comment 8 to § 9-317 is captioned “Purchase-Money Security Interests.” The first paragraph discusses subsection (e). The second paragraph states:

Section 9-311(b) provides that compliance with the perfection requirements of a statute or treaty described in Section 9-311(a) “is equivalent to the filing of a financing statement.” It follows that a person who perfects a security interest in goods covered by a certificate of title by complying with the perfection requirements of an applicable certificate-of-title statute “files a financing statement” within the meaning of subsection (e).

§ 9-317 cmt. 8.

⁸⁶ The same issue arises in § 9-315, which addresses continued perfection in proceeds. *See* § 9-315(d)(1)(A) (referring to “a filed financing statement”) and cmt. 6 (indicating that compliance with perfection requirements under a statute or treaty described in § 9-311 falls within the intended meaning of the phrase “filed financing statement”).

Article 9 offers evidence that the drafters knew how to cross-reference § 9-311 when discussing filed financing statements. *See, e.g.*, §§ 9-611(c)(3)(B), (C); 9-621(a)(2), (3).

D. “Policy considerations cannot override our interpretation of the text[.]”⁸⁷

In July, Socrates and Francois-Marie Arouet⁸⁸ were in a car accident. Arouet suffered a temporary injury to his back, which greatly affected his mobility. Arouet completed all necessary paperwork and obtained a six-month permit to park in spaces reserved for persons with physical disabilities. By October, Arouet’s mobility had returned to normal. Socrates, relying on the policy underlying the issuance of the six-month permit, believes that Arouet’s parking permit is no longer effective. He is surprised to discover, under the applicable law, that the parking permit remains effective for the full six months from issuance.

Relying on a statute’s underlying policy to determine its proper application may be a logical approach. But where there exists a disconnect between the policy and the text, such statutory construction may lead to incorrect conclusions. And that can happen in a Secured Transactions course.

Students learn early in the semester two basic priority rules: (1) a secured claim beats an unsecured claim; and (2) a perfected secured claim trumps an unperfected secured claim.⁸⁹ After mastering the rules for creating an enforceable security interest,⁹⁰ students turn their attention to the methods of perfection. The most popular method of perfecting a security interest is filing a financing statement with the appropriate state official,⁹¹ the purpose of which “is to give notice to creditors and other interested parties that a security interest exists in property of the debtor.”⁹²

⁸⁷ *Central Bank of Denver, N.A. v. First Interstate Bank of Denver, N.A.*, 511 U.S. 164, 188 (1994).

⁸⁸ Most readers know Arouet by his pen name, “Voltaire.” He once remarked, “Behind every successful man stands a surprised mother-in-law.” See *Voltaire Quotes*, <http://www.brainyquote.com/quotes/authors/v/voltaire.html> (last visited Oct. 24, 2009). He also owned sixty walking sticks, twenty more than another contemporary French philosopher, Jean Jacques Rousseau. See *Canes, So Much More Than Walking Sticks*, http://www.liveauctiontalk.com/free_article_detail.php?article_id=583 (last visited Oct. 24, 2009). Apparently their shared appreciation for the cane was the only matter on which they agreed. See Rit Nosotro, *Voltaire and Rousseau*, <http://www.hyperhistory.net/apwh/essays/comp/cw20/VoltaireRousseau32200123.htm> (last visited Oct. 24, 2009) (noting “their hatred for one another was well-known during their lifetime”).

⁸⁹ Article 9 codifies these two rules in §§ 9-201(a) and 9-322(a)(2), respectively.

⁹⁰ See generally § 9-203(a), (b).

⁹¹ See § 9-310(a) (stating the general rule that “a financing statement must be filed to perfect all security interests”) and cmt. 2 (noting “a central Article 9 principle” that “[f]iling a financing statement is necessary for perfection of security interests”).

⁹² *Peoples Bank v. Bryan Bros. Cattle Co.*, 504 F.3d 549, 559 (5th Cir. 2007) (quoting *In re Glasco, Inc.*, 642 F.2d 793, 795 (5th Cir. 1981)). See also *In re Perry Hollow Mgmt. Co., Inc.*, 297 F.3d 34, 40 (1st Cir. 2002) (referring to “the intended purpose of the filing requirement, which is to give proper notice that a secured interest exists”); *Allstate Fin. Corp. v. United States*, 109 F.3d 1331, 1335 (8th Cir. 1997) (“The purpose of filing

With this background in mind, assume that BankOne perfected a security interest in SmithCorp's current and after-acquired equipment, inventory, and accounts by filing a financing statement in July 2006 with the appropriate Texas official. In February 2007, SmithCorp sold a valuable piece of its equipment to JonesCorp without BankOne's knowledge or consent. In July 2007, SmithCorp changed its legal name to "HarrisCorp."

BankTwo made a secured loan to JonesCorp in November 2008 and obtained and perfected a security interest in JonesCorp's current and after-acquired equipment, inventory, and accounts by filing a financing statement that month with the appropriate Texas official. Before funding the loan, BankTwo searched the U.C.C. records and found no financing statements filed against JonesCorp.

BankThree made a secured loan to HarrisCorp in November 2008 and obtained and perfected a security interest in HarrisCorp's current and after-acquired equipment, inventory, and accounts by filing a financing statement that month with the appropriate Texas official. Before funding the loan, BankThree searched the U.C.C. records and found no financing statements filed against HarrisCorp.

In January 2009, two priority disputes erupt: BankOne and BankTwo are fighting over the piece of equipment sold by SmithCorp to JonesCorp in February 2007, and BankOne and BankThree are contesting the priority in a piece of equipment purchased by HarrisCorp in October 2007.

A student who is asked to resolve these two priority disputes might rule against BankOne in both cases. BankOne's security interest in the two pieces of equipment was perfected by a filed financing statement. BankTwo's search against "JonesCorp" failed to reveal BankOne's filing. BankThree's search against "HarrisCorp" yielded the same result. BankOne's filing failed to achieve its stated purpose, giving notice of its security interest. Because BankOne's filing failed to accomplish its notice function, its filing should be deemed ineffective. Absent an effective filing, BankOne's security interest is unperfected and is subordinate to the perfected claims held by BankTwo and BankThree.

financing statements is to put creditors on notice of existing interests in the debtor's property."); *Nw. Acceptance Corp. v. Lynnwood Equip., Inc.*, 841 F.2d 918, 921 (9th Cir. 1988) (observing that the financing statement "serves the purpose of putting subsequent creditors on notice that the debtor's property is encumbered" (quoting *Thorp Commercial Corp. v. Northgate Indus., Inc.*, 654 F.2d 1245, 1248 (8th Cir. 1981))); § 9-502 cmt. 2 (referring to the system of "notice filing").

Reformatted as a syllogism, the analysis might look like this:

First syllogism (addressing BankOne's continued perfection):

A financing statement that fails to accomplish its "notice" function is ineffective to perfect a security interest.

Searches performed by BankTwo and BankThree reveal that BankOne's financing statement failed to give "notice" of its security interest in equipment either sold by SmithCorp to JonesCorp or purchased by HarrisCorp following the name change from SmithCorp.

BankOne's financing statement, which failed to accomplish its "notice" function, is ineffective to perfect its security interest in the two pieces of equipment.

Second syllogism (addressing priority):

A perfected security interest has priority over an unperfected security interest.⁹³

The secured claims held by BankTwo and BankThree in the equipment are perfected by effective filings; the secured claim held by BankOne in the equipment is no longer perfected because the financing statement has become ineffective.

Perfected security interests held by BankTwo and BankThree enjoy priority over the unperfected security interest held by BankOne.

Whether in narrative form or syllogisms, the analysis—buttressed by policy—and conclusions drawn therefrom are logical. Nevertheless, BankOne enjoys priority over both BankTwo and BankThree. Its financing statement continues to perfect the security interest in the equipment sold by SmithCorp to JonesCorp, under a rule that states: "A filed financing statement remains effective with respect to collateral that is sold . . . and in which a security interest . . . continues."⁹⁴ Its financing statement also continues to perfect the security interest in the equipment purchased by HarrisCorp shortly after the name change from "SmithCorp," under the rule that states: "the financing statement is effective to perfect a security interest in collateral acquired by the debtor before, or within four months after, the [name] change" but "is not effective to perfect a security interest in collateral acquired by the debtor more than four

⁹³ For statutory authority, see § 9-322(a)(2).

⁹⁴ § 9-507(a). Observe that the financing statement remains effective post-disposition only if the underlying security interest survives the disposition, a matter addressed in § 9-315(a). BankOne's security interest in the equipment sold by SmithCorp to JonesCorp remains after the sale because BankOne did not authorize the disposition free of its security interest. See § 9-315 cmt. 2 (second paragraph). The moral of the story for BankTwo is that it "must inquire as to [JonesCorp's] source of title and, if circumstances seem to require it, search in the name of a former owner [e.g., "SmithCorp"]." § 9-507 cmt. 3.

months after the [name] change.”⁹⁵ Because BankOne filed its financing statement in 2006, long before BankTwo and BankThree filed their financing statements in 2008, BankOne continues to enjoy priority in the disputed collateral,⁹⁶ notwithstanding the fact that its continued perfection rests on a financing statement that failed to accomplish its notice function.

This discussion illustrates another possible hurdle to the effective use of syllogisms in Secured Transactions. Article 9 occasionally adopts rules that are inconsistent with policy-based analysis, which may lead to incorrect major premises and, accordingly, erroneous conclusions.

E. “Consistency, madam, is the first of Christian duties.”⁹⁷

Socrates and John Ford⁹⁸ have been discussing films which received the most Academy Award nominations in particular years. Ford mentions that *Gone With The Wind*, *All About Eve*, *Ben-Hur*, *Schindler’s List*, and *Titanic* received the most nominations in their respective years.⁹⁹ Each film also was

⁹⁵ § 9-507(c). This name-change rule requires BankThree to consider the past and the future. BankThree must determine prior names used by HarrisCorp and search for financing statements filed under those prior names (such as “SmithCorp”). BankThree must also monitor HarrisCorp for any future name changes and, if necessary, timely file a name-change amendment within four months of the change. See § 9-507(c)(2) and cmt. 4.

⁹⁶ See § 9-322(a)(1) (stating the general rule that a secured party who files or perfects first enjoys priority).

⁹⁷ Penned by Charlotte Bronte, and spoken by Mr. Brocklehurst to Mrs. Reed in Chapter 4 of *JANE EYRE*. See CHARLOTTE BRONTE, *JANE EYRE* 28 (Bantam Classic 1981) (1847). Perhaps Bronte’s most beloved work, *JANE EYRE* was originally published in 1847 under the pseudonym “Currer Bell.” See Charlotte Bronte, <http://academic.brooklyn.cuny.edu/english/melani/cs6/bronte.html> (last visited Oct. 24, 2009). One of the more famous film adaptations premiered in 1944 and features Joan Fontaine in the title role and Orson Welles as Edward Rochester. A very young Elizabeth Taylor appears in an uncredited role, and the screenplay is written, in part, by Orson Welles’ co-founder of The Mercury Theatre: John Houseman (who, as noted earlier in footnote 5, portrayed Professor Charles Kingsfield many years later). See *Jane Eyre* 1944, <http://www.imdb.com/title/tt0036969/> (last visited Oct. 24, 2009); Art Pierce, Orson Welles, The Mercury Theatre, and The Campbell Playhouse, <http://www.mercurytheatre.info/history> (last visited Oct. 24, 2009).

⁹⁸ John Ford won the Oscar for Best Director four times, a record. The films for which he won the award are *The Informer*, *The Grapes of Wrath*, *How Green Was My Valley*, and *The Quiet Man*. He also received a nomination for *Stagecoach*. Although no one has received as many wins, other famous directors have garnered more nominations, including Woody Allen (six), Frank Capra (six), Martin Scorsese (six), Steven Spielberg (six), Fred Zinnemann (seven), David Lean (seven), Billy Wilder (eight), and William Wyler (twelve). See Best Director Facts & Trivia, <http://www.filmsite.org/bestdirs.html> (last visited Oct. 24, 2009).

⁹⁹ See All About Oscar: Academy Award Trivia and Statistics, <http://www.reelclassics.com/Articles/General/oscar-trivia-article.htm> (last visited Oct. 24, 2009) (*All About Eve* and *Titanic* tied with fourteen nominations in their respective years while *Gone With the Wind* was a close runner-up with thirteen.); *Ben-Hur* 1959, <http://www.imdb.com/title/tt0052618/awards> (last visited Oct. 24, 2009) (accumulating twelve nominations); 1993 Academy Awards Winners and History, <http://www.filmsite>

named “Best Picture.”¹⁰⁰ Ford also informs Socrates that *A Streetcar Named Desire*, *Mary Poppins*, *Who’s Afraid of Virginia Woolf?*, *Reds*, and *Tootsie* received the most nominations in their respective years.¹⁰¹ These films were nominated for, but did not win, the “Best Picture” award.¹⁰² Ford then tells Socrates that *Dreamgirls* led all films released in 2006 with eight nominations.¹⁰³ Based on the preceding information, Socrates logically concludes that *Dreamgirls* was a “Best Picture” nominee, if not a winner. Socrates is wrong.¹⁰⁴ *Dreamgirls* is the only film to receive the most Academy Awards nominations in its year of eligibility and not receive a “Best Picture” nomination.¹⁰⁵ So much for precedent!

Article 9 has been known to address a specific concern with a series of rules that render a consistent and predictable result—except, for some unexplained reason, in one instance. As happened to Socrates in the preceding paragraph, this unanticipated departure from what all other related rules suggest is the norm may lead students to logically reach an incorrect conclusion.¹⁰⁶

.org/aa93.html (last visited Oct. 24, 2009) (*Schindler’s List* tied *Ben-Hur* with twelve nominations).

¹⁰⁰ See Academy Awards, USA, http://www.imdb.com/Sections/Awards/Academy_Awards_USA/ (last visited Oct. 24, 2009).

¹⁰¹ See 1951 Academy Awards Winners and History, <http://www.filmsite.org/aa51.html> (last visited Oct. 24, 2009) (*A Streetcar Named Desire* earned twelve nominations); 1964 Academy Awards Winners and History, <http://www.filmsite.org/aa64.html> (last visited Oct. 24, 2009) (*Mary Poppins* earned thirteen nominations); 1966 Academy Awards Winners and History, <http://www.filmsite.org/aa66.html> (last visited Oct. 24, 2009) (*Who’s Afraid of Virginia Woolf?* also earned thirteen nominations); 1981 Academy Awards Winners and History, <http://www.filmsite.org/aa81.html> (last visited Oct. 24, 2009) (*Reds* accumulated twelve nominations); 1982 Academy Awards Winners and History, <http://www.filmsite.org/aa82.html> (last visited Oct. 24, 2009) (*Tootsie* had ten nominations).

¹⁰² The respective “Best Picture” films were: *An American in Paris* (1951 Academy Awards Winners and History, *supra* note 101); *My Fair Lady* (1964 Academy Awards Winners and History, *supra* note 101); *A Man For All Seasons* (1966 Academy Awards Winners and History, *supra* note 101); *Chariots Of Fire* (1981 Academy Awards Winners and History, *supra* note 101); and *Gandhi* (1982 Academy Awards Winners and History, *supra* note 101).

¹⁰³ See 2006 Academy Awards Winners and History, <http://www.filmsite.org/aa06.html> (last visited Oct. 24, 2009).

¹⁰⁴ The five “Best Picture” nominees in 2006 were *The Departed* (winner), *Babel*, *Letters From Iwo Jima*, *Little Miss Sunshine*, and *The Queen*. See 2006 Academy Awards Winners and History, *supra* note 103.

¹⁰⁵ See No Best Picture Nomination for ‘Dreamgirls,’ <http://www.msnbc.msn.com/id/16767796/> (last visited Oct. 24, 2009).

¹⁰⁶ Asking a student to craft a rule by observing other rules may suggest that I am moving from deductive reasoning to inductive reasoning. Judge Aldisert describes inductive reasoning as:

a form of logic in which big, general principles are divined from observing the outcomes of many small events. In this form of inductive logic, you reason from multiple particulars to the general Thus, inductive reasoning is a

As a general rule, a filed financing statement is effective for five years.¹⁰⁷ Students may be asked to analyze whether certain post-filing events might shorten the five-year duration of effectiveness. Those events may include the following:

- * the debtor changes its name from “SmithCorp” to “JonesCorp”;
- * the debtor relocates its entire business from Houston to Dallas (the same state of the original filing) or Chicago (a different state from the original filing);
- * the debtor converts several units of “inventory” (collateral) to in-house use as “equipment” (not collateral); and
- * without the secured party’s consent, the debtor disposes of some collateral by selling its chattel paper to FinCo (incorporated under the laws of the same state of the original filing) and its accounts to PrimeCo (incorporated under the laws of a different state).

The name change has no immediate impact on the continued effectiveness of the original filing. Under the applicable rule, the filing remains effective through the end of the five-year period to perfect a security interest in collateral acquired by the debtor either before the name change or within four months after the name change.¹⁰⁸ The filing, however, will not remain effective to perfect a security interest in collateral acquired by the debtor more than four months after the name change unless the secured party files a name-change amendment within that four-month period.¹⁰⁹

The debtor’s in-state change of address has no immediate or future impact on the continued effectiveness of the original filing,¹¹⁰ presumably because a search against the debtor’s name after the address change will continue to reveal the original filing.¹¹¹ Whether an out-of-state address change may impact the effectiveness of an original filing is dictated by the type of debtor.

logic of probabilities and generalities, not certainties. It yields workable rules, but not proven truths.

Aldisert, *supra* note 3, at 12–13. Because my concern is with an existing statutory provision, rather than with the creation of a gap-filling rule that may be subject to some discretionary application, I do not believe that I have moved completely into the realm of inductive reasoning. That certainly is not my intent.

¹⁰⁷ See § 9-515(a).

¹⁰⁸ § 9-507(c)(1).

¹⁰⁹ § 9-507(c)(2). A name-change amendment filed more than four months after the name change is effective to perfect a security interest in collateral acquired by the debtor more than four months after the name change, but the effectiveness does not commence until the filing date of the amendment. *Id.* at cmt. 4. This gap in perfection exposes the secured party to a risk of nonpriority against former subordinate creditors who have climbed up the priority ladder.

¹¹⁰ § 9-507(b).

¹¹¹ Financing statements are indexed “according to the name of the debtor.” § 9-519(c)(1).

If the debtor is a corporation, limited partnership, or other “registered organization,”¹¹² then a mere address change has no effect. The correct place to file a financing statement against a registered organization is with the chartering state, whether or not the entity has a physical presence in that state,¹¹³ so the original filing continues to give notice. But notice is frustrated if the out-of-state relocation involves a natural person, or an organization other than a registered organization (*e.g.*, a general partnership), because the proper place to file a financing statement is the state in which the debtor has its principal residence or place of business, respectively.¹¹⁴ The financing statement will remain effective, but only until the end of the four-month period following the relocation; thereafter, the financing statement is ineffective.¹¹⁵

The debtor’s change in the use of its property which converts collateral (*e.g.*, “inventory”) to non-collateral (*e.g.*, “equipment”) has no effect on the continued effectiveness of the financing statement (which lists “inventory” but not “equipment” as collateral).¹¹⁶

The debtor’s sale of assets to an in-state purchaser has no impact on the original filing. The debtor’s sale of assets to an

¹¹² See § 9-102(a)(70) (defining “registered organization”) and cmt. 11 (listing corporations, limited liability companies, and limited partnerships as examples).

¹¹³ See §§ 9-301(1) (indicating that “while a debtor is located in a jurisdiction, the local law of that jurisdiction governs perfection”); 9-307(e) (stating that a registered organization “organized under the law of a State is located in that State”).

¹¹⁴ See §§ 9-301(1); 9-307(b)(1) (individuals), (2) (organizations with one place of business). If the debtor is an organization with multiple business locations, the debtor is located in the state of its “chief executive office.” § 9-307(b)(3) and cmt. 2 (defining “chief executive office” as “the place from which the debtor manages the main part of its business operations or other affairs” and “is the place where persons dealing with the debtor would normally look for credit information”).

¹¹⁵ § 9-316(a)(2). If the four-month period ends after the standard five-year period of effectiveness, then the out-of-state relocation renders the financing statement ineffective at the end of that five-year period. § 9-316(a) (stating that perfection continues “until the earlier of: (1) the time perfection would have ceased under the law of that jurisdiction; [and] (2) the expiration of four months after a change of the debtor’s location to another jurisdiction”).

¹¹⁶ § 9-507(b) (indicating that, unless either the debtor has changed its name or the facts involve a “new debtor” and application of § 9-508, “a financing statement is not rendered ineffective if, after the financing statement is filed, the information provided in the financing statement becomes seriously misleading”). See also § 9-338 cmt. 3 (“Except as provided in Section 9-507 with respect to changes in the debtor’s name, an otherwise effective financing statement does not become ineffective if the information contained in it becomes inaccurate.”).

Implicit in the analysis is the assumption that the property is never subject to a certificate-of-title law. A financing statement will perfect a security interest in a dealer’s inventory of motor vehicles. § 9-311(d). If, however, the dealer converts one of the vehicles to in-house use as equipment, the financing statement no longer perfects the security interest in that vehicle. Instead, the secured party must comply with the certificate-of-title law. See § 9-311 cmt. 4 (last sentence).

out-of-state purchaser does impact the original filing, although the impact is not immediate. The underlying security interests in the chattel paper and accounts survive the dispositions,¹¹⁷ and the original filing continues to perfect those security interests,¹¹⁸ even though a search against the purchasers will not reveal the filing against the seller/debtor. The financing statement remains effective to perfect the security interest in the chattel paper sold to the in-state purchaser until the end of its standard five-year duration. But the filing remains effective to perfect the security interest in the accounts sold to the out-of-state buyer for only one year.¹¹⁹

All of these post-filing events frustrate the notice function in whole (the financing statement cannot be found) or in part (the filing can be found but contains incorrect information). Nevertheless, these examples illustrate that post-filing mischief never has an *immediate* impact on the continued effectiveness of the financing statement. At best, the filing remains effective for the remainder of its standard five-year life; at worst, the filing remains effective for a grace period of either four months or one year. A secured party, therefore, should periodically monitor its debtor's behavior, but it has no concerns that any misbehavior will abruptly terminate the effectiveness of its filing.¹²⁰

So consider one more example. Lender makes a \$1 million loan to SmithCorp, a Texas corporation, secured by an enforceable security interest in SmithCorp's current and after-acquired inventory. Lender files its financing statement in Texas. Two years later, SmithCorp merges with and into WilliamsCorp, a Delaware corporation. WilliamsCorp, the

¹¹⁷ § 9-315(a)(1).

¹¹⁸ § 9-507(a). Nevertheless, the purchaser may have a superior claim to the chattel paper under § 9-330(a) or (b).

¹¹⁹ § 9-316(a)(3) (stating that, unless the standard five-year life of a filing will end earlier and trigger application of subsection (a)(1), a security interest remains perfected until "the expiration of one year after a transfer of collateral to a person that thereby becomes a debtor and is located in another jurisdiction"); § 9-316 cmt. 2 (observing that the longer one-year period (rather than a four-month period) "is necessary, because, even with the exercise of due diligence, the secured party may be unable to discover that the collateral has been transferred to a person located in another jurisdiction"). The purchaser of the accounts, PrimeCo, has become a "debtor" by acquiring an interest in property that remains "collateral" for the debt owed by the seller to its creditor. § 9-102(a)(28)(A). PrimeCo is "located in another jurisdiction" because it is incorporated under the laws of a state other than the state in which the financing statement was filed against the seller.

¹²⁰ The central issue is whether mischief which undermines the notice function of the filing creates a risk that should be borne by the filer, or the searcher. Perhaps the best discussion of this issue is found in PEB Commentary No. 3, dated March 10, 1990 (reprinted in THE AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE 1156-59 (2009)), which addresses whether unauthorized asset dispositions should create a refiling obligation for the filer, or a title-tracing obligation for the searcher.

surviving entity, acquires all of SmithCorp's assets and assumes all of SmithCorp's debts. Three months later, WilliamsCorp makes an inventory purchase of \$100,000. *Does Lender hold a perfected security interest in this post-merger inventory?*

Initially, an astute student may question whether Lender can claim *any* security interest in the inventory purchased by WilliamsCorp, as Lender's security agreement was with SmithCorp, not WilliamsCorp. And absent a security agreement, no security interest exists.¹²¹ But WilliamsCorp's acquisition of SmithCorp's assets and assumption of its debts makes WilliamsCorp a "new debtor,"¹²² so SmithCorp's security agreement binds WilliamsCorp and no new security agreement is necessary to create an enforceable security interest in the post-merger inventory acquired by WilliamsCorp.¹²³

The student then addresses the question asked: Does Lender have a *perfected* security interest in the post-merger inventory? Lender's filing is recorded against "SmithCorp" and will not be found by searching against "WilliamsCorp." Additionally, Lender filed its financing statement where SmithCorp is incorporated (Texas), but WilliamsCorp is incorporated in a different state (Delaware). Notwithstanding these concerns, however, the student may logically conclude—based on earlier observations of the effect, if any, that post-filing mischief may have on a financing statement—that Lender's filing remains effective for at least four months (if not longer). The student may even cite § 9-508, which states that the financing statement remains effective to perfect a security interest in inventory acquired by WilliamsCorp within four months after the merger, even though the difference between the name of the original debtor—"SmithCorp"—and the new debtor—"WilliamsCorp"—renders the filing seriously misleading.¹²⁴

But that result is wrong, although the error is far from obvious.

¹²¹ See § 9-203(b)(3)(A)–(D).

¹²² See §§ 9-102(a)(56) (defining "new debtor" as "a person [WilliamsCorp] that becomes bound as debtor under Section 9-203(d) by a security agreement previously entered into by another person [SmithCorp]"); 9-203(d) (stating that a "person [WilliamsCorp] becomes bound as debtor by a security agreement entered into by another person [SmithCorp] if, by operation of law other than this article or by contract . . . (2) the person [WilliamsCorp] becomes generally obligated for the obligations of the other person [SmithCorp] . . . and acquires . . . all or substantially all of the assets of the other person [SmithCorp]").

¹²³ See § 9-203(e) and cmt. 7.

¹²⁴ See § 9-508(b). Lender can perfect a security interest in inventory acquired after the four-month period by filing a financing statement against "WilliamsCorp" in Delaware. § 9-508(b)(2).

The four-month rule of § 9-508, codified in subsection (b) and paraphrased above, applies only if “a filed financing statement . . . is effective under subsection (a).”¹²⁵ Subsection (a) states in relevant part:

[A] filed financing statement naming an original debtor [SmithCorp] is effective to perfect a security interest in collateral in which a new debtor [WilliamsCorp] has or acquires rights *to the extent that the financing statement would have been effective had the original debtor [SmithCorp] acquired rights in the collateral.*¹²⁶

What does the quoted language mean? I read it as requiring the collateral acquired by the new debtor to fall within the description of the collateral found in the financing statement filed against the original debtor (an interpretation shared by one of the drafters¹²⁷). For example, a filing against SmithCorp’s “inventory” is effective to perfect a security interest in WilliamsCorp’s “inventory.” The language cannot address any concern with the difference between the names of the original debtor (“SmithCorp”) and the new debtor (“WilliamsCorp”), because that matter is expressly addressed by subsection (b). And if the language is intended to address any jurisdictional differences between the original debtor (located in Texas) and the new debtor (located in Delaware), one might reasonably argue that the drafters should have—and could have—been less obtuse,¹²⁸ especially since the difference in jurisdictions does not prevent the filing from remaining effective to perfect a security interest in the collateral transferred from the original debtor to the new debtor for as long as one year.¹²⁹ It will be the rare student who reads the statutory language of § 9-508 and appreciates that the change in jurisdiction *immediately* renders

¹²⁵ § 9-508(b).

¹²⁶ § 9-508(a) (emphasis added).

¹²⁷ See Harry C. Sigman, *The Filing System Under Revised Article 9*, 73 AM. BANKR. L.J. 61, 81 (1999) (noting that § 9-508 continues the effectiveness of a filed financing statement against a new debtor “to the extent that the financing statement would have been effective had the original debtor acquired rights in the collateral, e.g., *the collateral falls within the indication of collateral provided in the financing statement*”) (emphasis added); *id.* at 61 n.* (indicating that Mr. Sigman was the ALI representative on the NCCUSL Article 9 Drafting Committee).

¹²⁸ For example, subsection (a) § 9-508 could have been drafted as follows (revised language italicized):

Except as otherwise provided in this section, a filed financing statement naming an original debtor is effective to perfect a security interest in collateral in which a new debtor, *located in the same jurisdiction as the original debtor*, has or acquires rights to the extent that the financing statement would have been effective had the original debtor acquired rights in the collateral.

Alternatively, the following sentence could have been added to the end of subsection (c): “This section also does not apply if the original debtor and the new debtor are located in different jurisdictions.”

¹²⁹ See §§ 9-508(c), 9-507(a), and 9-316(a)(3).

the filing ineffective to perfect a security interest in any post-merger collateral. But buried in the Official Comments to § 9-508 is evidence of that intended result:

Moreover, if the original debtor and the new debtor are located in different jurisdictions, a filing against the original debtor would not be effective to perfect a security interest in collateral that the new debtor acquires or has acquired from a person other than the original debtor. See Example 5, Section 9-316, Comment 2.¹³⁰

Whether confronted with the confusing language of § 9-508(a), or the clear message tucked away in Official Comment 4, the student is left to revise her earlier analysis and conclude that Lender's original filing is not effective to perfect its security interest in any post-merger inventory acquired by WilliamsCorp.

The foregoing analysis reveals yet another possible obstacle to the effective use of syllogisms in Secured Transactions. Article 9 occasionally adopts a rule that, for no apparent reason, yields a result so unexpected and inconsistent with the norm that a logical approach to its understanding is undermined.

F. "If the law supposes that," said Mr. Bumble, . . . "the law is a ass—a idiot."¹³¹

While attending the Houston Livestock Show and Rodeo,¹³² Socrates purchases a pair of cowboy boots for \$200.¹³³ After

¹³⁰ § 9-508 cmt. 4. As Kenny Bania might exclaim, "That's gold, Jerry! Gold!" See *The Fatigues*, <http://www.seinfeldscripts.com/TheFatigues.html> (last visited Oct. 24, 2009).

¹³¹ CHARLES DICKENS, *OLIVER TWIST* 463 (The New American Library, Inc. 1961) (1838). Associate Justice Harry A. Blackmun is one of two members of the Supreme Court of the United States to drop this quote into an opinion. See *Estate of Wilson v. Aiken Industries, Inc.*, 439 U.S. 877, 880 n.3 (1978) (Blackmun, J., concurring). (Justice John Paul Stevens also borrowed the line in *Califano v. Goldfarb*, 430 U.S. 199, 223–24 n.10 (1977) (Stevens, J., concurring).) Justice Blackmun took his judicial oath on June 9, 1970, exactly one hundred years to the day of the death of Charles Dickens. See *Members of the Supreme Court of the United States*, <http://www.supremecourtus.gov/about/members.pdf> (last visited Oct. 21, 2009); Charles Dickens Biography and Works, <http://www.onlineliterature.com/dickens/> (last visited Oct. 21, 2009).

¹³² According to its website, the HLSR is the third largest fair or festival in North America (the general attendance in 2009 exceeded 1.8 million) and the largest rodeo in the world (26,285 livestock competitions and horse show entries in 2009). See *Frequently Asked Questions Houston Livestock Show and Rodeo*, <http://www.hlsr.com/about/faq.aspx> (follow the "How does the Show impact the city?" hyperlink and the "What's unique about the Houston Livestock Show?" hyperlink) (last visited Oct. 21, 2009). Entertainers from "A" to "Z" have performed on its concert stages through the years, including Ashanti, Tony Bennett, Johnny Cash, Bill Cosby, Bob Dylan, Alicia Keys, Barry Manilow, Olivia Newton-John, Elvis Presley, REO Speedwagon, Randy Travis, and ZZ Top. See *Entertainment History Houston Livestock Show and Rodeo*, <http://www.hlsr.com/concerts/past-performers.aspx> (last visited Oct. 21, 2009).

¹³³ Can you believe that someone would pay \$75,000 for a pair of cowboy boots? See Neal Santelmann, *Most Expensive Cowboy Boots*, April 4, 2006, http://www.forbes.com/2006/04/03/cowboy-boots-expensive_cx_ns_0404feat.html.

returning to his hotel, Socrates decides to take a walk in his new footwear. While waiting for a streetlight to change, a law enforcement officer hits the siren, jumps out of the car, and arrests Socrates. Dumbfounded, Socrates asks, "What's the charge, officer?" The officer replies, "It is a misdemeanor in Texas to wear cowboy boots while adorned in a toga." And Socrates retorts, "Well that's just silly."¹³⁴

Whatever your word choice (silly, stupid, idiotic, dumb, etc.), some laws just make no sense. When learning those laws, it may be best to park your deductive reasoning at the classroom door. That suggestion just may be appropriate when teaching one particular (and, in my opinion, "silly") provision of Article 9. Consider the following hypothetical.

Meredith, an architect, wants to buy a Steinway grand piano for her home. She cannot afford to buy a new piano from her local Steinway dealer,¹³⁵ so she has been reviewing advertisements posted by private owners in the weekend newspaper. Meredith visits the home of one seller, whose recently-deceased spouse was the family musician who played for personal satisfaction. After playing the piano for twenty minutes, she agrees to pay the asking price of \$22,000 and takes delivery that week. Eight months later, Friendly Finance contacts Meredith and informs her that it financed the seller's purchase of the piano and continues to hold a perfected security interest in it. Friendly Finance informs Meredith that if she does not pay the seller's unpaid debt of \$14,000, then Friendly Finance will exercise its contractual and statutory rights and remedies, including repossessing and selling the piano. Meredith had no prior knowledge of her seller's financing arrangement or Friendly Finance's security interest. She has no intention of paying \$14,000 to make this problem disappear and tells Friendly Finance to pursue collection from her seller, who

¹³⁴ To my limited knowledge, Socrates has not violated any actual law. Apparently, however, Texas does have (or has had) some "silly" laws. See *Stupid Texas Laws*, http://www.idiotlaws.com/dumb_laws/texas/ (last visited Oct. 21, 2009).

¹³⁵ Steinway & Sons, formed in 1853, builds approximately 5,000 pianos each year, and each piano takes about one year to craft. See *Steinway Piano: Through the Years*, May 21, 2008, <http://www.articlesbase.com/music-articles/-steinway-piano-through-the-years-421850.html>. I purchased a new Steinway baby grand piano in the mid-1990's for approximately \$24,000, so a new Steinway grand piano that the typical consumer would purchase for personal use probably costs at least \$30,000–\$35,000 today. Several years ago, singer George Michael paid over \$2 million for the upright Steinway piano on which John Lennon composed "*Imagine*." See Stephen M. Silverman, *George Michael Buys Lennon Piano*, PEOPLE MAGAZINE, Oct. 23, 1997, <http://www.people.com/people/article/0,,621306,00.html>. Readers interested in learning more about Steinway pianos may wish to read RONALD V. RATCLIFFE, *STEINWAY* (1989) or view the documentary movie *Note by Note: The Making of the Steinway L1037* (Plow Productions, LLC 2007).

received \$22,000 from Meredith. *Did Meredith acquire the Steinway grand piano subject to, or free and clear of, Friendly Finance's security interest?*

Most students will favor Meredith in this dispute. She purchased the piano for a fair price, and she had no knowledge of the security interest. Meredith is an innocent purchaser for value, so Steinway's recourse should be limited to the seller, who received sales proceeds more than sufficient to pay off the purchase debt.¹³⁶

Sounds great, right? That may be the result under Article 9. Then again, it may not. The correct result probably will turn on a single factor: whether Friendly Finance filed a financing statement.

As a general rule, a security interest remains effective against purchasers of the collateral, unless the secured party has authorized the sale free of (rather than subject to) its security interest.¹³⁷ Apparently Friendly Finance did not consent to its debtor's sale of the piano, so its security interest survived the disposition unless Article 9 provides buyers with protection.

The primary carve outs for buyers of goods are stated in § 9-320. Subsection (a) states that "a buyer in the ordinary course of business . . . takes free of a security interest created by the buyer's seller, even if the security interest is perfected and the buyer knows of its existence."¹³⁸ This rule permits consumers and non-consumers to purchase items at Wal-Mart, Target, Home Depot, Barnes & Noble, and other retailers without being concerned that the items serve as inventory collateral and may be subject to superior property claims held by an unknown creditor.¹³⁹ Meredith's seller was not in the business of selling pianos, so subsection (a) offers her no protection.

¹³⁶ As a general rule, a creditor has an enforceable (and often perfected) security interest in proceeds traceable to its collateral. See §§ 9-203(f); 9-315(a)(2), (c), (d). See also § 9-102(a)(64) (defining "proceeds").

¹³⁷ See §§ 9-201(a) (stating the general rule that "a security agreement is effective according to its terms . . . against purchasers of the collateral"); 9-315(a)(1) (stating the general rule that "a security interest . . . continues in collateral notwithstanding sale . . . or other disposition thereof unless the secured party authorized the disposition free of the security interest").

¹³⁸ § 9-320(a). See also §§ 1-201(b)(9) (defining "buyer in ordinary course of business"); 9-102(c) (incorporating Article 1 definitions).

¹³⁹ The buyer who is purchasing a unit of inventory may not need the protection of § 9-320(a). Many secured parties consent to inventory dispositions free and clear of their security interests; those security interests, then, effectively terminate at the point of sale. See § 9-315(a)(1). See also § 9-320 cmt. 6 ("If the secured party authorized the sale in an express agreement or otherwise, the buyer takes free under Section 9-315(a) without regard to the limitations of this section.").

But perhaps Meredith is protected by subsection (b), which permits her to acquire the piano free of the security interest held by Friendly Finance if five conditions are met. First, her seller must have used the piano “primarily for personal, family, or household purposes.”¹⁴⁰ Rephrased, the piano must have been a “consumer good” in the hands of her seller.¹⁴¹ Meredith can prove this condition, as the seller’s spouse played the piano for personal enjoyment. Second, Meredith must have bought the piano without knowledge of Friendly Finance’s security interest,¹⁴² a fact which she can prove. Third, she must give “value,”¹⁴³ met by her \$22,000 payment. Fourth, the piano must be a “consumer good” in Meredith’s hands.¹⁴⁴ Meredith is an architect, not a professional musician, so she can satisfy this condition. And fifth, Meredith must have purchased the piano before Friendly Finance filed a financing statement against it.¹⁴⁵

Subsection (b), sometimes referred to as the “garage sale” exception,¹⁴⁶ applies only if the debtor sells a consumer good. The overwhelming number of security interests in consumer goods are purchase-money security interests.¹⁴⁷ Because Friendly Finance financed its debtor’s purchase of the piano, which secures repayment of the credit extended, its security interest is a purchase-money security interest.¹⁴⁸ The debtor used the piano as a consumer good, so the purchase-money security interest was automatically perfected at attachment.¹⁴⁹ Therefore, Friendly Finance’s sole purpose in filing a financing statement is *not to perfect* its security interest, *but to protect* its security interest—

¹⁴⁰ § 9-320(b).

¹⁴¹ See § 9-102(a)(23) (defining “consumer goods” as “goods that are used or bought for use primarily for personal, family, or household purposes”).

¹⁴² § 9-320(b)(1).

¹⁴³ *Id.* at (b)(2). See also §§ 1-204 (defining “value” to include “any consideration sufficient to support a simple contract”); 9-102(c) (incorporating definitions from Article 1).

¹⁴⁴ § 9-320(b)(3).

¹⁴⁵ *Id.* at (b)(4).

¹⁴⁶ See WILLIAM H. LAWRENCE, ET AL., UNDERSTANDING SECURED TRANSACTIONS 268 (4th ed. 2007) (observing that § 9-320(b) is “sometimes referred to as the ‘garage-sale’ rule”); Robert M. Lloyd, *The New Article 9: Its Impact on Tennessee Law (Part II)*, 67 TENN. L. REV. 329, 342 (2000) (noting that former § 9-307(2)—the predecessor statute to § 9-320(b)—was “commonly called the ‘garage sale exception.’”). Professor Lloyd notes the irony of the label: “In spite of its name, the rule is not important in the garage sale situation, where the lender is unlikely to try to track down its collateral. Instead, the rule becomes important when consumers buy and sell big-ticket items like boats, expensive guns, or jewelry.” *Id.* at 342.

¹⁴⁷ There are at least two reasons for this, one practical and the other legal. First, most consumer goods have insufficient value to service any debt other than the purchase-money debt. Second, a secured party who claims a non-PMSI in a household good may incur the wrath of the Federal Trade Commission. See 16 C.F.R. § 444.2(a)(4) (2009).

¹⁴⁸ See § 9-103(a), (b).

¹⁴⁹ See § 9-309(1).

against a possible property claim asserted under § 9-320(b) by an unauthorized consumer purchaser of the collateral, such as Meredith.

So Meredith wins if Friendly Finance did not file, but she loses if Friendly Finance filed.

But why should Meredith's victory or defeat turn on whether Friendly Finance filed a financing statement? The purpose of filing a financing statement is to give notice of a possible security interest in the collateral.¹⁵⁰ Implicit in that policy, however, is the premise that the party to whom notice is directed *will search the public records and discover the filing*. But the average consumer who is purchasing an expensive item—from a neighbor, a co-worker, or a stranger, and in response to a casual conversation, a bulletin-board announcement, a newspaper advertisement, or an online solicitation—has never heard of U.C.C. Article 9 (I'm shocked!) or appreciate (much less know where and how to discover) that a financing statement ("financial statement?"¹⁵¹) may have been filed against the item. Article 9 should indeed resolve the potential dispute between Friendly Finance and Meredith, but Meredith's victory or defeat should not turn on a fortuitous event whose purpose is divorced from reality.¹⁵²

¹⁵⁰ See *supra* note 92.

¹⁵¹ As I grade Secured Transactions exams I tell myself that the one error for which I will subtract points is an incorrect reference to a "*financial* statement" instead of a "*financing* statement." But then I recall that judges often make the same mistake. See, e.g., *HomeAlert Corp. v. Concert Co.*, No. 08-912, 2008 WL 5134756, at *2 (W.D. Pa. Dec. 5, 2008) (referring to "the filing of a UCC Financial Statement Amendment"); *Whoops Transport, Inc. v. Blue Ribbon Commodity Traders, Inc.*, 2008 WL 2367213, at *1 (D. Puerto Rico June 6, 2008) ("Bancorp filed a financial statement concerning the security interest"); *Carlmont Capital Special Purpose Corp. II v. Healthcare for Women, S.C.*, No. 2:07-cv-1018-RLH-GWF, 2008 WL 150372, at *2 (D. Nev. Jan. 10, 2008) ("HfW filed UCC-1 financial statements"); *Heck v. Buhler*, 07-21-C, 2007 WL 4256967, at *4 n.5 (M.D. La. Nov. 30, 2007) ("he filed his UCC-1 Financial Statement"); *U.S. v. Pinkston*, SA-06-CA-0732 OG (NN), 2007 WL 1437690, at *2 (W.D. Tex. May 14, 2007) ("the government presented the UCC Financial Statement"); *Agrilience, L.L.C. v. Farmpro Serv., Inc.*, 328 F. Supp. 2d 958, 966 (S.D. Iowa 2003) ("filed the appropriate financial statement"); *EH Yacht, LLC v. Egg Harbor, LLC*, 84 F. Supp. 2d 556, 560 n.2 (D.N.J. 2000) ("the UCC financial statement"); *21 West Lancaster Corp. v. Main Line Restaurant, Inc.*, 614 F. Supp. 202, 203 (E.D. Pa. 1985) ("financial statements were timely filed"); *U.S. v. Trans-World Bank*, 382 F. Supp. 1100, 1103 (C.D. Cal. 1974) ("a financial statement filed with the California Secretary of State"); *In re Tamis*, 398 B.R. 124, 126 (Bankr. D.N.J. Dec. 17, 2008) ("The Chase security interest was perfected by the recordation of a UCC-1 financial statement"); *1st Bank v. Winderl*, 60 P.3d 998, 999 (Mont. 2002) ("two financial statements filed in 1990 and 1994"); *LBM, Inc. v. Rushmore State Bank*, 543 N.W.2d 780, 785 (S.D. 1996) ("a valid and filed financial statement"); *Texas Beef Cattle Co. v. Green*, 883 S.W.2d 415, 425 (Tex. Ct. App. 1994) ("filed financial statements"); *Simmons Oil Corp. v. Holly Corp.*, 796 P.2d 189, 197 (Mont. 1990) ("Holly filed UCC financial statements in Montana.").

¹⁵² See Linda J. Peltier, *Buyers of Used Goods and the Problem of Hidden Security Interests: A New Proposal to Modify Section 9-307 of the Uniform Commercial Code*, 36

The preceding discussion illustrates that Article 9 occasionally adopts a rule that appears to be so illogical as to render futile any attempt to understand it logically. And illogical rules can frustrate the ability to learn the core principles in a Secured Transactions course through syllogisms.

G. “There is no exception to the rule that every rule has an exception.”¹⁵³

Socrates and legendary actor Clark Gable are discussing the “leap year” phenomenon. Clark is a reluctant participant, given that on February 29, 1940, he failed to win the Oscar for Best Actor for his portrayal of Rhett Butler in *Gone With The Wind*.¹⁵⁴

HASTINGS L.J. 215, 253 n.164 (1985) (observing that “the outcome under § 9-307(2) [the predecessor statute to § 9-320(b)] would seem more dependent upon fortuity than reason”).

¹⁵³ James Thurber (1894-1961). See James Thurber Quotes, http://www.brainyquote.com/quotes/authors/j/james_thurber_3.html (last visited Oct. 21, 2009). Thurber had two brothers, one of whom shot James in the eye with an arrow during a game of “William Tell,” leading to blindness in that eye. See The Thurber House, <http://www.thurberhouse.org/james/james.html> (last visited Oct. 21, 2009). Almost ninety years earlier, another little boy—who later became famous throughout the world—had an eye accident. The boy was only three years old when he entered the workshop of his father, a harness maker. As he had seen his father do on many occasions, the boy took a sharp tool and tried to cut a piece of leather. The tool slipped and entered his eye. The injured eye became infected, and the infection spread to his other eye, leaving the boy blind. But his blindness led him, at the age of fifteen, to develop a system of communication that has been used by millions of blind people to read and write. The little boy’s name? Louis Braille. See A New Method: The Story of Louis Braille, <http://louisbrailleschool.org/resources/louis-braille/louis-braille-childhood/> (last visited Oct. 21, 2009).

¹⁵⁴ The award went to Robert Donat, for his portrayal of a shy “classics” teacher at an English boarding school, in *Goodbye, Mr. Chips*. See 1939 Academy Awards Winners and History, <http://www.filmsite.org/aa39.html> (last visited Oct. 21, 2009). *Gone With The Wind* did win several other Academy Awards, including the award for Best Picture, Best Director (Victor Fleming), Best Actress (Vivien Leigh), Best Screenplay, and Best Supporting Actress (Hattie McDaniel, the first African-American nominee and winner). *Gone With the Wind* (1939), <http://www.imdb.com/title/tt0031381/awards> (last visited Oct. 21, 2009). The film is the earliest of only four films to win the awards for Best Picture, Best Director, Best Actress, and Best Screenplay and not capture the award for Best Actor. The other three films are: *Mrs. Miniver* (Walter Pidgeon was nominated but lost to James Cagney in *Yankee Doodle Dandy*) (see 1942 Academy Awards Winners and History, <http://www.filmsite.org/aa42.html> (last visited Oct. 21, 2009)); *Annie Hall* (Woody Allen was nominated but lost to Richard Dreyfuss in *The Goodbye Girl*) (see 1977 Academy Awards Winners and History, <http://www.filmsite.org/aa77.html> (last visited Oct. 21, 2009)); and *Terms of Endearment* (Jack Nicholson was nominated for, and won, the Oscar for Best Supporting Actor; the Oscar for Best Actor was awarded to Robert Duvall in *Tender Mercies*) (see 1983 Academy Awards Winners and History, <http://www.filmsite.org/aa83.html> (last visited Oct. 21, 2009)). Twelve films have captured the awards for Best Picture, Best Director, Best Actor, and Best Screenplay but not the award for Best Actress: *The Best Years Of Our Lives*, *On The Waterfront* (although Eva Marie Saint won the Oscar for Best Supporting Actress), *The Bridge On The River Kwai*, *A Man For All Seasons*, *Patton*, *The French Connection*, *Kramer vs. Kramer* (although Meryl Streep won the Oscar for Best Supporting Actress), *Gandhi*, *Amadeus*, *Rain Man*, *Forrest Gump*, and *American Beauty*. See Oscar’s Greatest Films and Sweeps and its Most Surprising Loser, Jan. 29, 2008, http://www.associatedcontent.com/article/560801/oscars_greatest_films_and_sweeps_and.html?cat=37. Only three films in the history of the Academy Awards

Socrates states the general rule: a year divisible by four, such as 1960 and 2004, is a leap year. “So 1900 was a leap year, too,” suggests Clark. “No,” says Socrates. “The general rule is subject to an exception: a year divisible by 100 is not a leap year.” Clark replies, “Then 2000 was *not* a leap year.” “Sorry, no,” says Socrates. “A year divisible by 400 *is* a leap year.”¹⁵⁵ Clark, expressing some annoyance, responds, “So your general rule is subject to an exception, which itself is subject to an exception.” “Right!” exclaims Socrates. “What do you think of that?” To which Clark retorts, “Frankly, my dear, . . .”¹⁵⁶

When rules are subject to a myriad of exceptions, crafting the major premise of a syllogism becomes challenging. “All men are mortal” is a premise much easier to master than “All men are mortal, other than any male who (i) is a Southern Baptist with a law degree from BYU; (ii) teaches commercial law; (iii) follows the St. Louis Cardinals baseball team; (iv) enjoys playing the piano; (v) favors novels authored by Harlan Coben, Pat Conroy, Jeff Shaara, and Daniel Silva; and (vi) has two beautiful daughters and a ‘Proverbs 31’ wife.”¹⁵⁷ To a great extent, Article 9 raises the same concern: drawing a major premise from its rules can be difficult because many of its rules are riddled with exceptions. In fact, some variation of the phrase—“except as otherwise provided in subsection [x]”—appears *more than 100 times* in its text!¹⁵⁸

have captured the so-called “grand slam,” winning the Oscar for Best Picture, Best Director, Best Actress, Best Actor, and Best Screenplay: *It Happened One Night*, *One Flew Over The Cuckoo’s Nest*, and *The Silence Of The Lambs*. *Id.*

¹⁵⁵ For a general discussion of the leap year phenomenon, see Burkald Polster & Marty Ross, *Leaping to conclusions*, Mar. 3, 2008, <http://www.theage.com.au/news/education-news/leaping-to-conclusions/2008/02/29/1204226993941.html>.

¹⁵⁶ The full quote—“*Frankly, my dear, I don’t give a damn.*”—comes in at #1 on the American Film Institute’s list of “*AFI’s 100 Years . . . 100 Movie Quotes.*” See <http://connect.afi.com/site/DocServer/quotes100.pdf?docID=242> (last visited Oct. 21, 2009). Two other lines from *Gone With The Wind* also made the list: “*After all, tomorrow is another day!*” (#31) and “*As God is my witness, I’ll never be hungry again.*” (#59). *Id.* *Casablanca* has the most lines on the list (six). *Id.*

¹⁵⁷ Hey, I guess I must be immortal!

¹⁵⁸ See §§ 9-105(1), 9-108(a), 9-108(b), 9-108(b)(3), 9-108(b)(6), 9-108(d), 9-109(a), 9-201(a), 9-202, 9-203(b), 9-204(a), 9-207(a), 9-207(b), 9-207(c), 9-209(a), 9-301 (preamble), 9-301(1), 9-301(3), 9-305(a), 9-307(b), 9-307(f), 9-308(a), 9-309(1), 9-310(a), 9-311(a), 9-311(b), 9-311(c), 9-312(b), 9-312(b)(2), 9-313(a), 9-315(a), 9-316(d), 9-317(a)(2), 9-317(b), 9-317(c), 9-317(e), 9-319(a), 9-320(a), 9-320(b), 9-322(a), 9-322(c), 9-322(d), 9-323(a), 9-323(b), 9-323(d), 9-323(f), 9-324(a) (twice), 9-324(b) (twice), 9-324(d) (twice), 9-324(f) (twice), 9-325(a), 9-327(2), 9-327(3), 9-328(2), 9-330(c), 9-330(d), 9-334(d), 9-334(h), 9-335(c), 9-336(e), 9-340(a), 9-340(b), 9-341, 9-401(a), 9-403(b), 9-403(f), 9-404(b), 9-406(d), 9-406(f), 9-407(a), 9-407(b), 9-408(a), 9-501(a), 9-502(b), 9-506(b), 9-507(b), 9-508(a), 9-512(b), 9-513(d) (twice), 9-514(a), 9-514(b), 9-515(a), 9-515(b), 9-515(e), 9-516(a), 9-519(c), 9-525(a), 9-525(b), 9-601(a), 9-601(d), 9-601(g), 9-602, 9-611(b), 9-612(a), 9-620(a), 9-625(c), 9-626(a)(3), 9-702(a), 9-702(b), 9-703(b), 9-705(c), and 9-707(c). As used in the context of an exception, the phrase “subject to” also appears several times. See §§ 9-306(a), 9-322(d), 9-322(f), 9-324(b), 9-324(d), 9-326(a), 9-403(e), 9-404(b), 9-404(c),

Take, for example, the general rule that resolves priority disputes between two perfected secured creditors: priority is awarded to the secured party who first files its financing statement or otherwise perfects its security interest (whichever is earlier).¹⁵⁹ Given that a security interest in almost all forms of collateral can be perfected by a financing statement,¹⁶⁰ the practical effect is that the rule awards priority to the party who files first. The statute clearly warns the reader that the rule is subject to other exceptions found either in § 9-322¹⁶¹ or elsewhere in Part 3 of Article 9.¹⁶²

So consider the following, and rather typical, scenario. Lender makes a \$2 million loan on July 20¹⁶³ to BizCorp, a Texas corporation that sells and leases office and residential furniture through its stores located in Houston, Dallas, and San Antonio. To secure repayment, BizCorp executes on that date a written security agreement that grants to Lender a security interest in BizCorp's "*accounts, chattel paper, commercial tort claims, deposit accounts, documents, equipment, general intangibles, instruments, inventory, investment property, and letter-of-credit rights.*" The security agreement includes an after-acquired property clause and a future advance clause, and it additionally describes the commercial tort claims that exist on July 20.¹⁶⁴ Lender files its financing statement (accurate and complete in all respects and authorized by BizCorp) on July 20. A few days later, Lender's counsel receives a U.C.C. search report from the Texas Secretary of State dated July 30, showing no filings

9-405(c), 9-406(a), 9-406(b), 9-406(c), 9-406(d), 9-406(f), 9-406(g), 9-406(h), 9-502(a), 9-512(a) (twice), 9-604(b), 9-604(c), and 9-625(b).

¹⁵⁹ § 9-322(a)(1).

¹⁶⁰ Notable exceptions include collateral subject to perfection requirements stated in other state or federal law (e.g., motor vehicles which require lien notation on the certificate of title), deposit accounts and letter-of-credit rights (control only), and money (possession only). See §§ 9-311(a), 9-312(b).

¹⁶¹ See § 9-322(a) ("Except as otherwise provided in this section . . .").

¹⁶² See § 9-322(f)(1) (stating that the general rule of subsection (a) is "subject to . . . the other provisions of this part").

¹⁶³ July 20 is one of the most notable dates in world history. Neil Armstrong took mankind's first steps on the moon on July 20, 1969. Also on that date I celebrated my tenth birthday, an event of considerably less importance (except, perhaps at the time, to me).

Twenty-one years later, on July 20, 1990, William J. Brennan, Jr., concluded his distinguished career as an Associate Justice on the Supreme Court of the United States. See Members of the Supreme Court of the United States, <http://www.supremecourtusgov/about/members.pdf> (last visited Oct. 26, 2009).

¹⁶⁴ See § 9-108(e)(1) (stating that a reference to a "commercial tort claim" without additional language "is an insufficient description"). Notwithstanding the inclusion of an after-acquired property clause, future commercial tort claims will not be part of the collateral. See § 9-204(b)(2).

against BizCorp other than Lender's financing statement. BizCorp has no other secured debt or secured creditors.

Relying on the baseline "first to file or perfect" rule, a student may conclude that BizCorp has, and will continue to have, priority in all of the collateral in which a security interest may be perfected by filing.

But that conclusion is riddled with exceptions, including the following:

- Lender's security interest may not enjoy priority in a dispute with any future "buyer in the ordinary course of business" who claims the benefits afforded to such a party by § 9-320(a);
- Lender's security interest may not enjoy priority in a dispute with any future secured party claiming the superpriority status afforded to purchase-money security interests under § 9-324;
- Lender's security interest may not enjoy priority in a dispute with any secured party claiming priority under § 9-325 in future collateral acquired by BizCorp;
- Lender's security interest may not enjoy priority in a dispute with any secured party claiming priority under § 9-326 in collateral transferred by BizCorp to a "new debtor;"
- Lender's security interest may not enjoy priority in a dispute with any future secured party claiming an enforceable security interest in investment property perfected by control (§ 9-328(1));
- Lender's security interest may not enjoy priority in a dispute with any future "purchaser" who enjoys the benefits afforded to such a party by § 9-330;
- Lender's security interest may not enjoy priority in a dispute with any future "holder" or "purchaser" who enjoys the benefits afforded to such a party by § 9-331;
- Lender's security interest may not enjoy priority in a dispute with any future "transferee" who enjoys the benefits afforded to such a party by § 9-332;
- Lender's security interest may not enjoy priority in a dispute with any future creditor claiming a "possessory lien" under § 9-331; and
- Lender's security interest in collateral acquired by BizCorp more than four months after it makes a seriously misleading name change may become unperfected and lose priority unless Lender timely refiles (§ 9-507(c)).

No doubt there are others,¹⁶⁵ but the exceptions listed above are sufficient in number to illustrate the difficulty of stating a general rule in a manner that acknowledges its exceptions.

The preceding discussion illustrates another possible obstacle to the effective use of syllogisms in Secured Transactions. Article 9 occasionally adopts a rule that is subject to one or more exceptions. And a rule stripped of its exceptions may, when stated as a major premise, lead to an incorrect legal conclusion.

CONCLUSION

Judge Aldisert and his co-authors have called for a renewed emphasis on deductive reasoning in the classroom, with special attention given to the syllogism. Syllogistic reasoning requires the audience to craft a major premise (usually a legal rule) and a minor premise (often a statement of fact), and then draw from those two premises a conclusion. Students often find that correctly stating the legal rule can be quite challenging, but imperative to reaching a sound conclusion. That challenge dissipates in a statutory course, such as Secured Transactions, where the rules are codified. Perhaps more than many law school courses, then, Secured Transactions is a course in which the syllogism can be a useful learning tool in introducing, analyzing, and mastering the intricacies of Article 9. But, as illustrated in the preceding pages, Article 9 itself hinders the effective use of syllogistic reasoning.

Given these statutory roadblocks, then, should those of us who have been anointed to teach Secured Transactions simply dismiss Judge Aldisert's plea for more deductive reasoning in our classroom? Absolutely not. Notwithstanding these codified quirks, many of the legal issues introduced in the course can indeed be examined and resolved through syllogistic analysis. But that is a topic for another day—and a companion article.¹⁶⁶

¹⁶⁵ For example, the secured party may be concerned about the scope of its priority when the competing claimant is the Internal Revenue Service, a topic which I have explored in previous articles. See Timothy R. Zinnecker, *Resolving Priority Disputes Between the IRS and the Secured Creditor Under Revised UCC Article 9: And the Winner is . . . ?*, 34 ARIZ. ST. L.J. 921–66 (2002); Timothy R. Zinnecker, *When Worlds Collide: Resolving Priority Disputes Between the IRS and the Article Nine Secured Creditor*, 63 TENN. L. REV. 585–688 (1996).

¹⁶⁶ See *supra* note 15 and accompanying text.

Testing Open Source Waters: Derivative Works Under GPLv3

Joseph A. Chern*

INTRODUCTION

Richard Stallman¹ has always believed that sharing software is a moral obligation.² As such, in 1985, Stallman pioneered a movement towards free software and created the non-profit Free Software Foundation (“FSF”), a tax-exempt charity for free software development.³ The FSF, in turn, created the GNU General Public License (“GPL”) as an instrument to enforce free software principles.⁴ Although this license has become one of the most popular licenses in the world,⁵ legal uncertainties continue to plague its enforceability.⁶

Among the debated issues of the license, the FSF maintains that dynamically linking to a GPL-ed library creates a derivative work under copyright principles subject to the original license.⁷

* J.D. Candidate 2010, Chapman University School of Law; B.S. Computer Engineering, University of California, San Diego. I would like to thank Ryan Roemer for all of his constructive guidance and insight; Pamela Frohreich for her comments and encouragement; Kyhm Penfil for her edits and suggestions; and the Senior Articles Editors of the Chapman Law Review for all of their hard work. I also would like to thank my family for their love and support which made this possible.

¹ Software developer of GNU operating system and founder of Free Software Foundation. See Richard Stallman, *Richard Stallman's Personal Home Page*, <http://www.stallman.org/#serious> (last visited Oct. 7, 2009).

² Brian W. Carver, *Share and Share Alike: Understanding and Enforcing Open Source and Free Software Licenses*, 20 BERKELEY TECH. L.J. 443, 445 (2005) (describing Richard Stallman's belief that proprietary software was incompatible with “the golden rule”).

³ Free Software Foundation, *About Free Software and the GNU Operating System*, <http://www.fsf.org/about/> (last visited Oct. 7, 2009).

⁴ Free Software Foundation, *What is free software and why is it so important for society?*, <http://www.fsf.org/about/what-is-free-software> (last visited Oct. 7, 2009).

⁵ Free Software Foundation, *Licenses*, <http://www.gnu.org/licenses/licenses.html#GPL> (last visited Oct. 7, 2009) (reporting that the “GNU General Public License . . . is used by most GNU programs, and by more than half of all free software packages.”).

⁶ This article does not address all the legal issues with the GPL, but focuses on the issue of program linking. See *infra* Part IV.A; see also John Tsai, *For Better or Worse: Introducing the GNU General Public License Version 3*, 23 BERKELEY TECH. L.J. 547, 548 (2008) (analyzing other issues with GPLv3 including software patents and digital rights management).

⁷ David Turner, *LGPL and Java*, (2004), <http://www.fsf.org/licensing/licenses/lgpljava.html>.

There is, however, difficulty with this position because dynamic linking technically does not create a “copy” of any GPL-ed software.⁸ This article analyzes perceived arguments both for and against the FSF position. The legal validity of these arguments will be evaluated in relation to the current approach courts have taken with software copyright. This article argues that the FSF most likely does not have legal support for its position that dynamic linking to a GPL-ed library creates any GPL obligation. The implications of this holding are also examined and alternative consequences are considered.

Part I of this article discusses the development of the FSF and explores the history of the GPL. Part II briefly describes the application of the GPL and introduces the FSF’s interpretation of the extent of that application. Part III explains the technical complexity of computers, describes how copyright law has attempted to address the complex nature of software, and revisits the FSF’s interpretations in depth. Finally, Part IV analyzes perceived legal issues with the FSF position regarding current copyright law and examines alternative outcomes.

I. BACKGROUND

Before examining the legal issues with dynamic linking, it is helpful to understand the rationale behind the GPL and the free software movement.

A. Beginning of Free Software

The motivation for the free software movement began in 1970 when Stallman worked as a programmer for the Massachusetts Institute of Technology’s (MIT) Artificial Intelligence Lab.⁹ Frustrated with a paper jam in one of the lab printers, he decided to take action.¹⁰ Using the source code¹¹ freely provided with the centralized network printer, he modified the source code so that other users waiting would be alerted when the printer was jammed.¹² Eventually, the lab received a new Xerox printer, and Stallman tried to make the same

⁸ ABRAHAM SILBERSCHATZ ET AL., OPERATING SYSTEM CONCEPTS (8th ed. 2009).

⁹ SAM WILLIAMS, FREE AS IN FREEDOM: RICHARD STALLMAN’S CRUSADE FOR FREE SOFTWARE 1–12 (2002), available at <http://www.oreilly.com/openbook/freedom>.

¹⁰ *Id.* at 3.

¹¹ In computer programming, source code (also known as source or code singularly) is a collection of *human readable* statements and declarations that allow programmers to communicate with computers. Eventually, source code is compiled to object code, which makes up programs such as Microsoft Word, an Internet Browser, or a game. For more information on source code and an example of the first source code written in the C programming language, see BRIAN W. KERNIGHAN & DENNIS M. RITCHIE, THE C PROGRAMMING LANGUAGE 1–7 (2nd ed. 1988).

¹² WILLIAMS, *supra* note 9, at 3–5.

improvements; however, Xerox did not provide the printer's source code.¹³ Xerox supplied the hardware,¹⁴ but denied the release of their source code under contractual obligations.¹⁵

Consequently, Stallman set out to challenge the direction of software culture. In 1984, Stallman left MIT and wrote a complete operating system¹⁶ compatible with UNIX, the standard operating system at the time, and eventually called it "GNU."¹⁷ In 1985, Stallman founded the FSF to promote the project's concept of software freedom.¹⁸ The original focus of Stallman and the FSF was to bring a wholly free operating system into existence to give users an opportunity to inspect, share, and modify the source code of the software they used.¹⁹ According to the FSF, open source software hinges on the four software freedoms: (1) to run a program for any use; (2) to study the code and modify it for individual needs; (3) to redistribute; and (4) to improve a program for society.²⁰ It is important to note that the

¹³ *Id.* at 4–6.

¹⁴ Hardware, as distinguished from software, is the general term to describe the physical components of technology such as the components that make up a personal computer, including a printer. Software is the general term to describe the collection or programs, which function to tell hardware how to act. See WALTER SAVITCH, *JAVA: AN INTRODUCTION TO COMPUTER SCIENCE & PROGRAMMING* 4–5 (3d ed. 2004).

¹⁵ WILLIAMS, *supra* note 9, at 6.

¹⁶ An operating system is the software component of a computer system, which acts as a "manager" of the resources in the computer to coordinate between the hardware and software applications. See SAVITCH, *supra* note 14, at 8. Common operating systems are Windows for PC's and Tiger for Apple computers.

¹⁷ Free Software Foundation, *Overview of the GNU Operating System*, <http://www.gnu.org/gnu/gnu-history.html> (last visited Oct. 7, 2009) ("The GNU operating system is a complete free software system, upward-compatible with Unix. GNU stands for 'GNU's Not Unix.'").

¹⁸ What is free software and why is it so important for society?, *supra* note 4 ("in 1985, Stallman started the Free Software Foundation, a nonprofit with the mission of advocating and educating on behalf of computer users around the world.").

¹⁹ Free Software Foundation, *Philosophy of the GNU Project*, <http://www.gnu.org/philosophy/philosophy.html> (last visited Oct. 7, 2009) (describing the "philosophy of the free software movement, which is the motivation for our development of the free software operating system GNU"); see also Eben Moglen & Richard Stallman, *GPL Version 3: Background to Adoption*, <http://www.fsf.org/licensing/essays/gpl3-background.html> (last visited Oct. 7, 2009).

²⁰ The Free Software Foundation defines the four essential freedoms:

Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. More precisely, it means that the program's users have the four essential freedoms: The freedom to run the program, for any purpose (freedom 0). The freedom to study how the program works, and change it to make it do what you wish (freedom 1). Access to the source code is a precondition for this. The freedom to redistribute copies so you can help your neighbor (freedom 2). The freedom to improve the program, and release your improvements (and modified versions in general) to the public, so that the whole community benefits (freedom 3). Access to the source code is a precondition for this.

Free Software Foundation, *The Free Software Definition*, <http://www.gnu.org/philosophy/free-sw.html> (last visited Oct. 7, 2009).

movement's goals are "primarily social and political, not technical or economic."²¹ The term "free" did not mean monetarily free, but rather "free" to study and modify.²² FSF's often-quoted catchphrase is "'free' as in 'free speech,' not as in 'free beer.'"²³

B. General Public License, Versions 1–3

When Congress modified the Copyright Act to explicitly protect software programs,²⁴ the FSF became concerned about its efforts to promote shared source code since battling a copyright holder required great time and resources.²⁵ As a result, "companies, not hackers, held the automatic advantage[.]"²⁶ and provider corporations effectively could hold a monopoly on their software. Software typically builds upon the source code from previous programs;²⁷ therefore, tipping the scales in favor of large corporations places an arduous burden on the FSF's goals. Eventually, the FSF recognized it could actually use copyright licensing to its advantage.²⁸ A programmer could use a copyright license to bind others, who modify or distribute the programmer's source code, to the license's terms.²⁹ Specifically, open source programmers,³⁰ such as Stallman, could license their code with specific conditions, such as to allow others to copy it so long as they also publish any modified versions in return.³¹ This prompted the FSF to create a universal license making it possible for many projects to share the same code and keep programs

²¹ Moglen & Stallman, *supra* note 19.

²² GNU General Public License, version 2, <http://www.gnu.org/licenses/oldlicenses/gpl-2.0.html> (last visited Oct. 7, 2009) [hereinafter General Public License v2.0]; *see also* Carver, *supra* note 2, at 450–53 (using the term "open source" for software that refers to an open development process).

²³ FREE SOFTWARE, FREE SOCIETY: SELECTED ESSAYS OF RICHARD M. STALLMAN 41 (Joshua Gay ed., 2002).

²⁴ Act to Amend the Patent and Trademark Laws, Pub. L. No. 96-517, 94 Stat. 3015, 3028 (1980) (codified as amended at 17 U.S.C. § 117 (2006)); *see also* discussion *infra* Part III.B.

²⁵ WILLIAMS, *supra* note 9, at 123.

²⁶ *Id.*

²⁷ SAVITCH, *supra* note 14, at 12.

²⁸ WILLIAMS, *supra* note 9, at 123–26.

²⁹ *Id.*

³⁰ Open source programmers are generally defined as computer software programmers who use and make software readily available to the public domain. Open source software is to be distinguished from proprietary software. *See* Open Source Initiative, The Open Source Definition, <http://opensource.org/docs/osd> (last visited Oct. 7, 2009) (describing ten criteria for open source software).

³¹ WILLIAMS, *supra* note 9, at 124.

free.³² Thus, the first version of the GNU General Public License (“GPLv1”) was born.³³

GPLv1 attempted to solve two major problems that restricted the freedoms of “free software.”³⁴ First, many software companies released only binary versions—executable, but not modifiable or understandable by humans—of their source code.³⁵ GPLv1 solved this problem by requiring the release of human readable source code under a GPLv1 agreement.³⁶ Second, like the FSF, programmers could also benefit from copyright protection when they modified source code and redistributed it with additional restrictions to the existing license.³⁷ GPLv1 did not allow for combinations with software that might impose additional restrictions, because this would create a conflict.³⁸

Two years later, the FSF published the second version of the license called GNU GPL version 2 (“GPLv2”).³⁹ The underlying idea behind the second version was what the FSF calls “copyleft.”⁴⁰ A play on the phrase copyright, “copyleft” promoted rather than restricted the use of copyrighted materials.⁴¹ This change was best embodied in Section 7 of the GPLv2, which states that if a user has restrictions—such as court orders, agreements, or otherwise—placed on GPL-covered software that may hinder the rights of others, the user cannot distribute that code at all.⁴² Most importantly, “copyleft” principles subjected the author of a GPLv2 piece of software to two requirements: (1) to license the original and modified versions of the software under the same license; and (2) to provide the source code to any user of the license.⁴³ These provisions will be essential to the derivative work issue discussed below.⁴⁴

³² Robert W. Gomulkiewicz, *General Public License 3.0: Hacking the Free Software Movement's Constitution*, 42 HOUS. L. REV. 1015, 1024 (2005).

³³ The full text General Public License, version 1 is available at <http://www.gnu.org/licenses/old-licenses/gpl-1.0.html> (last visited Oct. 7, 2009) [hereinafter General Public License v1.0].

³⁴ *Id.*

³⁵ FREE SOFTWARE, FREE SOCIETY, *supra* note 23, at 4, 170–71.

³⁶ General Public License v1.0, *supra* note 33.

³⁷ *Id.*

³⁸ *Id.*

³⁹ Gomulkiewicz, *supra* note 32, at 1025–26 (reporting that Stallman created the second version to clarify misunderstandings and worries).

⁴⁰ See Free Software Foundation, *What is Copyleft?*, <http://www.gnu.org/copyleft/copyleft.html> (last visited Oct. 14, 2009).

⁴¹ *Id.*

⁴² General Public License v2.0, *supra* note 22, at § 7.

⁴³ *Id.* at §§ 2–3.

⁴⁴ See discussion *infra* Part IV.A.

After fourteen years, GPLv2 began to show its age in the form of many legal problems.⁴⁵ As a result, the final version of GPLv3, released on June 29, 2007, attempted to address some of the legal issues from the previous version while maintaining the core “copyleft” obligations.⁴⁶ Among the significant issues, GPLv3 tried to clarify the scope of the license by introducing newly defined terms and consolidated requirements.⁴⁷ GPLv2 borrowed many terms from U.S. copyright law, such as “derivative works” and “collective works”,⁴⁸ but this only led to ambiguities about how far the license should reach.⁴⁹ Furthermore, because the GPL is meant to be a global license that can apply in any country regardless of that country’s copyright laws, the GPLv3 abandoned the term “derivative work” and redefined the license to meet global standards.⁵⁰ The new license also includes “propagate” and “convey” to define what constitutes distribution⁵¹ and tries to clarify whether dynamic linking falls within the scope of the “copyleft” license.⁵²

II. SOFTWARE AND GPL

The resolution of the legal issues with GPL greatly depends on the FSF’s ability to advocate for definitions of legal terms as applied to software. Such definitions have yet to be evaluated by the courts. In addition, the unique nature of how a GPL is triggered can effectively limit its scope. This section first describes that unique nature and then previews the FSF’s legal interpretations. A more in-depth discussion of the FSF’s interpretation is revisited in Section IV.C, following a discussion of current case law.

⁴⁵ Moglen & Stallman, *supra* note 19.

⁴⁶ Not all of the legal issues are addressed here since the list could go on, but rather the most significant issues to this article will be discussed.

⁴⁷ Compare Free Software Foundation, *GNU Project: GNU General Public License Version 3*, <http://www.gnu.org/licenses/gpl-3.0.html> (last visited Oct. 14, 2009) [hereinafter General Public License v3.0], with General Public License v2.0, *supra* note 22.

⁴⁸ See General Public License v2.0, *supra* note 22, at §§ 0, 2(c); 17 U.S.C. § 101 (2006).

⁴⁹ Carver, *supra* note 2, at 458–60.

⁵⁰ Moglen & Stallman, *supra* note 19; compare GPLv3, *supra* note 47, at § 0, with GPLv2, *supra* note 22, at § 0.

⁵¹ General Public License v3.0, *supra* note 47, at § 0.

⁵² Tsai, *supra* note 6, at 566.

A. Triggering the GPL: “Distribution” and the “Viral” Effect

Pursuant to one of the “copyleft” provisions, modified versions⁵³ of GPL-ed software are subject to the same license as the original code⁵⁴ if released.⁵⁵ Consequently, critics of the license have described it as having a “viral effect.”⁵⁶ The effect of the license is metaphorically similar to that of a virus because in order to trigger the effects of the GPL, the only requirement is that any portion of the software be covered by the GPL.⁵⁷ If any portion of the software is covered by a GPL license, the remaining portions of the code, even if it was intended to be proprietary, become “infected” with the license.⁵⁸

Distributing any software under the license can cause this viral effect.⁵⁹ In order to use the license for a piece of software, the author only has to add two elements to his code: (1) “a copyright notice (such as ‘Copyright 1999 Terry Jones’),” and (2) “a statement of copying permission, saying that the program is distributed under the terms of the GNU General Public License.”⁶⁰ When these two elements are present, any distribution of that code will be subject to the same terms of the GNU GPL.⁶¹ For instance, assume software Program A contains the two elements and, as a result, is covered under the GPL. Anyone would be free to use Program A to create Program B. Accordingly, Program B is “infected” by the GPL of Program A

⁵³ “Modified versions,” as used here, is similar to the “derivative work” concept that was replaced from GPLv2. The term here assumes the work falls within the GPL definition of modified work.

⁵⁴ See Free Software Foundation, *supra* note 40.

⁵⁵ Among frequently asked questions, the Free Software Foundation provides: [t]he GPL does not require you to release your modified version, or any part of it. You are free to make modifications and use them privately, without ever releasing them But if you release the modified version to the public in some way, the GPL requires you to make the modified source code available to the program's users, under the GPL. Thus, the GPL gives permission to release the modified program in certain ways, and not in other ways; but the decision of whether to release it is up to you.

Free Software Foundation, *Frequently Asked Questions About the GNU Licenses*, Does the GPL require that source code of modified versions be posted to the public?, <http://www.fsf.org/licenses/licenses/gpl-faq.html#GPLRequireSourcePostedPublic> (last visited Oct. 24, 2009) [hereinafter FSF Frequently Asked Questions].

⁵⁶ Ron Phillips, *Deadly Combinations: A Framework for Analyzing the GPL's Viral Effect*, 25 J. MARSHALL J. COMPUTER & INFO. L. 487 (2008).

⁵⁷ General Public License v3.0, *supra* note 47, at §§ 0, 10 (describing the viral effect by automating licensing to any downstream recipients for works that are covered, which means any unmodified programs or worked based on the program).

⁵⁸ Phillips, *supra* note 56, at 492.

⁵⁹ *Id.* at 492–93; General Public License v3.0, *supra* note 47, at §§ 0, 10 (describing the viral effect by automating licensing to any downstream recipients for works that are covered, which means any unmodified programs or worked based on the program).

⁶⁰ Free Software Foundation, *How to use GNU licenses for your own software*, <http://www.gnu.org/licenses/gpl-howto.html> (last visited Oct. 24, 2009).

⁶¹ *Id.*

and must, at a minimum, have the same terms. In a simple situation, Program B typically contains a verbatim copy of Program A, which makes the license application a legal, rather than a technical, matter. In any case, the “viral effect” helps to achieve the goals of the free software movement by ensuring reciprocation of free software. It is important to note that neither private use nor modifications of a GPL code creates any licensing obligation, but obligations do apply if the code is publicly distributed.⁶²

B. FSF's Interpretations of Derivative Works

The FSF has been less than clear on its definition of what creates a derivative work. However, in order to promote the free software movement and the goals of a GPL, it is important to draw the line between “infected”⁶³ derivative works and works that are not “infected” that carry no “copyleft” obligations.

One thing is clear: the FSF has abandoned the term “derivative work” and has adopted “modified works” essentially to cover the same principles.⁶⁴ Although “modified works” is defined in the latest GPL,⁶⁵ the scope of the definition remains ambiguous and is still debated.⁶⁶ One possible interpretation of the FSF definition of “modified works” is that the definition mirrors the scope of the U.S. copyright law's definition of derivative work.⁶⁷ However, the drafters of the GPL seemingly urge a broader interpretation.⁶⁸ When answering questions

⁶² FSF Frequently Asked Questions, *supra* note 55.

⁶³ See *supra* Part II.A.

⁶⁴ See *supra* Part I.B.; compare General Public License v3.0, *supra* note 47, at § 0 (“The resulting work is called a ‘modified version’ of the earlier work or a work ‘based on’ the earlier work”) with General Public License v2.0, *supra* note 22, at § 0 (“The ‘Program,’ below, refers to any such program or work, and a ‘work based on the Program’ means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications . . .”).

⁶⁵ The Free Software Foundation defines modify as:

To ‘modify’ a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a ‘modified version’ of the earlier work or a work ‘based on’ the earlier work.

General Public License v3.0, *supra* note 47, at § 0.

⁶⁶ Tsai, *supra* note 6, at 548, 566–68 (arguing that new language creates new ambiguities and the open source community is still split over several key issues as to what is covered under GPL licenses).

⁶⁷ 17 U.S.C. § 101 (2006) (“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.’”); see also copyright discussion *infra* Part III.B.1.

⁶⁸ University of Washington School of Law, Derivative Works, <http://www.law.washington.edu/lct/swp/Law/derivative.html> (last visited Dec. 28, 2009) (arguing GPL drafters urge a broader definition of derivative works than US Copyright law); see also *infra* Part III.C; see also Tsai, *supra* note 6, at 555–56 (outlining a broader interpretation of the GPLv2 to include both collective works and compilations).

about whether programs are covered under the license, the FSF generally leans towards the answer that provides the broadest coverage.⁶⁹ Specifically, when dealing with a situation in which two separate programs are combined,⁷⁰ each with a different license, the FSF would argue that in most cases the programmer/user would be “combining them into one program” under the GPL.⁷¹ However, the FSF has admitted that “[t]his is a legal question, which ultimately judges will decide.”⁷²

III. PROGRAM LINKING AND COPYRIGHT

The FSF’s acknowledgment that courts ultimately will decide the scope of GPL licensing suggests the need for an analysis of the current legal interpretations of copyright protection by courts. Section III.A begins with a technical primer on software linking that may affect a court’s treatment of copyright protection for software. Section III.B examines traditional copyright law and basic copyright principles and analyzes courts’ approaches to derivative works. Section III.C revisits the FSF’s interpretation of “modified works” and focuses on the role software linking might play in this analysis.

A. Program Linking (Static and Dynamic): Technical Background

Developers of software often use pre-written code⁷³ and data when writing their own programs to avoid “reinvent[ing] the wheel.”⁷⁴ Software “libraries” are commonly used collections of reusable code and data that provide “common functions, such as printing, reading data from a disk, or opening a file.”⁷⁵ A computer will at some point “link” the relevant code from the library with the developer’s own code to create a larger functional

⁶⁹ FSF Frequently Asked Questions, *supra* note 55 (asserting adding modules to GPL programs are also covered, output from GPL covered programs can be covered by GPL, linking to libraries is also covered); *see also infra* Part III.C.

⁷⁰ *See* discussion *infra* Part III.A.

⁷¹ FSF, Frequently Asked Questions, *supra* note 55.

⁷² *Id.*

⁷³ *See* SAVITCH, *supra* note 14, at 12, 18 (stating programs are not typically written as one piece of code and may consist of several different parts, written by different people, combining other components that already exist).

⁷⁴ Alexandra J. Horne, *Shared Rights to Source Code: The Copyright Dilemma*, 32 SANTA CLARA L. REV. 497, 532 (1992) (“Furthermore, because of the complexity of the development process, the software developer can argue that the library routines must be afforded copyright protection because of the economic hardship the developer will face due to the prohibitive expense required to ‘reinvent the wheel’ for each new development project.”).

⁷⁵ Tsai, *supra* note 6, at 557.

program.⁷⁶ The timing of this linkage depends on whether the process is accomplished through *dynamic* or *static* linking.⁷⁷

In computer science, “human readable source code” typically goes through several steps before it is executed by the compiler.⁷⁸ Generally, these steps are divided into three “times”: compile time, load time, and execution time.⁷⁹ During compile time, the source code is translated into machine code to create an object module that is read by the computer.⁸⁰ What occurs during the load and execution times vary depending on hardware and software limitations.⁸¹

If an operating system supports only *static* linking, the libraries that the source code uses are combined with the object module during load time.⁸² Effectively, this is akin to replacing footnotes in an essay with a copy of the referenced material. The libraries are treated like another object module that is physically combined together with the new code into which it was integrated.⁸³ Conversely, if the operating system allows *dynamic* linking, the linking is “postponed until execution time.”⁸⁴ With dynamic linking, a “bookmark” is placed in the source code for each library reference, which indicates how to load the library code when it is actually being used.⁸⁵ Under this approach, there is no actual copy of the library code in the source code, but rather an address where the code can be accessed when needed.⁸⁶ The library and source code remain separate files on disk.⁸⁷

B. Traditional Copyright Law

As copyright law has developed, it has dynamically embraced new technologies. For example, courts,⁸⁸ and eventually Congress,⁸⁹ agreed that lithographs and photographs constitute

⁷⁶ SAVITCH, *supra* note 14, at 12 (describing the process of connecting several different pieces of code together as linking).

⁷⁷ SILBERSCHATZ, *supra* note 8, at 321.

⁷⁸ *Id.* at 318.

⁷⁹ *Id.*

⁸⁰ SAVITCH, *supra* note 14, at 9 (describing a compiler to translate the high-level language that humans can understand [source code] into low-level languages [machine code] that the computer can understand).

⁸¹ See SILBERSCHATZ, *supra* note 8, at 320.

⁸² See Tsai, *supra* note 6, at 557; SILBERSCHATZ, *supra* note 8, at 321.

⁸³ See SILBERSCHATZ, *supra* note 8, at 828.

⁸⁴ *Id.* at 321.

⁸⁵ *Id.*

⁸⁶ See *id.*

⁸⁷ See *id.*

⁸⁸ See, e.g., *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 58 (1883) (holding that there is “no doubt that the Constitution is broad enough to cover an act authorizing copyright of photographs, so far as they are representatives of original intellectual conceptions of the author”).

⁸⁹ 17 U.S.C. § 102(a)(5) (2006).

copyrightable subject matter in the early years. More recently, Congress convened the National Commission on New Technological Uses (“CONTU”)⁹⁰ to thoroughly explore the scope of copyright protection for software, a concept with which courts continue to struggle.⁹¹ This section will introduce relevant principles of copyright law as applicable to software and briefly examine how the courts currently address the changes.

1. Brief Overview of Copyright and Software: Basic Application

Copyright is the right of an author to control the reproduction of his or her original expression.⁹² Whether an author owns a valid copyright depends, *inter alia*, on whether the subject matter is copyrightable.⁹³ Copyrightable subject matter includes “original works of authorship fixed in any tangible medium of expression.”⁹⁴ Works of authorship can include anything, from literary works such as novels, to musical works such as compositions and lyrics.⁹⁵ However, although copyright law protects the expression of ideas,⁹⁶ copyrights do not extend to either the functional elements or the ideas underlying those expressions; this is an area typically left to patent law.⁹⁷ Because of the functional nature of software, Congress has been reluctant to allow for copyright protection for software. In fact, it was not until 1980 that Congress approved a statutory amendment to define a “computer program” in the Copyright Act.⁹⁸ Following the decision in *Apple Computer, Inc. v. Franklin Computer Corp.*, courts have continued to “reflect Congress’ receptivity to new

⁹⁰ NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS, FINAL REPORT (1978).

⁹¹ See, e.g., *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1247 (3d Cir. 1983); see also *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 516 U.S. 233 (1996); *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV 99-7654 HLH (BQRX), 2000 WL 525390, (C.D. Cal. Mar. 27, 2000).

⁹² ROBERT A. GORMAN & JANE C. GINSBURG, COPYRIGHT CASES AND MATERIALS 12 (7th ed. 2006).

⁹³ 17 U.S.C. § 102 (2006).

⁹⁴ § 102(a); see also *Feist Publ’n, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 346 (1991) (providing “original works” has not meant novel or unique, but rather independent creations with a minimal modicum of creativity).

⁹⁵ § 102(a).

⁹⁶ *Id.*

⁹⁷ § 102(b) (“In no case does copyright protection . . . extend to any idea, procedure, process, system, method of operation . . .”); compare 35 U.S.C. § 101 (2006) (discussing *patentable* inventions).

⁹⁸ Act to Amend the Patent and Trademark Laws, Pub. L. No. 96-517, 94 Stat. 3015, 3028 (1980) (codified as amended at 17 U.S.C. § 117 (2006)); see also *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1247 (describing that although computer programs are not expressly listed as a work of authorship, legislative history suggests that programs were considered copyrightable as literary works and software was not *per se* uncopyrightable purely based on functionality).

technology and its desire to encourage, through the copyright laws, continued imagination, and creativity in computer programming.”⁹⁹

Software copyrightability allowed programmers to become the beneficiaries of exclusive rights including the right to prepare derivative works.¹⁰⁰ The Copyright Act defines a derivative work as:

[A] work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgement, 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.’¹⁰¹

Historically, “derivative works” have been simpler to identify in areas such as choreography,¹⁰² visual art,¹⁰³ and literary sequels.¹⁰⁴ Defining a derivative work in the software context has not been as obvious and finds little clarity in the statutory definition or case law.¹⁰⁵ The recent trend in courts is that “copyright law today, in practice, affords narrow protection to software.”¹⁰⁶ This may be because “as software patents gain increasingly broad protection, whatever reasons there once were for broad copyright protection of computer programs disappear.”¹⁰⁷ Alternatively, copyright law may find difficulty affording broad protection because of the unique nature of software and the “inherent difficulty of fitting copyright[s] . . . to rapidly changing and variously described software technology approaches.”¹⁰⁸

⁹⁹ *Apple Computer*, 714 F.2d at 1254.

¹⁰⁰ 17 U.S.C. § 106 (2006) (enumerating the exclusive rights that the owner of a valid copyright has and can authorize, including reproduction, preparation of derivative works, distribution, and public performance of their copyrighted work).

¹⁰¹ 17 U.S.C. § 101 (2006).

¹⁰² *See, e.g.*, *Horgan v. MacMillan, Inc.*, 789 F.2d 157, 162 (2d Cir. 1986).

¹⁰³ *See, e.g.*, *Lee v. A.R.T. Co.*, 125 F.3d 580, 581 (7th Cir. 1995).

¹⁰⁴ *Suntrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257, 1259 (11th Cir. 2001).

¹⁰⁵ Carver, *supra* note 2, at 458 (noting the lack of case law and statutes to clarify derivative works for software).

¹⁰⁶ Emanuela Arezzo, *Struggling Around the “Natural” Divide: The Protection of Tangible and Intangible Indigenous Property*, 25 CARDOZO ARTS & ENT. L.J. 367, 411 (2007).

¹⁰⁷ Mark A. Lemley, *Convergence in the Law of Software Copyright?*, 10 HIGH TECH. L.J. 1, 26 (1995).

¹⁰⁸ Raymond T. Nimmer, *Legal Issues in Open Source and Free Software Distribution*, 885 PLI/P at 33, 91 (2006).

The difficulty in providing protection for software is further complicated by the “fair use” exception for copyright protection.¹⁰⁹ Essentially, fair use excuses “reasonable unauthorized appropriations” which is determined by evaluating four factors.¹¹⁰ The four factors are: (1) the purpose and character of the use;¹¹¹ (2) the nature of the copyrighted work;¹¹² (3) the amount and substantiality of the portion used in relation to the copyrighted work;¹¹³ and (4) the effect of the use upon the potential market.¹¹⁴ Because copying software is often for the public benefit, to promote the sciences without substantially impairing the value of the copied work, “fair use” has provided a venerable defense against infringement in the software context.¹¹⁵

Despite the struggle to find steady copyright ground in the computer context, it may be useful to examine how courts have dealt with the issue thus far.

2. Reverse Engineering and Non-Literal Copying Software Cases

Recent court cases¹¹⁶ have confirmed that the unique nature of software¹¹⁷ yields different copyright results with similar fact patterns. Although there is currently scant case law specifically discussing whether dynamic linking of software programs can implicate a derivative work, there is an extensive history of reverse engineering and non-literal copying software cases that

¹⁰⁹ 17 U.S.C. § 107 (2006).

¹¹⁰ *Id.*; see also *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578 (1994) (“All [factors] are to be explored, and the results weighed together, in light of the purposes of copyright.”).

¹¹¹ § 107(1); see also *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514 (9th Cir. 1992) (holding intermediate software copying is allowed if it is the only means to get to the uncopyrighted functional material).

¹¹² § 107(2); see also *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 561 (1985) (weighing sub-factors such as whether the copyrighted work has been published already since right to first publication is important under copyright).

¹¹³ § 107(3); see also *Harper & Row*, 471 U.S. at 601 (Brennan, J., dissenting) (weighing quality of portion used versus quantity).

¹¹⁴ § 107(4).

¹¹⁵ See *Sega*, 977 F.2d at 1523 (considering the public benefit that copying software for the functional requirements would increase creative expressions based on the unprotected ideas); see also *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 607 (9th Cir. 2000) (stating copyright law does not create monopolies over certain markets and legitimate software competitors may be welcomed so long as it does not merely supplant the original).

¹¹⁶ This is not an exhaustive list of all computer or software related cases, but rather a select few that reflect common principles within the scope of this article.

¹¹⁷ See also Andre R. Jaglom, *Internet Distribution, E-Commerce and Other Computer Related Issues: Current Developments in Liability On-Line, Business Methods Patents and Software Distribution, Licensing and Copyright Protection Questions*, SN019 A.L.I.-A.B.A. 535, 606 (stating that “[t]he difficulty is in establishing the scope of copyright protection, since unlike the traditional kinds of works protected by copyright, software has a significant functional component”).

are instructive to the ultimate linking question. In a series of reverse engineering¹¹⁸ cases, courts have held that copyright infringement existed even when there was no literal copying of any software, as long as the visuals created by the software were substantially similar in both idea and expression, and at least an exact description of the visuals existed in concrete form.¹¹⁹ Some courts also have determined that there was no copyright infringement where the software distributed was not copied, even though there had been intermediate copying to create the distributed work.¹²⁰

For example, in *Micro Star v. FormGen, Inc.*, the defendant owned the copyrights to a popular computer game.¹²¹ The game included an editor, which allowed players to create their own levels for the game.¹²² Without literally copying any of defendant's software, plaintiff packaged user-created levels for sale on a CD.¹²³ Plaintiff sought a declaratory judgment that there was no copyright infringement, while defendant counterclaimed for a preliminary injunction.¹²⁴ In reversing the district court, the Ninth Circuit granted the preliminary injunction and held that the defendant would likely succeed in showing that the audiovisual displays generated when running plaintiff's distributed levels would infringe on plaintiff's right to create derivative works.¹²⁵

The Court reasoned that the displays would be similar in both idea and expression.¹²⁶ Even though the plaintiff did not literally copy any of the software for the games, the plaintiffs were infringing on the defendant's right to create sequels by creating new tales told by the plaintiff's new levels.¹²⁷ However, the Court clarified that the work the defendant was infringing upon was the *story* of the video game characters.¹²⁸ It conceded that the software files themselves did not create the derivative

¹¹⁸ Reverse engineering is essentially looking at a final product or output and working backwards to uncover the underlying elements that make the product or output work. See *Sega*, 977 F.2d at 1514.

¹¹⁹ *Micro Star v. FormGen Inc.*, 154 F.3d 1107, 1111–12 (9th Cir. 1998); *Worlds of Wonder, Inc. v. Veritel Learning Sys., Inc.*, 658 F. Supp. 351, 355 (N.D. Tex. 1986); *Midway Mfg. Co. v. Arctic Int'l, Inc.*, 704 F.2d 1009, 1013 (7th Cir. 1983).

¹²⁰ See *Sega*, 977 F.2d at 1514; *Sony*, 203 F.3d at 600 (dismissing copying because it was only intermediate).

¹²¹ *Micro Star*, 154 F.3d at 1109.

¹²² *Id.*

¹²³ *Id.*

¹²⁴ *Id.*

¹²⁵ *Id.* at 1114.

¹²⁶ *Id.* at 1112.

¹²⁷ *Id.*

¹²⁸ *Id.*

works because they did “not, in fact, incorporate any of [plaintiff’s] protected [software].”¹²⁹

Similarly, in *Worlds of Wonder, Inc. v. Veritel Learning Systems, Inc.*, the defendant made cassette tapes compatible for use with a toy bear named Teddy Ruxpin.¹³⁰ The toy was animated by placing a cassette into the back of the animal, thus moving the eyes, nose, and mouth in synchronization with the story or song on the tape.¹³¹ Although the defendant did not actually copy any of the plaintiff’s tapes or software, the Northern District of Texas held that the “series of related images” that the tapes created were the copyrighted work in this case.¹³² Granting a preliminary injunction, the Court held that the defendant infringed the plaintiff’s right to prepare derivative works because defendant’s tape cassettes created a modified work substantially similar to the plaintiff’s “series of related images.”¹³³ Again, because the defendant’s tapes did not contain any copyrighted software, the Court emphasized that the defendant was infringing the *story* of the toy rather than the tapes.¹³⁴

Furthermore, courts have found no copyright infringement where there was literal intermediate copying of plaintiff’s software, done only to access uncopyrighted elements of the code. In *Sega Enterprises Ltd. v. Accolade Inc.*, the defendant copied the plaintiff’s video game code in order to reverse engineer the program to discover the requirements for compatibility with plaintiff’s video game console.¹³⁵ The defendant used the information to create its own games to sell in the market.¹³⁶ In the software code for these games, the defendant also included plaintiff’s initialization code¹³⁷ into a standard header file¹³⁸ to allow games to interact with the console code.¹³⁹

¹²⁹ *Id.*

¹³⁰ *Worlds of Wonder, Inc. v. Veritel Learning Sys., Inc.*, 658 F. Supp. 351, 352 (N.D. Tex. 1986).

¹³¹ *Id.*

¹³² *Id.* at 355.

¹³³ *Id.*

¹³⁴ *Id.* at 356.

¹³⁵ *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514 (9th Cir. 1992).

¹³⁶ *Id.* at 1515.

¹³⁷ *Id.* at 1516. Initialization code in the context of computer programming is code that gives variables a value prior to executing the code. For more discussion on initialization code, see KERNIGHAN & RITCHIE, *supra* note 11, at 40.

¹³⁸ *Sega*, 977 F.2d at 1516. In computer programming, particularly in the C and C++ programming language, header files are separate files from the source code, which typically contain declarations for variables, classes, subroutines, and other identifiers. For more discussion on header files, see KERNIGHAN & RITCHIE, *supra* note 11, at 33.

¹³⁹ *Sega*, 977 F.2d at 1514.

The Ninth Circuit ultimately held that the intermediate copying was fair use, and therefore was not infringement, because it was the only means to access the uncopyrightable elements, namely the compatibility procedure.¹⁴⁰ The Court reasoned that the defendant was only trying to “become a legitimate competitor in the field of Genesis-compatible video games” and that consumers would still continue to purchase multiple video games.¹⁴¹ The Court also found the distribution of the header files was purely functional and held that “[t]he initialization code is a functional feature of a Genesis-compatible game and Accolade may not be barred from using it.”¹⁴² The Court focused on the purely functional aspect of the software copied in determining that the third party could distribute copies of the functional code.¹⁴³

In a similar case, *Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.*, the plaintiff created a device called a Game Genie, which was sold for use with a Nintendo Entertainment Video Game system.¹⁴⁴ The device physically attached to the video games before they were placed into the console, in order to block data bytes sent to the console and replace the bytes with new values.¹⁴⁵ Effectively, this allowed users to modify aspects of the game by entering in “cheat codes.”¹⁴⁶

The Ninth Circuit held that because the Game Genie acted only as a window into the computer program, it did not create a derivative work and there was no infringement.¹⁴⁷ The Court focused on the fact that Game Genie, as distributed, did not contain any of the Nintendo code in “some concrete or permanent form” and therefore did not meet the statutory definition of a derivative work.¹⁴⁸ It also did not matter that the visual images the code generated could be reconstructed when the game was over by entering the same codes, because the images did not exist

¹⁴⁰ *Id.* at 1518.

¹⁴¹ *Id.* at 1523; *see also* Sony Computer Entm’t, Inc. v. Connectix Corp., 203 F.3d 596, 602 (9th Cir. 2000) (applying the analysis in *Sega* to determine that defendant’s copying of plaintiff’s code from a video game console was fair use because it was for purposes of a new video game, different in character, and that defendant was a legitimate competitor in the market for plaintiff’s games).

¹⁴² *Sega*, 977 F.2d at 1531.

¹⁴³ *See id.* at 1516.

¹⁴⁴ *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 967 (9th Cir. 1992).

¹⁴⁵ *Id.*

¹⁴⁶ *Id.*

¹⁴⁷ *Id.* at 968; *see also* Micro Star v. FormGen Inc., 154 F.3d 1107, 1111 (9th Cir. 1998) (distinguishing the case based on the software existing in some permanent or concrete form).

¹⁴⁸ *Id.*

permanently *at distribution* of the Game Genie.¹⁴⁹ Despite copyright law's lack of protection for functional elements, the Court also mentioned that the "Game Genie is useless by itself, it can only enhance" and "such innovations rarely will constitute infringing derivative works under the Copyright Act."¹⁵⁰

Whether or not one agrees with the holdings of these cases, they all provide insight into the various approaches of applying copyright principles to software. Courts first determine whether a work "incorporates" any portion of the copyrighted work.¹⁵¹ If a work does not incorporate any protected expression, it cannot be a derivative.¹⁵² Even when protected expressions have been incorporated into the new work, they must exist in some "concrete or permanent form" at distribution to be a derivative work.¹⁵³ Temporary reproductions following distribution are not enough to be "concrete or permanent."¹⁵⁴ Furthermore, even where there is intermediate copying, courts have suggested that this is permissible so long as it is used purely to gain access to the unprotected functional elements.¹⁵⁵ Finally, attempts to promote creative expression may also influence how a court ultimately rules, regardless of any adverse market effect.¹⁵⁶ Despite these opinions, the FSF appears to urge a different approach.¹⁵⁷

¹⁴⁹ *Id.* (characterizing the Nintendo case saying once the game was over, the audiovisual displays generated were gone and as a result, were not permanent).

¹⁵⁰ *Lewis Galoob Toys*, 964 F.2d at 969; *see also* 17 U.S.C. § 117 (2006) (codifying the exception that it will not be infringement for the owner of a computer program to make a copy or adaptation provided the new copy or adaptation is created as an essential step in the *utilization* of the computer program).

¹⁵¹ *See Micro Star*, 154 F.3d at 1110–11.

¹⁵² *Id.* at 1110.

¹⁵³ *Lewis Galoob Toys*, 964 F.2d at 967.

¹⁵⁴ *Micro Star*, 154 F.3d at 1111 (characterizing the Nintendo case saying once the game was over, the audiovisual displays generated were gone and as a result, were not permanent).

¹⁵⁵ *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1524 (9th Cir. 1992); *see also* 17 U.S.C. § 117 (2006) (codifying the exception that it will not be infringement for the owner of a computer program to make a copy or adaptation provided the new copy or adaptation is created as an essential step in the *utilization* of the computer program).

¹⁵⁶ *Sega*, 977 F.2d at 1523 (noting the court is "free to consider the public benefit resulting from a particular use notwithstanding the fact that the alleged infringer may gain commercially" and growth of creative expression can serve a public interest); *see also* *Lewis Galoob Toys*, 964 F.2d at 969 (recognizing that "technology often advances by improvement rather than replacement" as a reason to exclude works such as spell-checking or the Game Genie from infringement), *and* U.S. CONST. art. I, § 8 (granting Congress power to enact Copyright laws "to promote the progress of science and useful arts").

¹⁵⁷ *See supra* Part II.B; *see also infra* Part III.C.

C. FSF / GPL Interpretation of Static / Dynamic Linking

The FSF, and the drafters of the GPL in particular, likely would not agree with courts' treatment of derivative works.¹⁵⁸ Terminology aside, whether or not a work is a "derivative" or "modified" version of a GPL-ed source is critical to the success of the free software movement.¹⁵⁹ Because of the "viral nature" of the GPL, the narrower the scope of coverage under the GPL, the fewer "infections" and the fewer GPL-species works there will be.¹⁶⁰ Accordingly, the FSF has attempted to tackle the problematic issue of what falls under the scope of the license by addressing program linking.¹⁶¹

Software that uses static linking to a GPL library warrants minimal discussion that it is a derivative work under any definition. With software utilizing static linking, an actual copy of the GPL code is transcribed verbatim into the larger program.¹⁶² Because the software now contains the GPL code "or a portion of it, either verbatim or with modifications," the software is now subject to the provisions of the license.¹⁶³ Specifically, any future distribution and modification of that software must now include the source code and will be subject to the same license.¹⁶⁴ This much seems clear.

Dynamic linking has generated a different response to the question whether a derivative or modified work is subject to GPL's "copyleft" provisions.¹⁶⁵ This debate occurs in part because the GPL library is never physically combined with the larger program in which it is referenced.¹⁶⁶ As discussed above, the GPL library and the larger program exist on completely different disks, and the GPL software is never technically "modified."¹⁶⁷ Consequently, if no derivative work is created when dynamically linking to a GPL library, then there is no obligation to follow the license.¹⁶⁸ Under this analysis, software developers are free to retain proprietary software despite using GPL code.

¹⁵⁸ See Derivative Works, *supra* note 68 (arguing GPL drafters urge a broader definition of derivative works than US Copyright law).

¹⁵⁹ See *supra* Part II.B.

¹⁶⁰ See *supra* Part II.A.

¹⁶¹ See *infra* Part IV.A.

¹⁶² See *supra* Part III.A.

¹⁶³ General Public License v2.0, *supra* note 22, at § 0.

¹⁶⁴ *Id.* at §§ 1–3.

¹⁶⁵ Lothar Determann, *Dangerous Liasons – Software Combinations as Derivative Works? Distribution, Installation, and Execution of Linked Programs Under Copyright Law, Commercial Licenses, and the GPL*, 21 BERKELEY TECH. L.J. 1421, 1458–62, 1488 n.255 (2006) (stating that dynamic linking typically does not create a derivative work).

¹⁶⁶ See *infra* Part IV.A.

¹⁶⁷ See *supra* Part III.A.

¹⁶⁸ See Phillips, *supra* note 56, at 500–02.

In spite of this, the FSF maintains that dynamic linking to a GPL library can create a single work that is subject to the GPL terms.¹⁶⁹ When asked about the specific issue of program linking and its relation to GPLv3, Stallman expressly addressed that:

[I]t doesn't matter which kind of linking is being used. If there are two modules that are designed to be run linked together and it's clear from the design from one or the other that they are meant to be linked together then we say they are treated as one program and so I hope that will make it a little bit clearer although that's not really a change, it's a clarification. That's what we believe GPL version two means already.¹⁷⁰

The "clarification"¹⁷¹ to which Stallman was referring is a new express provision in GPLv3.¹⁷² Under the new license, the "copyleft" obligation now extends specifically to "dynamically linked subprograms" that require an "intimate data communication or control flow between these subprograms."¹⁷³ The FSF appears to take a firm position that dynamic linking undeniably creates a "copyleft" obligation in a single work. However, it weakens its position with the self-imposed condition that the dynamic link must involve an "intimate data communication."¹⁷⁴ Thus, while the FSF clarifies rather than changes its view that dynamically linked programs always have been within the scope of the GPL, courts must now define terms such as "intimate" to determine the status of dynamically linked programs in addition to depending upon principles of copyright.

IV. CHALLENGES AND ISSUES WITH GPL AND DYNAMIC LINKING

This section analyzes the likelihood of success of the FSF's argument, perhaps simplified, that dynamic linking creates a derivative work. Infringing upon an exclusive right such as

¹⁶⁹ Free Software Foundation, *supra* note 71 ("If the modules are included in the same executable file [static linking], they are definitely combined in one program. If modules are designed to run linked together in a shared address space [dynamic linking], that almost surely means combining them into one program.").

¹⁷⁰ Richard Stallman, *Transcript of Richard Stallman in Torino*, (March 18, 2006), <http://www.fsfeurope.org/projects/gplv3/torino-rms-transcript.en.html#q2-linking> (last visited Nov. 11, 2009).

¹⁷¹ *Id.*

¹⁷² General Public License v3.0, *supra* note 47, at § 5.

¹⁷³ *Id.* at § 6 (requiring one to supply corresponding source); *Id.* at § 1 (defining corresponding source to include "interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work").

¹⁷⁴ *Id.* at § 1; *see also* Tsai, *supra* note 6, at 566 (arguing that the FSF takes a "tenuous position, stating first that dynamically linked programs are within the scope of the [license], then pulling back with a cryptic condition"); Free Software Foundation, *supra* note 71 (stating that one relevant condition should be what kind of information is exchanged).

preparing a derivative work would subject one to the restrictions of the corresponding license, namely the GPL in this case. Section IV.A considers the courts' likely approach to the FSF's argument based on current legal theories; Section IV.B examines the implications of the courts' likely trend; and Section IV.C asks whether that approach is the optimal solution.

A. Is There Any "Copying"?

Courts specifically have never addressed the issue of whether non-GPL libraries can dynamically link to GPL libraries with or without the same "viral effect" of the license.¹⁷⁵ The FSF leaves questions unanswered,¹⁷⁶ and courts struggle to set standards for software in general.¹⁷⁷ However, as case law on the issue is presently developing, a court likely would not find that a dynamically linked program falls within the scope of a GPL license.

With dynamic linking, the FSF would support a more technical analysis of the definition of "intimate."¹⁷⁸ If a linked program, whether dynamic or static, does not share the requisite "intimate data communication," there is no "copyleft" obligation.¹⁷⁹ However, as seen above, courts would likely focus on the abstract inquiry of whether the challenged software is a derivative work rather than on the underlying technology.¹⁸⁰ In any case, it is unlikely that a court would find an "intimate relationship" if the court did not also find that the linked program was a derivative work. When a work does not incorporate any protected expression, it cannot be a derivative.¹⁸¹ Dynamic linking never incorporates any protected expression; it only makes references to it.¹⁸² One may argue that the operating system physically incorporates the code somewhere in memory during runtime when needed; however, while having some merit, this argument should fail for two reasons.

First, recall that the GPL applies only if the work is released.¹⁸³ If a program can be released or distributed separately from a GPL work, the license obligation does not take

¹⁷⁵ In a close case, *MySQL AB v. Progress NuSphere*, Plaintiff sued Defendant for copyright infringement by linking code, but the case settled. See Press Release, MySQL, FAQ on MySQL vs. NuSphere Dispute (Jul. 13, 2001) available at <http://www.mysql.com/news-and-events/generate-article.php?id=75>.

¹⁷⁶ See *infra* Part IV.C.

¹⁷⁷ See *supra* Part III.B.

¹⁷⁸ See General Public License v3.0, *supra* note 47.

¹⁷⁹ *Id.*

¹⁸⁰ See *supra* Part III.B.2.

¹⁸¹ *Micro Star v. FormGen Inc.*, 154 F.3d 1107, 1110 (9th Cir. 1998).

¹⁸² SILBERSCHATZ, *supra* note 8, at 268–71.

¹⁸³ See Free Software Foundation FAQs, *supra* note 55.

effect.¹⁸⁴ For example, assume Program A is meant to be maintained as a proprietary word processing program, while Program B is a GPL-ed printing program. Program A can be distributed for use to word process without any use of printing capabilities of Program B. Program A used alone does not “incorporate” any of Program B’s protected expression. If User X of Program A then wants to have printing functionality, he can independently acquire Program B and use it privately with no obligation. As a result, even if Program A and B are combined in memory when the computer is running both of them, there is no “copyleft” obligation for this private use.

Second, even if combined in memory when the user is running the program, copyright law requires the protected expression or, in the example above, Program B, to exist in some “concrete or permanent form.”¹⁸⁵ Although virtual computer memory may be considered “concrete or permanent,”¹⁸⁶ temporary recreations are not enough to meet this requirement.¹⁸⁷ The combination will exist only while the programs run¹⁸⁸ and, as in the *Nintendo* case, as soon as the programs end, the combination does as well.

A second argument could be made that even references to GPL-ed code can be considered to “incorporate” the protected expression. This argument should also fail. Courts have suggested that even intermediate copying is permissible in order to take advantage of the unprotected functional elements.¹⁸⁹ References do not even rise to the level of intermediate copying, but are, as the name suggests, merely pointers to a particular function. Similar to copying an entire video game source code to find out compatibility functions,¹⁹⁰ referencing to GPL-ed code is a mechanism to use the unprotected function and would likely not be considered “incorporation” of any protected material.

¹⁸⁴ See General Public License v3.0, *supra* note 47; Viral Code and Vaccination, *supra* note 58.

¹⁸⁵ See *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 968 (9th Cir. 1992).

¹⁸⁶ See also *Micro Star v. FormGen Inc.*, 154 F.3d 1107, 1111 (9th Cir. 1998) (stating that concrete and permanent forms can exist on CD-ROMs, which are a type of computer memory).

¹⁸⁷ See *Lewis Galoob Toys*, 964 F.2d at 967 (stating the effects of the toy were temporary).

¹⁸⁸ See *SAVITCH*, *supra* note 14, at 5 (stating computer memory is used for computer’s “intermediate calculations”).

¹⁸⁹ See *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514 (9th Cir. 1992); *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 600 (9th Cir. 2000) (dismissing copying because it was only intermediate).

¹⁹⁰ *Sega*, 977 F.2d at 1526.

Finally, fair use factors may weigh in favor of no copyright infringement for dynamic linking. Even if there is copying in computer memory, the four fair use factors would likely not favor the FSF.¹⁹¹ First, the nature and purpose of using GPL-ed code is typically to promote software efficiency and to create new programs with additional functionality.¹⁹² Second, the nature of the GPL-ed work is that it already has been published because it is freely distributed and encourages its use. Third, the amount copied would be only that which is necessary for the program to function. Finally, courts have established fair use where the effect on the market from the copying is to establish legitimate competition.¹⁹³ Software vendors could make these arguments in their favor to establish fair use.

As a result, a court would likely not find that a derivative work is created simply by dynamic linking because the GPL-ed code is neither “incorporated” nor exists in a “concrete or permanent form.” It is hard to imagine that the level of “intimate data communication” is met when one program is not a legal derivative of the other. Additionally, if a court finds “incorporation,” it is likely that the incorporation would be protected by fair use.

B. Implications Assuming No Obligations

Assuming dynamic linking to GPL code does not impose any “copyleft” obligations,¹⁹⁴ proprietary software vendors can plan to selectively incorporate open software without further obligation. However, the free software community would not benefit from misappropriation of their work products. While pure believers in free software principles may continue their efforts in providing an open development model, others who once supported the FSF for its reciprocal guarantee that they would receive the same benefit may reconsider. This could result in a range of outcomes, from a floodgate of legal challenges to a decrease in free software development. The worst-case scenario might be an end of free software.¹⁹⁵

¹⁹¹ See *supra* Part III.B.1.

¹⁹² Tsai, *supra* note 6, at 557; SAVITCH, *supra* note 14, at 12.

¹⁹³ *Sega*, 977 F.2d at 1514, 1524.

¹⁹⁴ See General Public License v3.0, *supra* note 47.

¹⁹⁵ Free Software Foundation, *Copyleft and the GNU GPL*, <http://www.gnu.org/gnu/thewnuproject.html> (last visited Jan. 11, 2010) (admitting that “to permit such combinations would open a hole big enough to sink a ship”).

C. Alternative Outcome and Implications

Although a court most likely would hold that the distribution of software that only dynamically links to a copyrighted work is not itself a derivative work, it is worth exploring legal scenarios in which a court *does* impose copyright obligations for such a distribution.

One possible scenario could use the same principle for which the FSF is fighting and use it offensively. If dynamic linking creates derivative works because the operating system combines the works into shared memory, could the creator of the operating system equally argue that loading the software into its operating system creates derivative works and thus subjects users to their own license? For instance, could Microsoft, owners of the common operating system Windows, assert that dynamic loading of libraries in Windows creates a Microsoft licensed product? Furthermore, the Microsoft license could impose conflicting restrictions with the FSF such as prohibiting any GPL-ed linking in its entirety.¹⁹⁶

A second scenario may involve two dynamic libraries that both have the same functionality: `lib_GPL`, which is GPL-ed, and `lib_BSD`, which has no license obligations. Consider Program A that is distributed for use with either library, but will not function without one of them. Inevitably, users of Program A will choose to combine either library, but choosing one over the other may have entirely different consequences. Can it be possible that Program A is obligated by the GPL regardless of which library the user chooses? This likely would lead to yet another debate between the FSF, who would answer the question in the affirmative, and proprietary software vendors, who likely would disagree.

Although the preceding thought experiments also result in potential problems for the FSF, it is worth considering which scenario will yield greater benefits for the public.¹⁹⁷

CONCLUSION

Software complexities have challenged the legal system to find workable standards in areas such as copyright law. While this area is still developing, the software industry has attempted to define its own responsibilities. However, these responsibilities may not always find legal support. The FSF, for instance,

¹⁹⁶ General Public License v1.0, *supra* note 33 (stating that GPLv1 does not allow for combining code with conflicting licenses and thus would result in less possible code combinations in the world).

¹⁹⁷ See *supra* Part III.B.2.

maintains that dynamic linking, just like static linking, to GPL-ed software can create a derivative work subject to a GPL license. Although dynamic and static linking perform the same functional goals from a programming perspective, their technical differences may produce different legal implications. To date, there is no case law addressing this specific issue; however, some opinions suggest that the FSF has questionable basis to assert that dynamic linking creates a derivative work. Unlike static linking, there is no requisite incorporation of protected material to create a derivative work. Furthermore, other copyright principles such as “fair use” likely weigh against the FSF’s position. With potentially both positive and negative consequences for not only free software but also proprietary software proponents, this unresolved intersection of technology and copyright law will provide a rich and interesting source of future legal developments.

Heads I Win, Tails You Lose: The Disparate Treatment of Similarly Situated Taxpayers Under the Personal Injury Income Tax Exclusion

*Habib E. Hanna**

INTRODUCTION

Personal injury tort litigation continues to dominate the legal landscape in the United States.¹ Whether at the hands of a jury verdict, court order or a structured settlement agreement, tort litigation leads to the exchange of vast amounts of money.² Subsequently, whenever there is an exchange of money between two or more people, the United States Tax Code (“Code”) will likely have an impact on all the parties involved. The Code is primarily concerned with taxing income, so it tends to focus on the recipient of a tort damages money award.³ The primary Internal Revenue Code (“I.R.C.”) provision that deals with money received subsequent to a personal injury tort damages award is I.R.C. § 104(a)(2) (“section 104(a)(2)”)⁴. Since it was enacted in 1918, section 104(a)(2) has generated a considerable level of criticism, dismay and confusion over its intent, definition, and application.⁵

* J.D. Candidate, Chapman University School of Law, May 2010; B.A. Political Science, California State University Fullerton, June 2006. I wish to thank Professor Frank Doti for his invaluable guidance and insight. I also thank Allison De Tal, Kari Kadomatsu, Evan Scher and Michelle Nguyen for their help. I sincerely thank my sisters Mona, Salwa, Marcel and Mariam, as well as my niece Mary and her husband Peter, my uncle, Dr. Samy Farag, my brothers-in-law and all my nieces and nephews, for their love, support and encouragement. This article is dedicated to my parents whose love and devotion is my constant source of hope and motivation.

¹ See generally John T. Nockleby, *How to Manufacture a Crisis: Evaluating Empirical Claims Behind 'Tort Reform,'* 86 OR. L. REV. 533 (2007).

² *Id.*

³ I.R.C. § 61 (West 2009) (“[G]ross income means all income from whatever source derived.”). Unless otherwise noted, all references to the “Code” are to the Internal Revenue Code of 1986, as amended.

⁴ I.R.C. § 104(a)(2) (West 2009) stating:

(a) In general. Except in the case of amounts attributable to (and not in excess of) deductions allowed under section 213 (relating to medical, etc., expenses) for any prior taxable year, gross income does not include . . . (2) the amount of any damages (other than punitive damages) received (whether by suit or agreement and whether as lump sums or as periodic payments) on account of personal physical injuries or physical sickness[.]

⁵ See Frank J. Doti, *Personal Injury Income Tax Exclusion: An Analysis and Update*, 75 DENV. U. L. REV. 61, 79 (1997) (“For nearly eighty years, taxpayers, their advisors, and the government have wrestled with the scope of the personal injury exclusion.”). See also Renee C. Harvey, Note, *Commissioner v. Schleier: An Unfair Interpretation of Section 104(a)(2)*, 30 U.S.F. L. REV. 313, 342

Congressional intent behind the enactment of section 104(a)(2) represents a foundational theme that often dictates the method and manner in which section 104(a)(2) is applied.⁶ Congress designed section 104(a)(2) with one primary goal in mind: to alleviate the tax burden for those who suffered a personal injury and received income as a result of tort litigation.⁷ At its essence, section 104(a)(2) excludes from taxation any amount of money damages recovered by a taxpayer “on account of personal physical injuries or physical sickness.”⁸ While this phrase seems relatively straightforward, it has proven difficult in real-life application.⁹ Two of the primary issues presented by section 104(a)(2) are the definition of the word “physical” and the phrase “on account of.”¹⁰ A quick example will show how section 104(a)(2) works. A person is injured in a car accident, sues for money damages and recovers an award for pain, suffering and lost wages. Because this taxpayer received money on account of a personal physical injury, that amount is fully excludable and the taxpayer owes no income tax on the amount recovered.

One of the main problems encountered under section 104(a)(2) involves a tort that causes purely emotional distress that later manifests into a physical injury.¹¹ Any damages received in such a scenario are not excluded under section 104(a)(2) because the damages are not received on account of the physical injury, rather they are received on account of the emotional distress injury.¹² This outcome appears to be discriminatory in that it treats similarly situated taxpayers—those who suffered physical injuries at one point or another—differently.¹³ This disparate result is due to the fact that one physical injury occurred immediately as a result of the

(Fall 1995) (stating that section 104(a)(2) “has caused the courts, taxpayers, and the [Internal Revenue] Service considerable consternation” since its inception); Margarita R. Karpov, Note, *To Tax or Not To Tax — That is the Question in the Midst of Murphy v. I.R.S.*, 23 AKRON TAX J. 143, 143–44 (2008) (“This code section has seen differences of opinion, amendments, and commentator mistrust on numerous occasions”); Robert J. Henry, *Torts and Taxes, Taxes and Torts: The Taxation of Personal Injury Recoveries*, 23 HOUS. L. REV. 701, 702 n.7 (1986) (stating that section 213(b)(6) of the Revenue Act of 1918 was the predecessor to the modern iteration of section 104(a)(2)).

⁶ See Douglas K. Chapman, *No Pain — No Gain? Should Personal Injury Damages Keep Their Tax Exempt Status?*, 9 U. ARK. LITTLE ROCK L.J. 407, 409 (1986–1987) (discussing the intent of Congress in relation to section 104(a)(2)).

⁷ *United States v. Burke*, 504 U.S. 229, 236 n.6 (1992).

⁸ I.R.C. § 104(a)(2) (West 2009).

⁹ See Doti, *supra* note 5, at 61–62 (referring to the phrase “on account of personal physical injuries or physical sickness” and stating that “there still are many uncertainties over the meaning of these terms”).

¹⁰ Benjamin T. Cory, Note, *Amos v. Commissioner: The Ambiguous and Ever-Changing Definition of What Constitutes a Personal Physical Injury Under Internal Revenue Code Section 104(A)(2)*, 66 MONT. L. REV. 247, 248–49 (Winter 2005).

¹¹ See Patrick E. Hobbs, *The Personal Injury Exclusion: Congress Gets Physical but Leaves the Exclusion Emotionally Distressed*, 76 NEB. L. REV. 51, 88–89 (1997).

¹² See *id.*

¹³ Nicholas M. Whittington, Note, *Against the Grain: An Interdisciplinary Examination of the 1996 Federal Statutory Changes to the Taxability of Personal Injury Awards*, 37 WASHBURN L.J. 153, 185 (Fall 1997) (concluding that the “physical” requirement leads to arbitrary application of the tax law).

tort while the other physical injury occurred at a later point in time as a consequence of the emotional distress that was immediately caused by the tort.¹⁴ This comment posits that section 104(a)(2) treats similarly situated taxpayers inequitably and leads to an application of the tax code that lacks uniformity. Specifically, this comment argues that those who suffer real, verifiable physical manifestations of emotional distress injuries should receive the same favorable tax treatment received by those who suffer purely physical injuries.

This comment acknowledges and accepts the fact that Congress specifically refused to extend the benefits of section 104(a)(2) to those who suffer purely emotional distress injuries.¹⁵ However, there are legitimate cases where taxpayers suffer verifiable physical manifestations of emotional distress and those individuals should be afforded the same tax treatment as those who suffer purely physical injuries.¹⁶ Part I outlines the background and evolution of section 104(a)(2). Part II presents the development of section 104(a)(2) by examining the important case decisions dealing with the interpretation and application of section 104(a)(2). Part III undertakes an exhaustive analysis of the inequity that results when section 104(a)(2) is applied to individuals who suffer physical manifestations of emotional distress. Part III also discusses the treatment of physical manifestations of emotional distress by medical science and the field of tort law. Part IV suggests a viable course of action that would eliminate this disparate treatment while upholding congressional intent. Finally, this comment concludes that the totality of the circumstances presented leads to the logical deduction that section 104(a)(2) treats similarly situated taxpayers in a disparate manner and that the best course of action is through the adoption of the recommendations made in Part IV.

I. BACKGROUND: TRACING THE DEVELOPMENT AND THE EVOLUTION OF SECTION 104(A)(2)

Section 61 of the Code states that, "Except as otherwise provided in this subtitle, gross income means all income from whatever source

¹⁴ *Id.* at 161:

If there is an original physical injury or physical sickness, then all damages, including emotional distress damages, that flow from that injury are treated as if they are on account of a physical personal injury or physical sickness. Recoveries for physical injuries caused by emotional distress, not linked to an original physical injury, are not excludable from income.

¹⁵ J. Martin Burke & Michael K. Friel, *Getting Physical: Excluding Personal Injury Awards Under the New Section 104(a)(2)*, 58 MONT. L. REV. 167, 177 (Winter 1997) (explaining that Congress specifically stated in I.R.C. § 104 that "emotional distress shall not be treated as a physical injury or physical sickness").

¹⁶ See Kurt A. Leeper, Note, *Arguably Arbitrary: Taxation and the Physical Injury Requirement of I.R.C. Section 104(a)(2)*, 55 CASE W. RES. L. REV. 1039, 1061 (2005) ("[B]y requiring a showing of 'traditional' physical injury, [section 104(a)(2)] fails to protect taxpayers who have suffered a physical harm brought about by a nonphysical injury.").

derived”¹⁷ While the word “all” implies an extraordinary level of inclusion, there are limited but significant exceptions to the rule.¹⁸ Those exceptions, however, are often interpreted in a narrow manner such that the breadth of section 61 remains extremely broad.¹⁹ Section 104(a)(2) represents an important exception to section 61 and addresses the exclusion of personal injury proceeds received subsequent to civil tort litigation or to a negotiated tort settlement agreement.²⁰ Section 104(a)(2) states that “gross income does not include . . . the amount of any damages (other than punitive damages) received (whether by suit or agreement and whether as lump sums or as periodic payments) on account of personal physical injuries or physical sickness.”²¹

Prior to 1996, section 104(a)(2) did not include the word “physical” in its language and referred only to an exclusion of income received on account of personal injuries or sickness.²² However, in 1996 Congress amended section 104(a)(2) and added the word “physical” to the statutory language.²³ This amendment resulted from the original statute’s vague congressional intent to prevent the recipients of money damages awards from using section 104(a)(2) if the award resulted from a physical manifestation of an emotional distress injury.²⁴ This seemingly minor change to section 104(a)(2) led to increased levels of apprehension among legal professionals and academics who were already critical of the pre-amendment version for its ambiguous language.²⁵ In addition, the courts have struggled²⁶ and wavered²⁷ in their attempts to find a bright line of separation between what is obviously physical, what may be partly physical and partly emotional, and what may be purely emotional.²⁸ There are two

17 I.R.C. § 61 (West 2009).

18 Karpov, *supra* note 5, at 146.

19 *See id.* at 146–47; *see also* Comm’r v. Schleier, 515 U.S. 323, 327 (1995) (stating that the Court has “repeatedly emphasized the sweeping scope” of I.R.C. § 61 (internal quotations omitted)).

20 *Schleier*, 515 U.S. at 328–29.

21 I.R.C. § 104(a)(2) (West 2009).

22 O’Gilvie v. United States, 519 U.S. 79, 85 (1996) (citing the Revenue Act of 1918, Ch. 18 § 213(b)(6), 40 Stat. 1066 as the predecessor to I.R.C. § 104(a)(2)).

23 Anthony C. Infanti, *Tax Equity*, 55 BUFF. L. REV. 1191, 1209 (2008).

24 Gregory L. Germain, *Taxing Emotional Injury Recoveries: A Critical Analysis of Murphy v. Internal Revenue Service*, 60 ARK. L. REV. 185, 200–01 (2007).

25 *See generally* Tamara Larre, *Pity the Taxpayer: The Tax Exemption for Personal Injury Damages as Disability Policy*, 33 QUEEN’S L.J. 217, 220–21 (2007). *See also* Cory, *supra* note 10, at 261 (“By inserting the word ‘physical’ into § 104(a)(2) Congress created an arbitrary and ambiguous standard.”).

26 Robert W. Wood, *Waiting to Exhale: Murphy Part Deux and Taxing Damage Awards*, SN059 A.L.I.-A.B.A. 717, 719–20 (Feb. 2008).

27 *Compare* *Murphy v. I.R.S.*, 460 F.3d 79 (D.C. Cir. 2006) (“*Murphy I*”) (holding that I.R.C. § 104(a)(2) violated the Sixteenth Amendment because it permitted the taxation of emotional damages awards which do not fall within the definition of “incomes” found in the text of the Sixteenth Amendment), *with* *Murphy v. I.R.S.*, 493 F.3d 170 (2007) (“*Murphy II*”) (reversing *Murphy I* based on the proposition that purely emotional damages and any physical injuries resulting thereafter are not within the definition of physical injuries as intended by Congress when it enacted the 1996 amendment to I.R.C. § 104(a)(2)). For an in-depth discussion of *Murphy I* and *II*, see *infra* Part III.E–F.

28 *See generally* Karpov, *supra* note 5.

lines of debate surrounding this issue. First, what are the exact parameters of the word physical.²⁹ Second, how does the nexus between physical and emotional injuries impact the excludability of income under section 104(a)(2).³⁰

The 1996 amendment to section 104(a)(2) had a disparate impact on the applicability of the exclusion.³¹ The amendment provided comfort for those who suffered obvious physical injuries but it also appeared to apply in a seemingly *inequitable* manner based primarily on the fact that those who suffer legitimate physical manifestations of emotional distress are forced to pay tax on every penny they receive in money damages.³² This disparate treatment led many to make the argument that the first *Murphy* court was correct in its assessment that the 1996 amendment was not uniformly applied and thus constituted a violation of the Sixteenth Amendment of the United States Constitution.³³

In addition, there was another aspect of the 1996 amendment to section 104(a)(2) that added even more fuel to the fire of the criticism surrounding the amended language: this involved the phrase “on account of” in section 104(a)(2).³⁴ Courts have interpreted this phrase to mean that *any* damages award that flows directly from an initial physical injury will be covered by section 104(a)(2) and thus will be deemed non-taxable to the recipient.³⁵ The Supreme Court interpreted this to mean that even purely emotional injuries that emanated from physical injuries were covered by

²⁹ See Wood, *Waiting to Exhale*, *supra* note 26, at 723–24 (inquiring about “what is and is not physical”).

³⁰ Germain, *supra* note 24, at 202–03.

³¹ Doti, *supra* note 5, at 62.

³² See Infanti, *supra* note 23, at 1209:

In the service of bias, Congress has enacted a version of 104(a)(2) that will result in the dissimilar treatment of similarly situated persons, a violation of the widely accepted doctrine of horizontal equity in tax policy analysis. Injured individuals are not treated similarly. Those who are physically harmed may recover all damages awards tax-free. Those injured by employment bias [or other non-physical injuries] may recover nothing tax free.

See also Karpov, *supra* note 5, at 151; Leandra Lederman, *Statutory Speed Bumps: The Roles Third Parties Play in Tax Compliance*, 60 STAN. L. REV. 695, 708–10 (2007) (arguing that inequitable treatment under 104(a)(2) may lead people to fabricate physical injuries to avoid paying taxes. This could lead to enforceability and verifiability issues for the I.R.S. and, ultimately, to “higher taxes or reduced spending” by the federal government).

³³ See Wood, *Waiting to Exhale*, *supra* note 26, at 720, 724–25 (stating that “most people read Judge Ginsburg’s first *Murphy* opinion as a statement that the 1996 amendments to section 104 (imposing the “physical” requirements) were invalid.” Wood goes on to describe how the *Murphy* II court did an “about-face” to avoid the constitutional question when it reversed its previous *Murphy* I decision).

³⁴ See Germain, *supra* note 24, at 197.

³⁵ O’Gilvie v. United States, 519 U.S. 79, 84 (1996) (stating that “pain and suffering damages, medical expenses, and lost wages . . . are covered by the statute and hence excluded from income not simply because the taxpayer received a tort settlement, but rather because each element . . . satisfies the requirement . . . that the damages were received *on account of* personal injuries or sickness” (quoting *Comm’r v. Schleier*, 515 U.S. 323, 330 (1995)) (internal quotations omitted) (emphasis added)).

the exclusion.³⁶ However, to the dismay of some critics of the 1996 amendment to section 104(a)(2), the exclusion did not apply in the converse situation.³⁷ That is to say, if someone suffered latent physical injuries that manifested themselves as a result of a purely emotional injury, then that person is not covered by the exclusion and is subject to taxation on any and all money damages received for either the emotional or the physical manifestations of those emotional injuries.³⁸

In addition to the criticism of the way section 104(a)(2) treated people who suffered purely emotional injuries or physical manifestations of purely emotional injuries there is another central concern regarding the applicability of the language added in the 1996 Amendments. This concern has to do with the definition of the word “physical” and the extent to which any non-visible physical injuries would be covered by the exclusion.³⁹ The Internal Revenue Service (“I.R.S.”) has not specifically addressed the issue. However, the I.R.S. has made one notable statement regarding the definition of “physical.”⁴⁰ In October 2007, the I.R.S. issued a Private Letter Ruling in which it stated that a physical injury had to involve touching that causes a bruise, cut, swelling or bleeding in order to qualify for exclusion under section 104(a)(2).⁴¹ This language projects the self-interested nature of the I.R.S.’s position on the matter, a position that naturally lends itself to interpreting the breadth of section 104(a)(2) as narrowly as possible.⁴²

However, the congressional intent behind the 1996 amendment to section 104(a)(2) tends to show that the legislators who penned the amendment may have envisioned a broader scope for the word “physical.”⁴³ The problem with defining the word “physical” as narrowly as the I.R.S. has in its Private Letter Ruling is that purely internal injuries

³⁶ *Schleier*, 515 U.S. at 329.

³⁷ See Germain, *supra* note 24, at 200–06.

³⁸ See Robert W. Wood, *Tax Treatment of Settlements and Judgments*, SN059 A.L.I.-A.B.A. 665, 669–70 (Feb. 2008) (asking the question “[w]hat if the defendant’s conduct did not involve physical touching, but did produce physical injury, particularly physical injury caused by intentional infliction of emotional distress?” and cautioning that the I.R.S.’s emphasis on physical touching places an “undue emphasis on the original harm” without taking into account the real possibility that purely emotional distress can lead to what may be classified as verifiable physical manifestation). See also Germain, *supra* note 24, at 204 (arguing that the “clear purpose of the new physical injury or physical sickness requirement was to make all amounts received for emotional distress damages, including amounts received on account of physical manifestations arising out of that emotional distress, includible in income”).

³⁹ See Doti, *supra* note 5, at 75.

⁴⁰ See Karpov, *supra* note 5, at 172 (stating that “[e]ven ten years after the 1996 amendment changes, many questions remain in the interpretation of ‘physical’ in § 104(a)(2)” and arguing that section 104(a)(2) “has not been applied consistently” by the courts or the I.R.S.).

⁴¹ I.R.S. Priv. Ltr. Rul. 200041022 (July 17, 2000).

⁴² See Wood, *Tax Treatment of Settlements*, *supra* note 38, at 672 (criticizing I.R.S. Private Letter Ruling 200041022 by stating that “[a] distinction between a case in which the plaintiff is touched and then injured as a result, compared with the plaintiff who is *not* touched but injured in the same way, seems artificial”).

⁴³ See Doti, *supra* note 5, at 73–74.

and pain are apparently not physical enough to qualify.⁴⁴ Critics of the narrow I.R.S. interpretation have posited that the approach is too simplistic in that it relies on obvious, visible signs of injury and disregards the very real possibility that personal injury litigants could suffer from physical injuries that do not necessarily manifest themselves in such visible ways.⁴⁵

So, who suffers as a result of the addition of the word “physical” in the 1996 amendment to section 104(a)(2)? Primarily those who are victims of purely emotional distress injuries and those who are victims of emotional distress injuries that lead to a manifestation of subsequent physical injuries.⁴⁶ Examples include victims of sexual harassment, wrongful termination, employment discrimination, unlawful arrest, wrongful incarceration, and sexual molestation, just to name a few.⁴⁷ These emotional distress victims often have a serious medically and psychologically documented impact on their physical and mental well-being.⁴⁸

Unfortunately, the applicability of section 104(a)(2) forces these emotional distress victims to pay taxes on any amount of damages they recover while someone who suffers a minor injury that bruises for a couple of days and then goes away is treated with proverbial “kid gloves” and escapes all or most of the tax liability associated with that injury.⁴⁹ While the background of section 104(a)(2) highlights its ambiguities and shortcomings, the next section considers various cases that illustrate how the courts have struggled to define workable parameters for the application of section 104(a)(2).

⁴⁴ See Wood, *Tax Treatment of Settlements*, *supra* note 38, at 671. Wood states that he is not sure the I.R.S. got it right in Letter Ruling 200041022 when they tried to:

draw a line between the various incidents of sexual harassment and touching that left no ‘observable bodily harm,’ and the various assaults (that they term beginning with the ‘First Pain Incident’). Although the [ruling] seems cogent enough, the truth is that very often it is difficult to separate exactly what causes trauma (and what [type] of trauma) and what does not.

⁴⁵ See *id.* See also Marianna G. Dyson, *Hot Topics in Fringe Benefits*, SG003 A.L.I.-A.B.A. 1, 9 (2001) (arguing that the narrow interpretation made by the Internal Revenue Service in Private Letter Ruling 200041022 is unworkable as a practical matter because “there can be [an] adverse impact on the physical being of an individual, whether or not physical symptoms are manifested”).

⁴⁶ See Infanti, *supra* note 23, at 1209 (“[T]he physical/non-physical dichotomy adopted in amended section 104(a)(2) guarantees disparate and disadvantageous treatment of recoveries for the physical and emotional harms to workers in job bias cases.”).

⁴⁷ See Stephen Cohen, *Why Civil Rights Lawyers Should Study Tax*, 22 HARV. BLACKLETTER L.J. 1, 2 (2006) (“It seems arbitrary and unfair, for example, that damages for sexual harassment are taxed if damages for a broken leg are tax-exempt.”).

⁴⁸ See *infra* Part III.B.

⁴⁹ See Cory, *supra* note 10, at 258 (discussing the fact that the petitioner in *Amos v. Commissioner* suffered what was labeled a minor injury at best and that “[d]espite the fact that Mr. Amos had no observable injury . . . the Tax Court concluded that Mr. Amos had suffered a physical injury”).

II. CASE HISTORY: KEY DECISIONS ILLUSTRATE THE JUDICIARY'S ATTEMPTS TO SET THE BOUNDARIES OF I.R.C. § 104(a)(2)

There are only three Supreme Court decisions dealing directly with section 104(a)(2).⁵⁰ However, there are also numerous federal and tax court cases dealing with the interpretation of section 104(a)(2).⁵¹ Most of these decisions, with the exception of one important outlier,⁵² demonstrate the growing trend that the courts should defer to Congress and the I.R.S. in matters requiring statutory interpretation of the Code and federal tax legislation.⁵³

A. United States v. Burke

One of the first Supreme Court cases dealing with section 104(a)(2) was *United States v. Burke* decided in 1992.⁵⁴ In *Burke*, the Court looked at the taxability of a money damages award received for violation of Title VII of the Civil Rights Act of 1964.⁵⁵ The taxpayer believed that the back pay portion of the settlement—received for a sex discrimination claim under Title VII of the Civil Rights Act of 1964—was not taxable because it was received on account of a personal injury.⁵⁶ The Court, however, held that the award was not excludable under section 104(a)(2) because there was no tort-like personal injury redressed by the suit.⁵⁷

The Court came to this conclusion based on the reasoning that the employees were awarded damages in the form of back pay which did not fall within the rubric of tort-like personal injury envisioned by the original drafters of section 104(a)(2).⁵⁸ More importantly, the *Burke* Court

⁵⁰ See *infra* Part II.A–C. Other than the primary cases discussed in this comment, there are a handful of other Supreme Court cases that dealt tangentially with I.R.C. § 104(a)(2).

⁵¹ See *infra* Part II.D–F.

⁵² See *infra* Part II.E. See also cases cited, *supra* note 27.

⁵³ See Myron C. Grauer, *Justice O'Connor's Approach to Tax Cases: Could She Have Led the Court Toward a More Collaborative Role for the Judiciary in the Development of Tax Law?*, 39 ARIZ. ST. L.J. 69, (Spring 2007) (discussing Justice O'Connor's influence over the Court's deferential approach to Congress in tax law matters). See also T. James Lee, Jr., *Section 104(a)(2) After Commissioner v. Schleier: Litigating the Excludability of Statutory Damages 'Received on Account of Personal Injuries'*, 1996 BYU L. REV. 531, 543 n.86 (1996) (examining Justice O'Connor's dissenting opinion in *Schleier* where "[s]he asserted that even though the interpretive rulings of the [Internal Revenue] Service do not rise to the authoritative level of regulations, the Court must give substantial deference to the Service's reasonable interpretations") (emphasis added).

⁵⁴ *United States v. Burke*, 504 U.S. 229 (1992). This case was decided before the 1996 amendments to section 104(a)(2), however, it remains an important decision because it established one of the first boundary lines between what is and what is not a personal injury.

⁵⁵ *Id.* at 230.

⁵⁶ *Id.* at 229.

⁵⁷ *Id.* at 241 ("[W]e cannot say that a statute such as Title VII, whose sole remedial focus is the award of back wages, redresses a tort-like personal injury within the meaning of § 104(a)(2).").

⁵⁸ *Id.* at 234 (citing 26 C.F.R. § 1.104-1(c) (1991) and stating that:

Neither the text nor the legislative history of § 104(a)(2) offers any explanation of the term personal injuries. Since 1960, however, IRS regulations formally have linked identification of a personal injury for purposes of § 104(a)(2) to traditional tort principles: The term 'damages received (whether by suit or agreement)' means an amount received . . . through

interpreted section 104(a)(2) as applying to both physical and non-physical injuries.⁵⁹ The Court based this reasoning on the fact that Congress amended section 104(a)(2) in 1989 to specifically preclude punitive damages from the exclusion, thus implying that both physical and non-physical injuries should be covered by the exclusion.⁶⁰ The *Burke* decision set the foundation for the requirement of tort or tort-like damages in a section 104(a)(2) analysis.⁶¹

B. Commissioner v. Schleier

The Court followed *Burke* with the ground-breaking 1995 decision in *Commissioner v. Schleier*.⁶² This decision involved money damages awarded for back wages and liquidated damages in an age discrimination suit filed under the Age Discrimination in Employment Act of 1967 (“ADEA”).⁶³ In *Schleier*, the taxpayer—with the *Burke* decision likely in mind—included the amount received for back wages in his federal income tax return.⁶⁴ However, the taxpayer excluded the portion received for liquidated damages because he claimed that the liquidated damages were received on account of personal injuries under section 104(a)(2).⁶⁵ Using logic similar to that in *Burke*, the Court declared: “Our consideration of the plain language of [I.R.C. §] 104(a), the text of the regulation implementing [I.R.C. §] 104(a)(2), and our reasoning in *Burke* convince us that a recovery under the ADEA is not excludable from gross income.”⁶⁶

The *Schleier* Court reasoned “that Congress intended the ADEA’s liquidated damages to be punitive in nature; thus, they serve no compensatory function and cannot be described as being ‘on account of personal injuries.’”⁶⁷ *Schleier* is significant because the Court annunciated a two-prong test for use in analyzing section 104(a)(2) exclusions.⁶⁸ The test requires that, in order to be excludable under section 104(a)(2), the money damages award (i) must be received through prosecution or

prosecution of a *legal suit or action based upon tort or tort type rights*, or through a settlement agreement entered into in lieu of such prosecution. (emphasis added) (some internal quotations omitted)).

⁵⁹ *Id.* at 235–36 n.6.

⁶⁰ *Id.* (“The enactment of this limited amendment addressing only punitive damages shows that Congress assumed that other damages (i.e., compensatory) would be excluded in cases of both physical and nonphysical injury.”).

⁶¹ See Leeper, *supra* note 16, at 1052 (stating that *United States v. Burke* established the “tort-like” requirement).

⁶² *Comm’r v. Schleier*, 515 U.S. 323 (1995). Like *Burke*, the *Schleier* decision also predated the 1996 amendments to section 104(a)(2) when Congress added the “physical” requirement to personal injuries. However, this case remains important and applicable because the two-prong test it established continues as the threshold test in applying section 104(a)(2).

⁶³ *Schleier*, 515 U.S. at 324–25.

⁶⁴ *Id.* at 327.

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.* at 323.

⁶⁸ See Harvey, *supra* note 5, at 341.

settlement of an action “based upon tort or tort type rights” and (ii) received “on account of personal injuries or sickness.”⁶⁹ However, in adopting the “on account of personal injuries” language, the Court added to the confusion surrounding section 104(a)(2) when it failed to provide any guidelines as to what constitutes a personal injury or sickness.⁷⁰

C. *O’Gilvie v. United States*

Following the trend it established in *Burke* and *Schleier*, the Court placed another boundary line on the section 104(a)(2) exclusion in the 1996 *O’Gilvie* decision.⁷¹ This case dealt with punitive damages received in a wrongful death cause of action based on products liability tort law.⁷² The taxpayer claimed that the punitive damages were excludable under section 104(a)(2) because the damages award was received on account of a personal injury, specifically death.⁷³ The Court disagreed and held that the “punitive damages received here were not received ‘on account of’ personal injuries; hence the provision does not apply, and the damages are taxable.”⁷⁴

The *O’Gilvie* Court’s reasoning mirrored that in *Burke* and *Schleier* in that it relied heavily on congressional intent, and the I.R.S.’s interpretation of section 104(a)(2), in making its decision.⁷⁵ The *O’Gilvie* Court first conceded that the damages received were indeed based on tort or tort-type injuries, thus satisfying the first prong of the *Schleier* test.⁷⁶ However, the Court agreed with the Government’s position that “such [punitive] damages were not ‘received . . . on account of’ the personal injuries, but rather were awarded ‘on account of’ a defendant’s reprehensible conduct and the jury’s need to punish and to deter it.”⁷⁷ The Court, in deciding *Burke*, *Schleier*, and *O’Gilvie*, limited the reach of section 104(a)(2) but failed to provide a clear directive as to what is or is not a personal injury.⁷⁸ As a result, the

⁶⁹ *Schleier*, 515 U.S. at 337.

⁷⁰ See Harvey, *supra* note 5, at 341–42 (arguing that the Court erred when it failed to clearly define the term “on account of personal injuries or sickness,” and when it avoided “the ambiguity of the statute, which since its inception has caused the courts, taxpayers, and the [Internal Revenue] Service considerable consternation”) (emphasis added).

⁷¹ *O’Gilvie v. United States*, 519 U.S. 79 (1996).

⁷² *Id.* at 79.

⁷³ *Id.* at 81.

⁷⁴ *Id.*

⁷⁵ See *id.* at 86–90. See also F. Patrick Hubbard, *Making People Whole Again: The Constitutionality of Taxing Compensatory Tort Damages for Mental Distress*, 49 FLA. L. REV. 725, 742 (1997) (discussing congressional intent—as well as the *O’Gilvie* Court’s view of punitive damages vis-à-vis section 104(a)(2)—and concluding that there is “no evidence that congressional generosity or concern for administrative convenience provide a reason for exempting these noncompensatory [i.e. punitive] damages”) (internal quotations omitted).

⁷⁶ *O’Gilvie*, 519 U.S. at 82.

⁷⁷ *Id.* at 83.

⁷⁸ See Hobbs, *supra* note 11, at 53–54. See also Lee, *supra* note 53, at 551 (“Though the Court provided an appropriate grounding in the plain language of the Code, it failed to provide additional guidance as to the scope of the term ‘personal injuries’ for purposes of s[ection] 104(a)(2).”); Donald J. Zahn, *Personal Injury Exclusion: Is the Slashing of Wrists Necessary?*, 13 AKRON TAX J. 129, 148–49

perception that emanated from cases dealing with section 104(a)(2) is that a tort-based money damages award—received for personal physical *and* non-physical injuries—is excludable as long as the damages award was compensatory in nature, that is, not for liquidated or punitive damages.⁷⁹

D. *Amos v. Commissioner*

After Congress amended section 104(a)(2) in 1996 and added the “physical” requirement, the courts continued to struggle in setting the boundary lines between what is and what is not a personal injury.⁸⁰ In the 2003 *Amos* decision, the United States Tax Court addressed the severity required in order for an injury to qualify as “physical” within the meaning intended by Congress when it amended section 104(a)(2).⁸¹ In *Amos*, the taxpayer claimed that the entire award of damages received for an intentional kick to the groin was for physical injuries.⁸² The I.R.S. argued that none of the award should be excludable because the taxpayer’s physical injuries were minimal and because the authenticity of the injury was questionable.⁸³ The *Amos* court held that the taxpayer did suffer a physical injury as is required by section 104(a)(2).⁸⁴

The *Amos* court reasoned that “it is the nature and character of the claim settled, and *not its validity*, that determines whether the settlement payment is excludable from gross income under section 104(a)(2).”⁸⁵ The court also found that the defendant’s “dominant reason in paying the settlement amount at issue was petitioner’s claimed physical injuries as a result of the incident.”⁸⁶ In coming to its conclusion the court pointed to congressional intent behind the 1996 amendment where House members indicated that purely emotional distress injuries do not fall within the ambit of the amended section 104(a)(2) exclusion.⁸⁷ While the *Amos* court

(1997) (discussing the *Schleier* decision and stating that “[t]he Court, in effect, broke from traditional notions of personal injuries and many years of precedent. The Court indicated that even though tort litigation allows for liquidated damages, personal injuries of a legal character are not sufficient to bring the claim within the regulation.” Zahn goes on to discuss the negative impact of the three major Supreme Court decisions on section 104(a)(2)—*Burke*, *Schleier* and *O’Gilvie*—and concludes that “[a]s [federal trial and tax] courts diverted their focus from legal precedent to flights of fancy, the state of the personal injury award exclusion deteriorated”) (emphasis added).

⁷⁹ See Cory, *supra* note 10, at 251 (“Relying on a broad definition of ‘personal injury,’ many rulings and court decisions during the 1970’s and 1980’s allowed exclusion for damages arising from both physical and non-physical personal injuries.”).

⁸⁰ See *Burke & Friel*, *supra* note 15, at 168 (“[T]he remedy chosen to limit an overbroad statute, the drawing of a line between physical and nonphysical injuries, has introduced its own difficulties and is not supportable from a tax policy standpoint.”). See also Doti, *supra* note 5, at 62 (“Most of the confusion and controversy surrounding the personal injury exclusion would have been avoided if Congress had carefully considered the scope of the exclusion from its inception.”).

⁸¹ *Amos v. Comm’r*, T.C. Memo 2003-329 (2003).

⁸² *Id.* at 1, 5.

⁸³ *Id.* at 5.

⁸⁴ *Id.* at 6.

⁸⁵ *Id.* at 6 (emphasis added).

⁸⁶ *Id.*

⁸⁷ *Id.* at 4.

indicated that there had to be some physical injury, it failed to address lingering questions about the veracity of the underlying claim and the severity of the injury.⁸⁸ The court's decision left open the question of whether the physical injury, no matter how slight, had to take place as an immediate consequence of the tort or whether the physical injury could result from a manifestation of an emotional distress injury.⁸⁹

E. *Murphy v. Internal Revenue Service* (2006) (*Murphy I*)

Murphy I is a case that sent shock waves throughout the tax world.⁹⁰ In *Murphy I*, the Circuit Court of Appeals for the District of Columbia addressed the applicability of the amended section 104(a)(2)—and the “physical” injury requirement—to purely emotional distress tort damages awards.⁹¹ The *Murphy* taxpayer suffered emotional distress after she was “blacklisted” by her employer for filing a complaint against the employer with state authorities.⁹² The taxpayer subsequently filed a tort claim and received compensatory damages for emotional distress and injury to professional reputation.⁹³ The taxpayer then claimed that this money damages award was not taxable because it was received on account of physical injury because she had suffered some physical manifestations of her emotional injury, including “teeth grinding . . . which may cause permanent tooth damage[,] . . . shortness of breath, and dizziness.”⁹⁴ The taxpayer also claimed that section 104(a)(2) was unconstitutional because it failed to exclude revenue that is not income under the Sixteenth Amendment.⁹⁵

⁸⁸ See Cory, *supra* note 10, at 257–58 (arguing that the *Amos* court interpreted section 104(a)(2)'s physical injury requirement in a broader manner than that put forth by the IRS and concluding that “[i]n accordance with *Amos*, it can be argued that the minimum requirements of §104(a)(2) are that an individual must experience (1) a physical touching, and (2) that as a result of the touching the individual experiences pain”).

⁸⁹ See Doti, *supra* note 5, at 75 (“It is not clear, however, whether more serious physical manifestations of emotional distress, such as a nervous breakdown or heart attack, will constitute physical injury.”). See also Sharon E. Stedman, Note, *Congress's Amendment to Section 104 of the Tax Code Will Not Clarify the Tax Treatment of Damages and Will Lead to Arbitrary Distinctions*, 21 SEATTLE U. L. REV. 387, 405 (1997) (arguing that “the tax treatment of damages should not depend on the arbitrary distinction of whether the taxpayer is fortunate (or unfortunate) enough to have sustained a physical injury”).

⁹⁰ *Murphy v. I.R.S.*, 460 F.3d 79 (D.C. Cir. 2006) (“*Murphy I*”). See Germain, *supra* note 24, at 186 (declaring that “[a] unanimous panel of the Circuit Court of Appeals for the District of Columbia recently dropped a bombshell on federal income tax jurisprudence by holding in *Murphy v. I.R.S.* [(*Murphy I*)] that Congress violated the United States Constitution”) (emphasis added); Karpov, *supra* note 5, at 144 (stating that the *Murphy I* court “rendered a monumental decision that could change the course of damages taxation”); Wood, *Waiting to Exhale*, *supra* note 26, at 719 (stating that “in the wake of the first iteration of *Murphy*, many observers had feared that once one domino in the unconstitutionality chain toppled, virtually no tax would be safe from constitutional attack” and that because of *Murphy I* “there was near hysteria in some sectors about undermining the scope of congressional taxing powers”) (emphasis added).

⁹¹ *Murphy I*, 460 F.3d at 80–81.

⁹² *Id.* at 81.

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ *Id.* at 84–85.

The court disagreed with the taxpayer on the first point and decided that—despite the fact that the taxpayer “suffered from certain physical manifestations of emotional distress”—the written record indicated that the award she received was for “mental pain and anguish” and, therefore, was not received on account of a physical injury or sickness.⁹⁶ However, the court also held that section 104(a)(2) was unconstitutional because Congress was barred by the Sixteenth Amendment from taxing tort compensation that was deemed not income.⁹⁷ The court argued that the damages awarded to the taxpayer “were awarded to make Murphy emotionally and reputationally ‘whole’ and not to compensate her for lost wages or taxable earnings of any kind.”⁹⁸ This decision is important for two reasons. First, the court acknowledged that the word “physical” is ambiguous and that the possibility exists that some physical manifestations of emotional distress may rise to the level of the “physical” injury requirement of section 104(a)(2).⁹⁹ Second, the court concluded that section 104(a)(2)’s disparate treatment of similarly situated taxpayers is a violation of the Constitution because it excluded revenue that is deemed not income—such as that received for a physical injury—while it taxed other revenue that is similarly not income—such as that received for a physical manifestation of a non-physical injury.¹⁰⁰ The serious constitutional implications of *Murphy I* meant that its staying power was doubtful as is evidenced by the next episode in the *Murphy* saga.

F. *Murphy v. Internal Revenue Service (2007) (Murphy II)*

Less than one year after the *Murphy I* decision, the same Circuit Court of Appeals for the District of Columbia reversed itself in the *Murphy II* decision.¹⁰¹ Looking at the same facts and essentially the same arguments as in *Murphy I*, the court came to a remarkably different conclusion.¹⁰² Similar to its previous decision, the court first reasoned that the taxpayer’s

⁹⁶ *Id.* at 84.

⁹⁷ *Id.* at 88.

⁹⁸ *Id.*

⁹⁹ *See id.* at 84; Wood, *Waiting to Exhale*, *supra* note 26, at 721 (“At least *Murphy I* noted that the question [of what is a “physical” injury] was confusing and that the IRS has done little to remedy that posture.”). *See also* Karpov, *supra* note 5, at 149 (suggesting that the *Murphy I* court simply followed the Internal Revenue Service’s early interpretations of the predecessor to section 104(a)(2) where the Service determined that “compensation for a nonphysical tort . . . constituted a replacement, and not a gain, of human capital, and thus was not within the definition of income”).

¹⁰⁰ *Murphy I*, 460 F.3d at 89:

That emotional distress and loss of reputation were both actionable in tort when the Sixteenth Amendment was adopted supports the view that compensation for these nonphysical injuries was not regarded differently than was compensation for physical injuries and, therefore, was not considered income by the framers of the Amendment and the state legislatures that ratified it.

¹⁰¹ *Murphy v. I.R.S.*, 493 F.3d 170 (2007) (“*Murphy II*”). The court—acting on its own and without any prompting from either of the parties—vacated the *Murphy I* decision and ordered a rehearing where it ultimately overturned the *Murphy I* decision.

¹⁰² *Id.* at 171.

recovery was intended as recompense for emotional distress suffered and not for physical injury or physical manifestations of emotional distress.¹⁰³ Next, the court concluded that the imposition of tax on a tort damages award for non-physical personal injury did indeed fall within Congress' tax powers as prescribed by the Constitution and by the Supreme Court's interpretation of what constitutes income.¹⁰⁴ The court reached this conclusion by applying the sweeping definition of income conveyed by the Supreme Court in *Glenshaw Glass*.¹⁰⁵

The *Murphy II* court's sudden reversal led some scholars to speculate that the court was under considerable external pressure to reconsider its earlier position.¹⁰⁶ This speculation is bolstered by the reasoning and ease with which the *Murphy II* court dismissed the Sixteenth Amendment argument it applied in *Murphy I*.¹⁰⁷ It is important to note that the issue of what is a "physical" injury—and whether physical manifestations of emotional distress fall within the rubric of section 104(a)(2)—was not fully addressed in either opinion, leaving little judicial guidance on the meaning of "physical."¹⁰⁸ However, the court implied, in both cases, that the definition of "physical" and the applicability of the "on account of" phrasing of section 104(a)(2) remained questionable.¹⁰⁹ This section has highlighted the difficulty the courts have experienced in applying section 104(a)(2). The next section analyzes some of the underlying principles, as well as the tort law, medical science, ethical, and prudential reasons that make section 104(a)(2) an inequitable tax provision.

¹⁰³ *Id.* at 176.

¹⁰⁴ *Id.* at 179 ("[A]lthough the Congress cannot make a thing income which is not so in fact, . . . it can label a thing income and tax it, so long as it acts within its constitutional authority.").

¹⁰⁵ *Id.* (stating that I.R.C. § 61(a) defines gross income as "all income from whatever source derived" and citing *Comm'r v. Glenshaw Glass Co.*, 348 U.S. 426, 431 (1955), where the Supreme Court defined income to include all instances of undeniable accessions to wealth, clearly realized, and over which the taxpayers have complete dominion. This supported the Court's conclusion that Murphy's tort recovery is included in both the statutory and the interpretational definitions of income).

¹⁰⁶ See Germain, *supra* note 24, at 187–88 n.8; Karpov, *supra* note 5, at 182; Wood, *Waiting to Exhale*, *supra* note 26, at 724.

¹⁰⁷ See Wood, *Waiting to Exhale*, *supra* note 26, at 724–25:

The about-face, the court says, is attributable to a new and novel argument the IRS raised for the first time in its petition for rehearing en banc (of course, the court denied that motion). This new and novel argument, says the court, is that while Murphy's recovery may not be income within the meaning of the 16th Amendment, Congress can still tax it without violating the Constitution.

¹⁰⁸ See *id.* at 721 (stating that "[t]he first Murphy opinion did not answer the question of just what constitutes physical injuries" and that "the court doesn't offer any comments about this issue in the second go-round [in *Murphy II*]").

¹⁰⁹ Karpov, *supra* note 5, at 184 ("The *Murphy II* decision has not given an acceptable answer to the taxability of nonphysical personal injury awards because the court's reasoning was noticeably result-driven and weaved in and out of arguments with a motive to end with a taxable result.").

III. ANALYSIS: HOW THE “ON ACCOUNT OF” REQUIREMENT OF I.R.C.
§ 104(a)(2) LEADS TO THE DISPARATE TREATMENT OF SIMILARLY
SITUATED TAXPAYERS

The problem at the heart of the current iteration of section 104(a)(2) is in the seemingly arbitrary, one-way direction in which the exclusion applies.¹¹⁰ The three words “on account of” are interpreted to mean that *any* consequence stemming from an initial physical injury may be excludable regardless of how indirect the connection.¹¹¹ For example, a young girl is injured in an accident at a negligently supervised swimming pool and is partially paralyzed as a result. The child rightfully recovers for her injuries based on a tort claim against the owner of the pool.¹¹² Clearly, any damages—except for punitives—recovered by the young girl are fully excludable.¹¹³ However, the child’s parents also suffer purely emotional distress both from witnessing the accident and from dealing with the subsequent condition of their daughter. If the parents sue and recover damages for their purely emotional injuries, then, according to the *Schleier* decision, the parents are also allowed to exclude the recovered damages under section 104(a)(2).¹¹⁴ It is important to note that the parents’ money damages recovery is *not based on any physical injury or sickness* to either of the parents themselves. Rather, this seemingly counterintuitive result is based solely on the notion that section 104(a)(2) applies in a forward-looking manner. Any injury or harm that results after the initial physical injury—including a subsequent purely emotional distress injury to a different person, such as a parent—may be excludable because that subsequent injury was suffered “on account of” the initial personal physical injury to the child.¹¹⁵

¹¹⁰ See *id.* at 149–52 (explaining that prior to the addition of the “physical” injury requirement, section 104(a)(2) applied to people who suffered physical manifestations of emotional distress). See also Stedman, *supra* note 89, at 387–89 (stating that Congress slightly over-reacted when it added the “physical” requirement in the 1996 amendment and that this over-reaction was likely due to the fact that Congress wanted to clarify the terms of the exclusion. Congress also acted out of fear that the tax code might be abused by people who filed and recovered tort damages based on fraudulent emotional distress claims. Stedman argues that the addition of the “physical” requirement swung the pendulum too far in the other direction thus leading to the disparate treatment of people who recover tort damages for legitimate, verifiable physical manifestations of emotional distress by denying them the benefits of the exclusion. Stedman concludes that “[t]he ambiguous term ‘personal injury’ has been replaced by ‘personal physical injury,’ a term that is equally ambiguous”).

¹¹¹ See Wood, *Waiting to Exhale*, *supra* note 26, at 722 (explaining that “[a] payment can be ‘on account of’ physical injuries or sickness even if the plaintiff is not injured but recovers on behalf of an injured party” and stating that “[e]xamples include recoveries for loss of consortium (based on physical injury to a spouse) and wrongful death”).

¹¹² The relevant tort claim would likely be based in negligence and/or products liability.

¹¹³ See *supra* Part II.B. The young girl’s recovery would qualify for the section 104(a)(2) exclusion because—under the *Schleier* test—her claim is a tort or tort-like claim and she suffered a significant physical injury. In addition, the 1989 amendment to section 104(a)(2) prohibited punitive damages from excludability.

¹¹⁴ Wood, *Waiting to Exhale*, *supra* note 26, at 722–23.

¹¹⁵ *Id.*

The converse situation is where the real problem arises.¹¹⁶ If a person suffers a purely emotional injury that later causes or leads to the manifestation of a *verifiable* physical injury, then any recovery that person receives, whether for the emotional or the physical injury, is *not* excludable under section 104(a)(2) and, thus, is fully taxable.¹¹⁷ For example, an employee is fired for acting as a whistle-blower and revealing fraudulent business practices to a regulatory agency. As a result of the increased stress caused by the firing, the employee cannot find another job and subsequently files suit for wrongful termination. The employee continues to suffer emotional problems and is soon diagnosed with Post Traumatic Stress Disorder (“PTSD”) and severe depression. Within six months of being fired, the employee begins to suffer serious physical manifestations of emotional distress. These physical manifestations include serious weight loss, hypertension, and stomach ulcers. The employee eventually wins the wrongful termination lawsuit and recovers money damages for the emotional and physical injuries, as well as for lost wages and punitive damages. However, because the *initial* injury was purely emotional the wrongfully terminated employee will have to pay tax on the *total* recovery.¹¹⁸

When Congress enacted the 1996 amendments to section 104(a)(2) and added the word “physical,” its primary intention was to eliminate the exclusion for purely emotional injuries and to narrow the exclusion to those who suffered real, verifiable physical injuries.¹¹⁹ This change in direction was based largely on Congress’ distrust of tort claimants who could “easily” falsify or exaggerate an emotional injury in order to recover money damages.¹²⁰ However, Congress left the issue of the origin of the physical injury in question.¹²¹ If the intent of Congress was to provide tax relief for people who suffered physical injuries, then the way the physical injury came about should be, in theory, secondary to the fact that there is

¹¹⁶ See Hubbard, *supra* note 75, at 745 (questioning the motive and propriety of Congress’s actions in limiting the applicability of section 104(a)(2) to those situations where the tort recovery is on account of a physical injury while physical manifestations of emotional distress are not excludable).

¹¹⁷ See Whittington, *supra* note 13, at 176 (“The tax code also does not allow a taxpayer to exclude amounts received for physical injuries that flow from an emotional injury.”).

¹¹⁸ See generally Wood, *Tax Treatment of Settlements*, *supra* note 38.

¹¹⁹ Doti, *supra* note 5, at 72–74 (discussing Congress’ intent to narrow the scope of section 104(a)(2) as the major driving force behind the 1996 amendments).

¹²⁰ See Leeper, *supra* note 16, at 1057 (arguing that, in enacting the 1996 amendments to section 104(a)(2), “Congress [may have] feared a flood of frivolous claims (i.e., suits alleging that a large portion of damages came from emotional distress and other difficult-to-prove, nonphysical injuries) were it to allow the exclusion to encompass all types of harms”). See also Hobbs, *supra* note 11, at 88 (“A variety of reasons were given [as the intent behind the 1996 amendments to section 104(a)(2)], including concerns over proof and valuation, the fear of fraudulent claims, and a flood of litigation.”).

¹²¹ See Cory, *supra* note 10, at 248 (stating that “[d]ue to the vague nature of §104(a)(2), the courts have had difficulty limiting the scope of, as well as defining, ‘personal injuries’ under §104(a)(2)” and that “deciphering what constitutes a ‘personal physical injury’ has remained largely ambiguous”).

indeed a physical injury.¹²² The intent of Congress would be satisfied as long as there is a provable causal link between the initial incident and the subsequent physical injury thus making the recovery excludable under section 104(a)(2).¹²³

The United States Tax Court addressed the issue of verifiability and the possibility of fraud in the *Amos* case.¹²⁴ The *Amos* court made a rather revealing statement when it declared: “[w]hether the settlement payment is excludable from gross income under section 104(a)(2) depends on the nature and character of the claim asserted, *and not upon the validity of that claim.*”¹²⁵ So, a physically injured individual whose claim may or may not be valid is fully entitled to reap the benefits of the section 104(a)(2) exclusion.¹²⁶ However, an individual who has suffered a legitimate, verifiable physical manifestation of an emotional distress injury is not entitled to any favorable tax treatment regardless of the validity of that person’s claim.¹²⁷ The absurdity of these highly disparate outcomes is self-evident.¹²⁸ If the issue of verifiability was Congress’ main reason for adding the “physical” injury requirement, then the fact that *invalid* claims can and do exist in both purely physical and purely emotional injury claims shows that this was not likely the predominant reason why Congress added the “physical” requirement.¹²⁹

If two equally legitimate tort claims arise, one stemming from an initial physical injury, the other emotional at first and physical later, then both claimants should be entitled to the same treatment under the tax law. If the intent of Congress was to provide tax relief for people who suffered physical injuries, then it should not matter how the physical injury came about. As long as there is a provable causal link between the initial tort for which money damages were obtained and the subsequent physical injury.¹³⁰

¹²² Stedman, *supra* note 89, at 408–09 (discussing the absurdity of distinguishing between different types of physical injuries).

¹²³ See Leeper, *supra* note 16, at 1065 (arguing that a causal link between the tort and the injury was the primary method used before the 1996 amendment and concluding that “by requiring courts to distinguish between physical and nonphysical injury and examine sometimes unclear chains of causation, the ambiguity of the physical injury requirement of section 104(a)(2) creates the possibility of arbitrariness in application”).

¹²⁴ *Amos*, T.C. Memo 2003-329 at 5–6. The petitioner in *Amos* suffered an internal injury to the groin, however, there were no visible signs of injury and the severity of the injury was highly questionable. For a detailed analysis of *Amos*, see *supra* Part II.D.

¹²⁵ *Id.* at 5 (emphasis added).

¹²⁶ *Id.*

¹²⁷ See Whittington, *supra* note 13, at 176–77.

¹²⁸ See e.g. Stedman, *supra* note 89, at 405 (stating that “[a] victim of . . . sexual harassment has suffered no less of an injury than a taxpayer who broke a leg in a car accident” and concluding that “Congress should not treat a taxpayer who is a victim of sexual harassment differently from a taxpayer who is a victim of a car accident”).

¹²⁹ See Leeper, *supra* note 16, at 1071 (“Whatever the validity of such claims, requiring a showing of a physical injury for tax exclusion can only create the potential for disputes and differing treatment of taxpayers who have suffered similar harms.”).

¹³⁰ See generally Hobbs, *supra* note 11, at 55–56 (discussing the motives and the intent of Congress in adding the “physical” injury requirement and concluding that section 104(a)(2)’s disparate

However, section 104(a)(2) essentially discriminates between similarly situated taxpayers and leads to inequitable treatment in a system founded on principles of uniformity.¹³¹ While several courts have faced the question of whether section 104(a)(2) should apply to tort damages awarded for a physical manifestation of an emotional distress injury, to date, in *all* of these decisions these courts have declined to extend the application of section 104(a)(2) to physical manifestations of emotional distress.¹³² It is worth noting, however, that none of these cases dealt with a relatively serious form of physical manifestation caused by emotional distress. The unidirectional—and inherently inequitable—impact of section 104(a)(2) is best illustrated through the real-life examples presented in the next section.

A. Hypothetical Situations: Two Detailed Scenarios that Illustrate the Inequity of Section 104(a)(2)

Elena is a 42 year old single woman. She is a full-time employee at a leading biotech company where she is a clinical data manager. Elena's immediate supervisor, Richard, has repeatedly made unwelcome sexual advances toward her. Richard has frequently touched Elena despite her constant refusal to submit to his advances. Richard's non-consensual

treatment of similarly situated taxpayers may only be resolved via a complete repeal of section 104(a)(2)).

¹³¹ See *Infanti supra* note 23, at 1198–201 (discussing the notion that the Tax Code was founded on principles of equity which entail uniform treatment of similarly situated individuals who are subject to the provisions of the Code).

¹³² See *Germain, supra* note 24, at 204–05 n.82 (supporting the claim that no court has yet to apply the section 104(a)(2) exclusion to physical manifestations of emotional distress by citing several decisions including:

Lindsey v. Commissioner, 422 F.3d 684, 688 (8th Cir. 2005) (holding business tort claim recoveries not excludable because taxpayer's "hypertension and stress-related symptoms, including periodic impotency, insomnia, fatigue, occasional indigestion, and urinary incontinence . . . relate to emotional distress, and not to physical sickness"); *Goode v. Commissioner*, 91 T.C.M. (CCH) 901 (2006) (rejecting "physical injury" characterization in settlement agreement, and holding that employee's claimed "repeated, vehement verbal assaults" which caused him physical pain and suffering did not constitute physical injury); *Ndirika v. Commissioner*, T.C.M. (RIA) 2004-250 (2004) (holding there was no credible evidence employment settlement to compensate for previous miscarriage or pregnancy-related pain and suffering); *Lindsey v. Commissioner*, 87 T.C.M. (CCH) 1295 (2004) (holding that hypertension leading to "fatigability, occasional indigestion, and difficulty sleeping—are the types of injuries or sicknesses that Congress intended to be encompassed within the definition of emotional distress"); *Shaltz v. Commissioner*, 85 T.C.M. (CCH) 1489 (2003) (holding that sexual harassment settlement is not excludable because no evidence of physical injury, only "depression, anxiety, stress, and recurrent past stressors" caused by "depression and trauma at work"); *Henderson v. Commissioner*, 85 T.C.M. (CCH) 1469 (2003) (holding that \$5,000 settlement of credit reporting claim not excludable despite "life-threatening, pre-existing physical illness" exacerbated by the harm to his personal reputation); *Prasil v. Commissioner*, 85 T.C.M. (CCH) 1124 (2003) (holding that allegations of severe physical manifestations from sex discrimination and harassment not sufficient to establish physical injury or physical sickness); *Witcher v. Commissioner*, 84 T.C.M. (CCH) 582 (2002) (holding that physical manifestations caused by defamation and business torts are not physical injury or physical sickness)).

touching, however, does not cause any physical injuries to Elena. Richard has also consistently threatened Elena that he will make her work life miserable if she does not submit to his sexual advances. Elena endures this harassment for over two years and she starts seeing a psychiatrist because she experiences elevated stress levels, mild anxiety, and depression. She files complaints with the management, as well as human resources, but nothing is done. Eventually, the stress causes Elena to start missing work. Finally, Elena's job performance deteriorates to the point where she is placed on probation and eventually fired. Elena's life is turned upside down as she struggles to make ends meet. Her anxiety worsens and her depression becomes so severe that she sometimes spends the entire day in bed without a will or an inclination to do anything productive. In addition, Elena eats very little, begins to lose weight, and has acute insomnia.

Ultimately, all these emotional injuries begin to take a serious toll on Elena's physical health. She loses forty pounds of weight—to the point where her doctors classify her as anorexic, she develops several stomach ulcers that lead to moderate digestive system pain, and she has recurring migraine headaches of moderate to extreme severity that vary in duration from a few minutes to several hours at a time. Elena continues to see her psychiatrist throughout this ordeal. Her relatives convince her to find an attorney and she eventually files a sexual harassment lawsuit against her former employer. Her case is strengthened by the fact that the discovery process exposes two other sexual harassment complaints against her previous supervisor. In addition, Elena's attorney finds out that her previous supervisor was fired subsequent to an internal investigation that revealed highly damaging evidence. Elena's lawyer and her previous employer eventually reach a settlement agreement whereby Elena is paid \$200,000 in damages. The settlement agreement states that the payment is for Elena's pain, suffering, and lost wages.

Under the current interpretation and application of section 104(a)(2), Elena will have to pay taxes on the entire amount of the settlement. This is due to the fact that her initial injuries were purely emotional and that her physical injuries manifested themselves as a result of the emotional distress and not as a result of the sexual harassment itself. Elena's injuries started out as depression and anxiety but these conditions eventually manifested themselves into serious physical conditions such as the pain from the ulcers, the migraine headaches, and the dangerous drop in body weight. Nonetheless, Elena is taxed *on the entire amount of her recovery* because her recovery was not based "on account of a physical injury or sickness." After paying attorney's fees (33 percent) and federal taxes (35 percent), Elena is left with approximately \$87,000—less than half of what she recovered.

Now, a look at the other end of the spectrum. Ron is a healthy young man who works as a delivery truck driver. He exercises on a regular basis at his local fitness center. During one of his exercise sessions Ron is

injured when a weightlifting machine malfunctions. His injuries are not severe but he is unable to drive as a result of pain in his arms and shoulders. Ron files a negligence tort claim against the owner of the fitness center. Ron seeks damages for pain, suffering and lost wages. The two sides eventually reach a settlement agreement through non-binding arbitration. The settlement is for \$200,000.

The tax treatment for Ron's recovery is straightforward. Under section 104(a)(2) Ron *does not have to pay any tax* on the amounts recovered by virtue of the fact that his recovery was on account of a physical injury. After attorney's fees (33 percent), Ron is left with \$134,000. Consequently, Ron's injuries healed fairly quickly and he returned to work three months after the accident. Elena, on the other hand, is faced with several years of medical and psychiatric treatment before she can return to the job market. In addition, the long term damage to Elena's physical and psychological well being will impact her quality of life into the foreseeable future.

The unfair tax treatment afforded to these two hypothetical people shows the lack of uniformity and the inequity that lies at the heart of section 104(a)(2). It is a given that both Ron and Elena suffered injuries for which they were entitled to tort damages recovery. However, the seriousness of the injuries suffered and the long term impact on the lives of these two people are clearly different. Elena keeps just over one-third of what she received while Ron keeps two-thirds. However, Elena is faced with a long and arduous uphill battle before she returns to any semblance of her life as she used to know it, while Ron returns to work and leads a normal life after a short period. There is no logical tax policy or humanitarian reason why Elena has to pay income taxes on her recovery while Ron pays no income taxes on his. In contrast to the way section 104(a)(2) treats physical manifestations of emotional distress, medical science has established a link between emotional and physical well-being.¹³³

B. Medical Science: There is a Nexus Between Emotional Injuries and Physical Manifestations

Through the one-sided application of section 104(a)(2), the Code treats equally situated individuals differently by failing to recognize that legitimate physical injuries can result from emotional distress.¹³⁴

¹³³ See *infra* Part III.B.

¹³⁴ See Leeper, *supra* note 16, at 1067–68 (“Not only is the statute [section 104(a)(2)] itself arbitrary by drawing without logical support a distinction between sufferers of physical and nonphysical injury, but by doing so it also creates ambiguity and opens the door to interpretation and thus potentially arbitrary application by the judicial system.”); Timothy R. Palmer, *Internal Revenue Code Section 104(a)(2) and the Exclusion of Personal Injury Damages: A Model of Inconsistency*, 15 J. CORP. L. 83, 127 (1989) (arguing that section 104(a)(2) was applied—both by the courts and by the I.R.S.—in an arbitrary manner and that “[t]he time has come for the adoption of consistent principles in determining

Historically, the link between mental and physical health was the subject of much research and hypothesizing.¹³⁵ Philosophers, including Aristotle, Descartes, and Kant, theorized about the relationship between mind and body with the debate centering on whether the mind and the body are inseparably linked, or whether each exists notwithstanding the other.¹³⁶ However, in modern times, the world of physical and psychological medicine has reached a consensus that an affirmative relationship exists between emotional distress and physical manifestations of that distress.¹³⁷ Many members of the psychological care community continue to emphasize the importance of studying, understanding, and addressing the impact of emotional distress on the physical condition of a patient.¹³⁸

There are many examples of physical ailments that can be brought on by emotional distress.¹³⁹ One author, Shelley Taylor, states that “[a]nxiety, depression, and other psychological disorders are accompanied by a number of physical symptoms.”¹⁴⁰ Taylor goes on to provide some examples: “Anxiety can produce diarrhea, upset stomach, sweaty hands, shortness of breath, difficulty in sleeping, poor concentration, and general agitation. Depression can lead to fatigue, difficulty in performing everyday activities, listlessness, loss of appetite, and sleep disturbances.”¹⁴¹ In addition, many of these symptoms, such as loss of appetite, anxiety, and stress, can lead to other, more serious physical injuries, such as stomach ulcers, high blood pressure, diabetes, and extreme pain.¹⁴² Accordingly, the Code should not treat these serious physical injuries differently, regardless

the tax consequences of a damage award received for personal injuries”).

¹³⁵ See generally JOHN G. BENJAFIELD, A HISTORY OF PSYCHOLOGY (2d ed. 2005).

¹³⁶ See JACK H. ORNSTEIN, THE MIND AND THE BRAIN 3–19 (1972) (discussing Descartes’ view of the mind-body relationship). See also BENJAFIELD, *supra* note 135, at 15–42 (detailing the development and evolution of the modern understanding of the mind-body relationship and the origins of the study of psychology, including an examination of the views of Aristotle, René Descartes and Immanuel Kant).

¹³⁷ See SHELLEY E. TAYLOR, HEALTH PSYCHOLOGY 3–7 (2d ed. 1991) (describing the scientific belief, spanning many centuries, that there is an integral mind-body connection, one through which harm to the mind can lead to physical injury and declaring that “[i]t is now known that physical health is inextricably interwoven with the psychological and social environment: All conditions of health and illness . . . are influenced by psychological and social factors”). See also Burke & Friel, *supra* note 15, at 179 (stating that “scientific research has brought an awareness of the potentially devastating physical impacts of psychological injuries”); Hubbard, *supra* note 75, at 751–52 (discussing the valid and serious implications of emotional distress on the physical well-being and stating that “mental trauma is a real experience; those who suffer mental distress suffer an actual loss”).

¹³⁸ Infanti, *supra* note 23, at 1237–38 (discussing a recent trend by psychologists to study the mental health effects of discrimination and concluding that “[a]mong the negative mental health effects reported by these studies were emotional distress, depression, obsessive-compulsive symptoms, anxiety, stress, and somatization (i.e., the conversion of anxiety into physical symptoms)”).

¹³⁹ See, e.g., Germain, *supra* note 24, at 193 (describing the physical ailments that afflicted the petitioner in *Murphy v. I.R.S.* including “anxiety attacks, shortness of breath, and dizziness” that resulted from Murphy’s initial emotional distress injury).

¹⁴⁰ See TAYLOR, *supra* note 137, at 275–76.

¹⁴¹ *Id.* at 276.

¹⁴² *Id.* at 192, 347 (stating that stress can lead to an increase in heart rate, blood pressure and blood sugar and adding that “anxiety-producing conditions . . . can lead to the formation of ulcers” while “anxiety can increase pain”).

of whether they are caused by an initial physical injury or subsequent to an emotional distress injury.¹⁴³ The expansive field of tort law has also recognized the integral relationship between emotional distress and physical manifestations of that emotional distress.¹⁴⁴

C. Tort Law: Physical Manifestations of Emotional Distress are Valid Forms of Tort Recovery

As opposed to the relative ease of recovery in physical injury claims, the field of tort law has grappled with the notion of plaintiff recovery for *purely* emotional distress—and physical manifestations of emotional distress—injuries.¹⁴⁵ The historical progression of emotional distress recoveries in tort law has evolved from one with an absolute physical injury requirement to the modern iteration of stand-alone emotional distress claims.¹⁴⁶ This transformation took place despite the reluctance of some courts to acknowledge purely emotional distress claims primarily due to the misperception that the number of fraudulent claims would increase.¹⁴⁷ Today, the notion that non-frivolous tort claims can be brought for purely emotional distress claims—in addition to claims based on physical manifestations of emotional distress—has become a well-established rule.¹⁴⁸

Analysis of the “on account of” language in section 104(a)(2)—in the context of tort law—has frequently pointed to causation as an issue in

¹⁴³ See Doti, *supra* note 5, at 79–80; Burke & Friel, *supra* note 15, at 195–96; Stedman, *supra* note 89, at 405; Whittington, *supra* note 13, at 185.

¹⁴⁴ See *infra* Part III.C. See also Whittington, *supra* note 13, at 185 (“As the areas of . . . tort law . . . and medicine have taken steps towards the recognition that emotional distress is a very real injury, Congress[, by adding the physical requirement to section 104(a)(2),] has caused the tax code to take a tremendous leap backwards.”).

¹⁴⁵ KENNETH S. ABRAHAM, *THE FORMS AND FUNCTIONS OF TORT LAW* 29–30 (3d ed., Foundation Press 2007) (1997). See also Calvert Magruder, *Mental and Emotional Disturbance in the Law of Torts*, 49 HARV. L. REV. 1033, 1035 (1936) (describing the state of emotional distress claims during the early history of tort law and stating that “it is nevertheless true that the common law has been reluctant to recognize the interest in one’s peace of mind as deserving of general and independent legal protection, even as against intentional invasions”).

¹⁴⁶ See generally JON R. ABELE, *EMOTIONAL DISTRESS: PROVING DAMAGES* (2003). See also JOHN D. WINER ET AL., *PROVING MENTAL & EMOTIONAL INJURIES* §§ 2:31–2:33 (James Publishing 2005).

¹⁴⁷ See ABELE, *supra* note 146, at 70. See also Hobbs, *supra* note 11, at 88 (“Courts also discovered that their earlier concerns with proof and valuation were overstated; such claims more often than not presented the same difficulties as other components of personal injury.”); Leeper, *supra* note 16, at 1070 (arguing that the fear of frivolous lawsuits, evident in Congress’ addition of the “physical” requirement to section 104(a)(2), is “unfounded” because tort litigants are more likely to be motivated by the tort recovery itself rather than the tax consequences of such recovery).

¹⁴⁸ ABELE, *supra* note 146, at 70 (explaining that most “[j]urisdictions that impose the physical manifestation rule require that the plaintiff exhibit a physical injury from the emotional distress suffered”). See also Hobbs, *supra* note 11, at 88 (discussing the evolution of tort law and the transition of most courts from resistance to acceptance of purely emotional distress causes of action and concluding that “[t]oday, a clear majority of states recognize emotional distress as an independent cause of action”); Whittington, *supra* note 13, at 174–76 (discussing the fact that many courts have moved away from the physical manifestation rule to more liberal rules allowing purely emotional injury claims and that the “claimant must still prove he has suffered an emotional distress type injury”).

2009]

Heads I Win, Tails You Lose

183

proving that a physical injury was caused by emotional distress.¹⁴⁹ However, the vast majority of courts have recognized that a legitimate causal link can and often does exist between an initial emotional distress incident and subsequent physical manifestations.¹⁵⁰ Clearly, the field of tort law has accepted and even embraced the notion that legitimate, provable cases of physical manifestations of emotional distress are deserving of the very same monetary damages recoveries afforded to those who suffer purely physical injuries.¹⁵¹

Unfortunately, under the current interpretation and application of section 104(a)(2), the tax code runs contrary to tort law—and, by extension, to the world of medical science—because the tax code denies the personal injury tax exclusion for individuals who suffer real, verifiable physical injuries that manifest themselves as a consequence of emotional distress.¹⁵² In addition to medical science and tort law, there are two conceptions of justice—horizontal equity and return of human capital—that illustrate the unfairness of section 104(a)(2).

D. Horizontal Equity and Human Capital: How I.R.C. § 104(a)(2) is

¹⁴⁹ See Lee, *supra* note 53, at 552, 555 (stating that “a causal connection must exist between the personal injury and the damages received” and that the “on account of language ostensibly employs a causal analysis”); Whittington, *supra* note 13, at 161 (“When pleading, proving or settling emotional distress damages in a tort suit, the tort practitioner is well advised to make certain the emotional distress damages are causally linked to a demonstrable and recoverable physical injury. Otherwise, the client’s emotional distress recovery will be taxable.”).

¹⁵⁰ See Haralson v. State Farm Mut. Auto. Ins. Co., 564 F. Supp. 2d 616, 622 (N.D. Tex. 2008) (stating that “[t]he clear weight of authority holds that physical symptoms of emotional distress constitute a “bodily injury” in the insurance context” and citing numerous cases in support of that statement including:

Twin City Fire Ins. Co. v. Colonial Life & Accident Ins. Co., 124 F. Supp. 2d 1243, 1247 (M.D. Ala. 2000) (emotional trauma can constitute “bodily injury” if accompanied by physical symptoms); General Star Indemnity Co. v. Schools Excess Liability Fund, 888 F. Supp. 1022, 1027 (N.D. Cal. 1995) (physical injury resulting from emotional distress constitutes “bodily injury”); Garvis v. Employers Mutual Casualty Co., 497 N.W.2d 254, 257 (Minn. 1993) (emotional distress with appreciable physical manifestations can qualify as a “bodily injury” within the meaning of an insurance policy); Voorhees v. Preferred Mutual Ins. Co., 607 A.2d 1255, 1262 (N.J. 1992) (emotional distress resulting in headaches, stomach pains, nausea, and body pains constituted “bodily injury” under homeowner’s policy); Western Casualty & Surety Co. v. Waisanen, 653 F.Supp. 825, 832 (D.S.D. 1987) (emotional distress accompanied by high blood pressure was sufficient to establish physical harm); Kufalk v. Hart, 636 F. Supp. 309, 311–12 (N.D. Ill. 1986) (emotional distress resulting in high blood pressure, headaches, stomach pains, and diarrhea constituted “bodily injury” under general liability policy); McGuire v. American States Ins. Co., 491 So.2d 606, 608 (Fla. Dist. Ct. App. 1986) (mental distress accompanied by headaches and muscle spasms, requiring treatment and medication, could be construed as “bodily injury” under insurance policy); Holcomb v. Kincaid, 406 So.2d 646, 649 (La. Ct. App. 1981) (emotional distress resulting in rash, hair loss, weight loss, and symptoms of stroke fell within definition of “bodily injury” in homeowner’s policy).

¹⁵¹ See ABELE, *supra* note 146, at 70–71; Germain, *supra* note 24, at 196; Hubbard, *supra* note 75, at 748.

¹⁵² See, e.g., Murphy v. I.R.S., 493 F.3d 170 (D.C. Cir. 2007) (“*Murphy II*”). See also Burke & Friel, *supra* note 15, at 168 (“[N]o substantial policy-based justification was advanced for distinguishing between physical and nonphysical injuries.”).

Inherently Unfair

Horizontal equity and return of human capital are two inter-related notions of basic equity and justice that permeate the foundation of our various systems of laws, including the tax code.¹⁵³ Horizontal equity describes the “notion of justice that similarly situated people should be treated equally and that people with equal incomes should pay equal taxes.”¹⁵⁴ Looking back at Elena and Ron in the hypothetical situations above, it is obvious that section 104(a)(2) violates the concept of horizontal equity because it treats similarly situated individuals in an inequitable manner. The concept of return of human capital involves the notion that money received to help restore a person’s body and mind to its state before the tortious act took place should not be taxed because there is no gain—and therefore no taxable income—for that individual.¹⁵⁵ Again, in the two hypothetical situations, Elena and Ron both received tort damages awards to restore them to their previous state of well-being. Yet, the fact that one pays taxes and the other does not further illustrates the inequity of section 104(a)(2).

This analysis has focused on the inequitable treatment of similarly situated individuals under section 104(a)(2). Those individuals are ones who suffered real, verifiable physical injuries and recovered money damages awards under established notions of tort law. Yet, this analysis has shown that these individuals are treated differently by section 104(a)(2) simply because of the timing of the physical injury suffered. If the tort directly caused the physical injury, then the money damages received are not taxed. However, if the tort caused purely emotional distress that later leads to physical manifestations of that distress, the money damages received are fully taxed. This analysis has presented two hypothetical situations to illustrate the nature and severity of this inequity. In addition, the evidence presented shows that both medical science and the field of tort law have recognized that a positive link does exist between emotional distress and physical manifestations of that distress. Subsequently, basic notions of fairness and justice—horizontal equity and return of human capital—indicate that section 104(a)(2) treats similarly situated taxpayers in a disparate manner. Finally, numerous courts have faced this very issue and, out of deference to the perceived congressional intent discussed throughout this comment, have refused to extend the application of section 104(a)(2) to physical manifestations of emotional distress.¹⁵⁶ The next section introduces a proposal that can resolve this inequity and continue to

¹⁵³ See Lederman, *supra* note 32, at 701–04 (discussing the nature of horizontal equity and its influence on tax policy). See also Doti, *supra* note 5, at 63–65 (explaining the significance of human capital as a basis for fair, equitable treatment under the tax code).

¹⁵⁴ Chapman, *supra* note 6, at 407 n.3.

¹⁵⁵ See Doti, *supra* note 5, at 63–65.

¹⁵⁶ See Germain, *supra* note 24, at 204–05 n.82 (citing several examples of decisions where the courts have refused to apply section 104(a)(2) to physical manifestations of emotional distress).

2009]

Heads I Win, Tails You Lose

185

uphold the intentions of Congress when it created, modified, and retained section 104(a)(2).

IV. PROPOSAL: A SIMPLE ADDITION TO I.R.C. § 104(a)(2) WILL LEAD TO
MORE EQUITABLE TAX LAW TREATMENT FOR SIMILARLY SITUATED
TAXPAYERS

Based on the authorities presented in this comment it is clear that the unidirectional application of section 104(a)(2) is inequitable, inherently unfair to similarly situated individuals and should be revised. This comment has not argued, nor does it support the proposition that purely emotional injury tort recoveries should be excluded under section 104(a)(2). The intent of Congress was clear in that the drafters of the 1996 amendment to section 104(a)(2) were cognizant of the possibility that purely emotional injuries are very difficult to prove and may often be based on fraudulent claims. Instead, this comment focuses on one aspect of the application of section 104(a)(2): the fact that section 104(a)(2) does not apply to individuals who suffer legitimate, verifiable physical manifestations of emotional injuries and that section 104(a)(2) has led to a lack of uniformity and inequitable treatment under the tax code.

This comment proposes that a simple addition to section 104(a)(2) will resolve the inequitable treatment currently rendered upon similarly situated taxpayers. The semi-colon at the end of section 104(a)(2) should be deleted and the following phrase should be added: “or personal physical injuries proximately caused by emotional distress;” This simple addition would allow those who have legitimate claims and subsequent recoveries to exclude from their taxes those amounts received on account of the physical injuries proximately caused by the emotional distress. The proximate causation requirement mandates that the physical manifestation be proven according to the preponderance of evidence burden of proof currently used in civil law courts.¹⁵⁷ This eight-word addition to section 104(a)(2) would reinstate the balance of equity for those who have suffered verifiable physical injuries regardless of whether the physical injury came first or resulted from an emotional distress injury.

This addition would also preserve the underlying intentions of Congress when it enacted, modified, and retained section 104(a)(2). The evidence presented shows that Congress intended to help those who suffered physical injuries and that Congress expressly did not intend the exclusion to apply to those who suffered purely emotional distress injuries. This proposal maintains that intention because it allows those who suffer physical manifestations of emotional distress to enjoy the benefits of the section 104(a)(2) exclusion while still maintaining the prohibition of the exclusion for those who suffer purely emotional distress injuries.

¹⁵⁷ See Leeper, *supra* note 16, at 1068.

CONCLUSION

This comment has examined the lack of uniformity with which section 104(a)(2) is applied to similarly situated taxpayers. The argument centered around the fact that those who suffer real, verifiable physical manifestations of emotional distress do not receive the same tax treatment as those who suffer purely physical injuries. If section 104(a)(2) was intended to alleviate the tax burden for those who suffer physical injuries, then it should not matter whether the physical injury took place immediately after the tort or later in time. As long as an affirmative causal link is established between the initial emotional distress injury and the subsequent physical manifestations, then the section 104(a)(2) exclusion should apply.

The background and history of section 104(a)(2) show that it has experienced a rather tumultuous existence. This turbulent existence became more evident through an analysis of the case decisions that have shaped the modern understanding of section 104(a)(2). Evidence was presented to demonstrate that there are solid legal, medical, prudential, and ethical grounds to show that section 104(a)(2) does indeed treat similarly situated taxpayers differently. A simple proposal was offered—allowing application of section 104(a)(2) to those who have physical manifestations proximately caused by emotional distress—in order to make section 104(a)(2) more equitable and uniform in application. Allowing someone who suffers legitimate physical manifestations of emotional distress to receive the same tax treatment as someone who suffers purely physical injuries will *reinforce* the notion that the tax code is founded on basic notions of fairness and uniformity.

Chemical Castration for Child Predators: Practical, Effective, and Constitutional

*Elizabeth M. Tullio**

If legislation and punishment alone cannot fully solve the problem, medicine and science need to be called into action. And if society can be made safer by such means, why not use them?¹

INTRODUCTION

Every year approximately 100,000 to 500,000 children are sexually molested in the United States.² This results in 10–25% of children being sexually abused by the age of eighteen—out of which 30–40% are females and 10–15% are males.³ With such high rates, child sexual abuse can be classified as an epidemic across the country. Friends, family members, and strangers commit these crimes.⁴ There are those who are attracted solely to children and others who are mainly attracted to adults, but occasionally attack children.⁵ While a number of different punishments and treatments have been implemented, there is no cure for such a condition.⁶ People are left wondering what to do, how to stop these offenders, and how they can keep their young children safe from such unspeakable horror.

* B.A., University of California, Irvine; J.D. expected May 2010, Chapman University School of Law. I would like to thank my family for their love and support throughout this process. I would also like to thank Professor Jennifer Robinson in the School of Social Ecology at the University of California, Irvine for all of her help and guidance in putting this article together. See also Peter Tullio, *Jurisdiction Obtained by Forcible Abduction: Reach Exceeds Due Process Grasp*, 67 J. CRIM. L. & CRIMINOLOGY 181 (1976) (the author of this article has been a source of motivation and inspiration in my life, without whom none of this would be possible).

¹ Karen Harrison, *The High-Risk Sex Offender Strategy in England and Wales: Is Chemical Castration an Option?*, 46 HOW. J. CRIM. JUST. 16, 28 (2007) (quoting F.S. Berlin).

² Bhagwan A. Bahroo, *Pedophilia: Psychiatric Insights*, 41 FAM. CT. REV 497, 497 (2003); Ariel Rosler & Eliezer Witztum, *Treatment of Men with Paraphilia with a Long-Acting Analogue of Gonadotropin-Releasing Hormone*, 338 NEW ENG. J. MED. 416, 416 (1998).

³ Bahroo, *supra* note 2, at 497.

⁴ *Id.* at 499.

⁵ Richard I. Lanyon, *Theory and Treatment in Child Molestation*, 54 J. CONSULTING & CLINICAL PSYCHOL. 176, 177–78 (1986).

⁶ John Cloud, *Pedophilia*, TIME, Apr. 29, 2002, at 44–45; Bahroo, *supra* note 2, at 503 (“Pedophilia tends to be a chronic condition, and recidivism rates are high.”).

Over the years, a number of treatments have been implemented including physical castration and invasive, almost cruel therapies.⁷ The crime of raping a child is so abhorrent that many states adopted laws that called for the death penalty for the commission of such crimes.⁸ However, sentencing these offenders to death is no longer constitutional,⁹ and those who are released often reoffend and end up back in prison.¹⁰ In reality, there has never been a widely used treatment that could be described as both effective and humane in treating pedophiles and child molesters.¹¹ However, a certain treatment has been used and experimented with since the mid twentieth century that is both humane and incredibly effective in treating these offenders.¹² This treatment is the next step in sex offender therapy and the best possible option for any child molester or pedophile. This treatment is chemical castration.

While chemical castration sounds almost barbaric, it is one of the more civilized forms of treatment that has been used on sexual predators.¹³ It is merely a type of hormone therapy that takes away the offender's sexual desire.¹⁴ There can be some unpleasant side effects, but they are mostly reversible, and, overall, there is little pain and suffering associated with the procedure.¹⁵ Even the Catholic Church fully supports the procedure and set up the St. Luke Institute in the United States in 1985 where pedophilic priests undergo a combination of counseling and chemical castration for pedophilia.¹⁶ In addition to the effectiveness of chemical castration, it is also exponentially less expensive than the cost of keeping these individuals in prisons and hospitals, making it an almost perfect solution.¹⁷

While a number of constitutional concerns arise with the use of such a procedure, this comment proposes that chemical castration does not violate the Constitution. For those states

⁷ See *infra* Part II.C.

⁸ Joanna H. D'Avella, Note, *Death Row for Child Rape? Cruel and Unusual Punishment under the Roper-Atkins "Evolving Standards of Decency" Framework*, 92 CORNELL L. REV. 129, 131 n.7 (2006).

⁹ *Kennedy v. Louisiana*, 128 S. Ct. 2641, 2650–51 (2008).

¹⁰ Bahroo, *supra* note 2, at 503 ("Neither treatment nor incarceration seems to significantly affect recidivism.").

¹¹ See *infra* Part II.C.

¹² Charles L. Scott & Trent Holmberg, *Castration of Sex Offenders: Prisoners' Rights Versus Public Safety*, 31 J. AM. ACAD. PSYCHIATRY L. 502, 502 (2003).

¹³ Harrison, *supra* note 1, at 21–28. A name change may be in order as the term "castration" will always "summon up images of pain and suffering." *Id.* Chemical castration is not as invasive or "barbaric" as surgical castration. *Id.*

¹⁴ See *infra* Part III.A.

¹⁵ Harrison, *supra* note 1, at 21.

¹⁶ *Id.* at 23.

¹⁷ See *infra* note 144 and *infra* Part III.C.

that require an offender to undergo such a procedure, the health and safety of the offender is taken into consideration and a number of procedural safeguards protect individuals from abuse.¹⁸ Because of these procedural safeguards, it is difficult to think of the treatment as cruel and unusual in any way. Additionally, there is a fundamental right to procreate as well as a right to refuse treatment that must be taken into consideration.¹⁹ However, chemical castration is not in violation of either right as it does not necessarily strip an individual of reproductive capabilities.²⁰ Furthermore, keeping children safe from sexual predators is the absolute definition of a compelling state interest. Therefore, requiring chemical castration as a condition of parole or probation for pedophiles and child molesters is practical, effective, and constitutional.

A number of issues will be addressed throughout the text of this comment. The next section explains the true makeup of a child molester and how he is a different breed of sex offender. It will also discuss how it is necessary to do more than imprison such offenders, as rates of recidivism are incredibly high. Section III will examine different legislation that has been passed and treatments that have been implemented in dealing with these offenders, and why they have not solved the problem. Section IV will discuss chemical castration and how it is an effective treatment that can succeed in treating these offenders where others have failed. Lastly, section V will look at the constitutionality of requiring such offenders to undergo treatment.

I. THE PROBLEM

A. Getting to Know Your Friendly Neighborhood Child Molester

Child molesters and pedophiles are different from other sex offenders. A child molester is an older person “whose conscious sexual desires and responses are directed, at least in part, toward dependent, developmentally immature children and adolescents who do not fully comprehend these actions and are unable to give informed consent.”²¹ In fact, those who commit sexually deviant crimes against children, commonly referred to as pedophiles, are recognized as suffering from a disability.²²

¹⁸ See *infra* notes 196, 197, 198 and accompanying text.

¹⁹ See *infra* notes 218, 219, 220 and accompanying text.

²⁰ See *infra* notes 134, 135, 136, 137 and accompanying text.

²¹ Lanyon, *supra* note 5, at 176.

²² AMERICAN PSYCHIATRIC ASSOCIATION, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 571–72 (4th ed. 2000) (stating that the DSM-IV is a text commonly

The latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) characterizes a pedophile as one who has suffered from “recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity” with prepubescent children.²³ The individual has either “acted on these sexual urges” or such urges and fantasies cause incredible difficulty or distress in their life.²⁴ Pedophilia is one of the most common paraphilias listed in the DSM-IV.²⁵ There are both male and female pedophiles, however, the rates of female pedophilia are so incredibly low and there is so little research on the subject that many doubt their existence.²⁶ It should be noted that not every pedophile actually acts on their urges, nor does every person who commits such deviant sexual acts qualify as a pedophile under the DSM-IV.²⁷

There are two main groups of child sexual abusers.²⁸ Those sex offenders who are primarily sexually attracted to children are known as “fixated” or “preference” offenders and those who are normally attracted to adults, but who also attack children are known as “regressed” or “situational” offenders.²⁹ Situational offenders are more likely to blame their behavior toward children on drugs, alcohol, or other outside factors.³⁰ Their behavior is impulsive and each attack can usually be linked to significant life stressors.³¹ These offenders will commit sexual acts with children as if the child is a surrogate for an adult companion.³² There are also violent pedophiles that attack children due to anger, the need to feel power, violence, or just sadistic pleasure.³³

Some of these regressed or situational offenders realize that what they are doing is not right and hate their sexual proclivities toward children.³⁴ However, many offenders cannot stop

used in the United States by doctors, researchers, and even health insurance companies to diagnose, understand, and further study those inflicted with mental disorders).

²³ *Id.* at 572.

²⁴ *Id.*

²⁵ Ariel Rosler & Eliezer Witztum, *Pharmacotherapy of Paraphilias in the Next Millennium*, 18 BEHAV. SCI. & L. 43, 44 (2000) (stating that a paraphilia is a mental disorder where a person is aroused by abnormal sexual behaviors); Pedophilia, the sexual preoccupation with children, is a type of paraphilia. Bahroo, *supra* note 2, at 498.

²⁶ J. Paul Fedoroff et al., *A Case Series of Women Evaluated for Paraphiliac Sexual Disorders*, 8 CAN. J. HUM. SEXUALITY 127, 127–28 (1999).

²⁷ T. Howard Stone et al., *Sex Offenders, Sentencing Laws, and Pharmaceutical Treatment: A Prescription for Failure*, 18 BEHAV. SCI. & L. 83, 90–91 (2000).

²⁸ Lanyon, *supra* note 5, at 177–78. See also Harrison, *supra* note 1, at 26.

²⁹ Lanyon, *supra* note 5, at 177.

³⁰ Harrison, *supra* note 1, at 26.

³¹ Lanyon, *supra* note 5, at 177–78.

³² *Id.* at 177.

³³ Harrison, *supra* note 1, at 26.

³⁴ See Cloud, *supra* note 6, at 44 (quoting Reverend Stephen Rosetti who runs St. Luke Institute, a psychiatric hospital, who stated, “People don’t grow up and say, ‘I want

themselves and it is as if they are addicted to these children.³⁵ Alternatively, a number of offenders see nothing wrong with what they are doing.³⁶ Some think that the child wanted the experience, enjoyed it, and that it was educational for the child.³⁷ Fixated offenders often view children as being seductive and teasing the offender.³⁸ There are many theories as to why one develops a sexual attraction towards children, but no one knows exactly why this happens.³⁹ It has been attributed to a number of factors which include, but are not limited to: difficulties in forming intimate relationships, brain abnormalities, being sexually abused as a child, and even having a brother.⁴⁰ In the end it does not matter what category of child predator the offender falls into or why, but that every single one of them is incredibly dangerous.⁴¹

B. Recidivism Rates

Merely imprisoning child predators will not stop the attacks on children.⁴² The average prison sentence for a child sexual abuser is eleven years and, once released, there are incredibly high re-offense rates.⁴³ Prison is not much of a deterrent for sex offenders as many of these individuals “revert to earlier patterns of behavior” without further treatment.⁴⁴ Mere imprisonment may deter some offenders, but most are still going to offend regardless of what the penalty is.⁴⁵ Being that, in a number of cases, the abused child and the sexual predator are often the only ones that know of the offense, the risk of actually being caught is

to be a pedophile. . . . All the people I’ve ever talked to hate it.”); Lanyon, *supra* note 5, at 178.

³⁵ Harrison, *supra* note 1, at 26 (describing the idea that the offender is addicted to the child is raised by Dr. Gillian Mezey who claims that preferential offenders are “constantly preoccupied with children and will look on children as sexual prey, so when they see a child they will be considering ways in which they can gain access to that child”).

³⁶ Lanyon, *supra* note 5, at 178.

³⁷ Bahroo, *supra* note 2, at 500.

³⁸ *Id.*

³⁹ Cloud, *supra* note 6, at 45; Rosler & Witztum, *supra* note 2, at 416.

⁴⁰ Cloud, *supra* note 6. See also Lanyon, *supra* note 5, at 178–79.

⁴¹ Keith F. Durkin & Allison L. Digianantonio, *Recidivism Among Child Molesters: A Brief Overview*, 45 J. OFFENDER REHAB. 249, 250 (2008) (reviewing a number of case studies which have collectively found that children who are sexually victimized suffer in a number of different ways from the experience including physical injuries, fear, anxiety, depression, low self esteem, poor academic performance, social maladjustment, aggressive behavior, and acting out sexually).

⁴² See Bahroo, *supra* note 2, at 503.

⁴³ Cloud, *supra* note 6; Durkin & Digianantonio, *supra* note 41, at 252.

⁴⁴ Lita Furby et al., *Sex Offender Recidivism: A Review*, 105 PSYCHOL. BULL. 3, 3 (1989).

⁴⁵ Bahroo, *supra* note 2, at 503, 506.

very low.⁴⁶ Therefore, while the punishment for such an offense is often quite severe, it does not stop the offender from sexually victimizing children: many are driven by their urges and do not view getting caught as a reality.⁴⁷

Case studies conducted on released child sexual abusers have uncovered appalling results including that approximately 50–70% of those released eventually re-offend.⁴⁸ While some researchers argue that these rates are much lower, findings of low rates can be attributed to data collected from unreliable sources.⁴⁹ Being that the numbers in these studies are all based on self-reports and police or conviction records, a majority of attacks have gone unreported.⁵⁰ Studies have found that offenders commit two to five times more offenses than those for which they are arrested or convicted.⁵¹ Often, before an offender is released, assessments are done to determine if he presents a substantial risk of reoffending. However, so many factors are linked to sexual recidivism that it is impossible to predict who really presents a substantial risk, making life imprisonment the only way to ensure that the offender will not reoffend.⁵²

II. INADEQUATE SOLUTIONS

A. Sex Offender Registration

Since implementing a life sentence for every sex offender, or even just for every child molester and pedophile is unreasonable and unrealistic, states have taken steps to keep communities safe from released sex offenders. In 1947, California was the first state to have any kind of registry system for sex offenders, and it

⁴⁶ *Id.* at 503.

⁴⁷ *Id.*

⁴⁸ Durkin & Digianantonio, *supra* note 41, at 252.

⁴⁹ Joseph J. Romero & Linda Meyer Williams, *Recidivism Among Convicted Sex Offenders: A 10-Year Followup Study*, 49 FED. PROBATION 58, 58 (1985).

⁵⁰ *Id.*; Durkin & Digianantonio, *supra* note 41, at 251–52.

⁵¹ Durkin & Digianantonio, *supra* note 41, at 251 (declaring that most sex offenses are never reported and those arrested for such offenses reported committing at least twice as many offenses); Romero & Williams, *supra* note 49, at 58 (reporting that in one case study, it was found the incarcerated rapists and pedophiles had committed two to five times more offenses than they had been convicted for).

⁵² Robert Prentky et al., *Risk Factors Associated with Recidivism Among Extrafamilial Child Molesters*, 65 J. CONSULTING & CLINICAL PSYCHOL. 141, 147 (1997) (citing studies assessing factors that could help predict sexual recidivism and finding that the degree of sexual preoccupation with children, having other paraphilias, and numerous prior sexual offenses all indicated the individual was likely to reoffend); Durkin & Digianantonio, *supra* note 41, at 253 (reviewing a number of case studies which found that sexual recidivism was more likely when certain factors were present such as a higher number of offenses, the offender not being related to the victim, offending at an early age, going after very young children, using physical force, personality disorders, impulsivity, lack of empathy toward the victim, alcohol abuse, low IQ, and learning disabilities).

now has more registered sex offenders than any other state.⁵³ Today, there are currently over 11,000 registered sex offenders residing in Los Angeles County alone.⁵⁴ Throughout the late Twentieth Century, the laws regarding sex offender registration have evolved, in large part, due to the numerous attacks on children.⁵⁵ It was not until the 1990's, after the brutal rape and murder of seven-year-old Megan Kanka in New Jersey, that a public outcry was heard for something more to be done to protect the nation's children.⁵⁶ Unbeknownst to Megan or her family, their new neighbor across the street was a recently released sex offender.⁵⁷ When Megan went over to her neighbor's house to see a puppy, she ended up being raped and strangled to death with a leather belt.⁵⁸ No one in the community had any idea that they lived near a sexual predator.⁵⁹

After Megan's death, New Jersey passed an entirely new kind of sex offender registration law.⁶⁰ By the late 1990's, implementing sex offender registration laws was not anything new: over 25 states had already put such laws into action.⁶¹ However, there was a big difference between the laws enacted in New Jersey after Megan's death and those throughout the rest of the nation.⁶² New Jersey became the first state to require public notification in addition to the individual registering as a sex offender.⁶³ New Jersey's new law caught political attention as the Federal Omnibus Crime Bill was passing through Congress, which contained the Jacob Wetterling Crimes Against Children and Sex Offender Registration Act ("Jacob Wetterling Act").⁶⁴

President Clinton amended the Jacob Wetterling Act by signing Megan's Law so that, in addition to requiring every state to compile a registry for sex offenders, the states were required to

⁵³ Sex Offender Registration and Exclusion Information, <http://www.meganslaw.ca.gov/sexreg.aspx?lang=ENGLISH> (last visited Sept. 6, 2009).

⁵⁴ California Sex Offender Locator Map—Los Angeles, http://www.meganslaw.ca.gov/search_main.aspx?searchBy=county&county=los%20angeles&lang=ENGLISH (last visited Sept. 6, 2009).

⁵⁵ See Bahroo, *supra* note 2, and accompanying text.

⁵⁶ Maureen S. Hopbell, *Balancing the Protection of Children Against the Protection of Constitutional Rights: The Past, Present and Future of Megan's Law*, 42 DUQ. L. REV. 331, 336, 339 (2004).

⁵⁷ *Id.* at 332.

⁵⁸ *Id.* at 333; State v. Timmendequas, 737 A.2d 55, 68–69 (N.J. 1999).

⁵⁹ Hopbell, *supra* note 56, at 332.

⁶⁰ *Id.* at 336; N.J. STAT. ANN. § 2C:7–8 (West 2005).

⁶¹ Hopbell, *supra* note 56, at 337.

⁶² *Id.*

⁶³ *Id.*; N.J. STAT. ANN. § 2C:7–8 (West 2005).

⁶⁴ Hopbell, *supra* note 56, at 338 (reporting that years before Megan was murdered, Jacob Wetterling was kidnapped at gunpoint and neither he nor the man who abducted him were ever found). See generally 42 U.S.C. § 14071 (2006).

release relevant information to the public about such individuals residing in the state.⁶⁵ Compliance with the act was deemed so crucial that 10% of a state's federal funding was made contingent upon the timely passage of similar laws in each state.⁶⁶ Sex offenders required to register have challenged these laws as an invasion of privacy.⁶⁷ Given that the purpose of these laws is to protect the community and keep children safe from sexual predators, the states have held that such regulations are legitimate.⁶⁸ The U.S. Supreme Court even ruled that laws requiring sex offender registration and community notification are constitutional.⁶⁹ The Court held that the "mere injury to reputation" that these laws may cause, "even if defamatory, does not constitute the deprivation of a liberty interest."⁷⁰ While these laws are a step in the right direction, more needs to be done to ensure the safety of our children. Even with public notification systems in place, a child predator remains dangerous.

B. Capital Child Rape Statutes & Their Unconstitutionality

A number of states have realized that a community's awareness of convicted child molesters will not, by itself, ensure the safety of its children. In fact, starting in 1993, a number of states went a little too far and enacted statutes that made it a capital crime to rape a child under the age of twelve.⁷¹ Almost twenty years before these states began enacting capital child rape statutes, the U.S. Supreme Court held in *Coker v. Georgia* that issuing the death penalty for the rape of an adult woman is unconstitutional as it is a disproportionate punishment for a crime in which there is no loss of life.⁷² In *Kennedy v. Louisiana*, in 2008, the U.S. Supreme Court applied similar logic in finding that implementing the death penalty in cases of child rape was a grossly excessive punishment which violated the Cruel and Unusual Punishment Clause of the Eighth Amendment.⁷³ At the time the decision was handed down, Louisiana, South Carolina, Montana, Georgia, and Oklahoma all had laws which deemed the rape of a child to be a capital offense⁷⁴

⁶⁵ 42 U.S.C. § 14071 (2006).

⁶⁶ *Id.*; U.S. CONST. amend. VIII.

⁶⁷ Hopbell, *supra* note 56, at 342–43. *See also* Doe v. Poritz, 662 A.2d 367, 368, 372 (N.J. 1995).

⁶⁸ Hopbell, *supra* note 56, at 343.

⁶⁹ Conn. Dep't of Pub. Safety v. Doe, 123 S.Ct. 1160, 1162, 1164, (2003).

⁷⁰ *Id.* at 1164.

⁷¹ Kennedy v. Louisiana, 128 S. Ct. 2641, 2647–49 (2008).

⁷² Coker v. Georgia, 433 U.S. 584, 584–85 (1977).

⁷³ Kennedy, 128 S. Ct. at 2649–51.

⁷⁴ *Id.*

Arguably, these states found the rape of a child to be such a deplorable act that even though the offender had not taken a life, he deserved to have his life taken for the commission of such an act. In fact, the Governor of Louisiana, Bobby Jindal, stated in an interview that he found the Court's decision in *Kennedy* to be an "awful decision."⁷⁵ In reaction to the Court's decision, Governor Jindal stated: "That's ridiculous . . . We don't want anybody in Louisiana harming our children. We think these monsters need to be stopped . . . If there is any other crime other than taking human life . . . that screams out for the death penalty it's those criminals that harm our children."⁷⁶

While the state requested a rehearing on the case due to certain factual errors, the Court denied the request in October 2008.⁷⁷ The Supreme Court's decision in *Kennedy v. Louisiana* still stands, so implementing the death penalty for the rape of a child constitutes cruel and unusual punishment in violation of the Eighth Amendment.⁷⁸ Society's abhorrence to crimes of this nature is obvious. However, the dilemma as to what to do with those who commit sexual crimes against children remains. Imprisonment keeps pedophiles away from children. However, those released back into the community often end up reoffending. The death penalty is far too harsh a punishment, yet more needs to be done to protect our children. Punishing offenders simply is not enough; it is necessary to take all steps possible to rehabilitate them.⁷⁹

C. Treatment, Surgery & Disappointment

In reality, "active pedophiles who find their way into the few treatment programs around the country turn out to be less of a risk than those who are locked up for a while and released."⁸⁰ Scientists in the 1940's realized that "[p]unishment alone is a

⁷⁵ *Your World With Neil Cavuto* (Fox News television broadcast June 26, 2008) (interview with the Governor of Louisiana, Bobby Jindal) available at <http://www.foxnews.com/search-results/m/20265419/dead-wrong.htm?q=jindal> (last visited Oct. 26, 2009).

⁷⁶ *Id.*

⁷⁷ Linda Greenhouse, *In Court Ruling On Executions, A Factual Flaw*, N.Y. TIMES, July 2, 2008, at A1 (reporting the Supreme Court did not take into account that in 2006 the Uniform Code of Military Justice had been revised to deem child rape as a capital offense under federal law); *U.S. Supreme Court Denies Rehearing in Kennedy v. Louisiana Opinion*, DEATH PENALTY INFORMATION CENTER, <http://www.deathpenaltyinfo.org/us-supreme-court-denies-rehearing-kennedy-v-louisiana-opinion> (last visited Sept. 6, 2009); *Kennedy v. Louisiana*, 129 S. Ct. 1 (2008), *rehearing denied* (Oct. 1, 2008) (denying based on the Uniform Code of Military Justice not conflicting with the Court's reasoning or decision).

⁷⁸ *Kennedy v. Louisiana*, 128 S. Ct. 2641, 2641 (2008).

⁷⁹ Bahroo, *supra* note 2, at 506.

⁸⁰ Cloud, *supra* note 6.

wasteful and useless procedure” for a child sexual abuser.⁸¹ Time spent in prison does nothing to treat the offender’s sexual proclivities toward children.⁸² These offenders need serious help in controlling their urges and leaving children alone.⁸³

A number of treatments have been used on pedophiles and child molesters, some of which are bizarre and even seem absurd. Aversion therapy, also known as covert association or covert sensitization, is a common cognitive behavioral treatment that has been used to treat child molesters.⁸⁴ Here, the deviant sexual behavior or fantasy is associated with unpleasant stimuli so that the offender will associate the deviant fantasies with the unpleasant stimuli and no longer find them arousing.⁸⁵ Even though aversion therapy is promising for some offenders, the effects are not long lasting.⁸⁶ The unpleasant associations “lose their pervasive effectiveness quite quickly” and need to be strengthened continuously, which is difficult to do.⁸⁷

Other treatments include directed masturbation and satiation procedures.⁸⁸ Directed masturbation is where an individual masturbates to pictures of adults or to adult sexual fantasies in an attempt to associate the pleasure of having an orgasm with these fantasies and images.⁸⁹ The hope is to increase the offender’s ability to become sexually aroused by adults.⁹⁰ However, this treatment is useless against the situational offender who is already attracted to adults, but still goes after children. As for the satiation procedures, there are two different types: verbal and masturbatory. Verbal satiation is where the offender is directed to say deviant fantasies out loud until it becomes boring and tedious.⁹¹ Masturbatory satiation is a procedure in which the offender is directed to masturbate to deviant sexual fantasies for prolonged periods of time without orgasm.⁹² The idea is that the uncomfortable experience will be

⁸¹ Benjamin Apfelberg et al., *A Psychiatric Study of 250 Sex Offenders*, 100 AM. J. PSYCHIATRY 762, 762 (1944).

⁸² Bahroo, *supra* note 2, at 506.

⁸³ *Id.* at 503 (“Neither treatment nor incarceration seems to significantly affect recidivism.”).

⁸⁴ *Id.* at 504; W.L. Marshall, *Covert Association: A Case Demonstration with a Child Molester*, 6 CLINICAL CASE STUD. 218, 218 (June 2007).

⁸⁵ Bahroo, *supra* note 2, at 504 (stating that the unpleasant stimuli can be anything such as noxious odors and electric shocks); Marshall, *supra* note 84, at 218.

⁸⁶ Marshall, *supra* note 84, at 219.

⁸⁷ *Id.*

⁸⁸ Bahroo, *supra* note 2, at 504; N. McConaghy, *Science and the Mismanagement of Rapists and Pedophiles*, 4 PSYCHIATRY, PSYCHOL. & L. 109, 114 (1997).

⁸⁹ McConaghy, *supra* note 88, at 114.

⁹⁰ *Id.*

⁹¹ Bahroo, *supra* note 2, at 504.

⁹² *Id.*; McConaghy, *supra* note 88, at 114.

associated with the “deviant sexual interest” thereby diminishing it.⁹³ While these procedures are undoubtedly unpleasant, they appear to do little to lessen the offender’s attraction to children.⁹⁴

Other treatments including anger management, conflict resolution, social skills training, and victim empathy have also been implemented.⁹⁵ One case study even used music therapy and drum playing to treat a child molester.⁹⁶ The proponents of music therapy say that it allows the offender to express himself, develop listening skills, learn self control, and reduce the offender’s resistance to therapy and level of denial.⁹⁷ However, while the case study showed that he became more open and learned to play the drums, nothing actually indicated that the offender was less likely to sexually victimize children.⁹⁸

Unfortunately, psychological and behavioral treatments alone are not effective in treating *all* pedophiles and child predators.⁹⁹ First of all, a number of released pedophiles will not even seek out treatment unless their urges cause them great distress or a court has forced them to undergo treatment.¹⁰⁰ Others who want help may be afraid to talk to anyone about what they have done.¹⁰¹ Pedophiles and sexual predators who have attacked large numbers of children usually have not disclosed all of their offenses.¹⁰² They have been trained to deny their offenses, being that in the legal system, further disclosure of any unknown offenses will only lead to lengthier prison sentences.¹⁰³ With society’s abhorrence to sexual crimes against children, it is not uncommon for many such offenders to be afraid of telling anyone about their offenses, including therapists, for fear of incarceration.¹⁰⁴ Also, a number of those who do undergo such psycho-behavioral treatment do not actually receive any

⁹³ McConaghy, *supra* note 88, at 114.

⁹⁴ *Id.* at 118.

⁹⁵ Bahroo, *supra* note 2, at 504.

⁹⁶ See generally Vaughn A. Kaser, *Music Therapy Treatment of Pedophilia Using the Drum Set*, 18 ARTS IN PSYCHOTHERAPY 7 (1991).

⁹⁷ See *id.*

⁹⁸ *Id.* at 15. The author states in the conclusion that this method should be implemented to treat pedophilia and related personality disorders, but the offender’s changes in playing showed nothing more than an improvement in playing the drums, social skills, a possible change in mood swings, and possibly becoming more comfortable with the researcher or therapist in general. *Id.*

⁹⁹ Stone et al., *supra* note 27, at 94–95.

¹⁰⁰ Bahroo, *supra* note 2, at 503.

¹⁰¹ Jon Kear-Colwell & Douglas P. Boer, *The Treatment of Pedophiles: Clinical Experience and the Implications of Recent Research*, 44 INT’L J. OFFENDER THERAPY & COMP. CRIMINOLOGY 593, 594 (2000).

¹⁰² *Id.*

¹⁰³ *Id.*

¹⁰⁴ *Id.*

benefit from it.¹⁰⁵ Research has revealed that while some offenders are effectively treated with just psychotherapy and behavioral reassessment therapies, this same treatment is completely ineffective on other sexual offenders.¹⁰⁶ These offenders need more than just a few therapy sessions to prevent them from reoffending.

On the more drastic side, physical castration and neurosurgery are two options that have also been considered for the treatment of child molesters.¹⁰⁷ Neurosurgery “involves removal of part of the hypothalamus to . . . decrease sexual arousal and impulsive behaviors.”¹⁰⁸ The procedure was more common in Europe than it ever was in the United States. However, it was rarely ever implemented and, in reality, has not been used in decades.¹⁰⁹ Additionally, “this procedure has a high failure rate and adverse [consequences].”¹¹⁰ Plus, mandating that part of the offender’s brain be removed crosses ethical and moral lines.¹¹¹ Therefore, because brain surgery is ineffective, highly invasive, and unethical, it cannot be considered as a serious treatment for pedophiles and other child predators.

Physical castration is a drastic measure, although it is nowhere near as dangerous, ethically unsound, or constitutionally questionable as neurosurgery. Though physical castration has its advantages, a number of flaws remain. Physical castration, or orchiectomy, is a surgical procedure in which a man’s testes are removed.¹¹² This causes a dramatic drop in the levels of testosterone in the body which severely decreases a man’s sex drive.¹¹³ Studies have shown that the offender will feel calmer, happier, more passive, and less likely to reoffend.¹¹⁴ The offender is better able to suppress violent and aggressive urges making it easier for him to live in society.¹¹⁵ Additionally, while the sexual recidivism rate for non-castrated

¹⁰⁵ Stone et al., *supra* note 27, at 95.

¹⁰⁶ *Id.*

¹⁰⁷ Bahroo, *supra* note 2, at 505.

¹⁰⁸ *Id.*

¹⁰⁹ See Luk Gijs & Louis Gooren, *Hormonal and Psychopharmacological Interventions in the Treatment of Paraphilias: an Update*, 33 THE J. SEX RES. 273, 274 (Fall 1996) (going over history of and current trends of medical treatment for paraphilias).

¹¹⁰ Bahroo, *supra* note 2, at 505.

¹¹¹ *Id.*

¹¹² Stone et al., *supra* note 27, at 92.

¹¹³ *Id.* at 92–93. See also Robert A. Prentky, *Arousal Reduction in Sexual Offenders: A Review of Antiandrogen Interventions*, 9 SEXUAL ABUSE: J. RES. & TREATMENT 335, 335 (1997) (reporting that the testes produce 95% of testosterone in the body).

¹¹⁴ Harrison, *supra* note 1, at 18.

¹¹⁵ *Id.*

offenders is around 50%, the re-offense rate for a sex crime among castrated offenders is usually 1–5%.¹¹⁶

While physical castration proves to be highly effective, there are a number of ethical and constitutional concerns associated with it. There is no problem if a patient voluntarily undergoes such treatment completely out of their own free will, which some patients have done in order to stop their deviant sexual urges and live normal lives.¹¹⁷ However, there are a number of problems when physical castration is statutorily mandated for those who commit such offenses. First, it is a permanent procedure.¹¹⁸ While certain effects can be reversed by injections of testosterone, nothing can replace what was removed through this invasive surgical procedure. Physical castration is literally a life sentence.¹¹⁹ Therefore, just like executing an innocent individual who was wrongly convicted, nothing can be done to remedy a wrongly convicted sex offender who was forced to undergo surgical castration. In addition, there are a number of adverse mental and physical side effects including metabolic changes, depression, and suicidal tendencies.¹²⁰ While certain states allow for sex offenders to voluntarily undergo physical

¹¹⁶ Prentky, *supra* note 113, at 336 (reporting results from a number of studies conducted in Europe throughout the 20th century which assessed groups of sex offenders and found sexual recidivism rates before castration were 50% in a group of 738 sex offenders and 1.4% after castration, group of 1,036 sex offenders with an 84% recidivism rate before castration and 2.3% after, 30 year follow up study of 900 people had a post castration rate of 1.1%, group of 127 offenders after 5 years who had an offense rate pre-castration of 76.8% and a post castration recidivism rate of 7.4%); Stone et al., *supra* note 27, at 93 (reporting that studies of recidivism rates for sexually castrated offenders found pre surgical and non castration rates of sexual recidivism were 46% and 58% and rates for those castrated were only 3% and 2.9%); Rosler & Witztum, *supra* note 25, at 44 (claiming that “post castration recidivism rates are among the lowest rates for all forms of treatment of paraphilias” and find that in 11 European case studies that 3,589 sex offenders who were castrated had a mean sexual recidivism rate of 2.2% with findings ranging from 0–7.4%); Harrison, *supra* note 1, at 18–19 (reporting on numerous case studies which found that from 1930–1969 over 400 castrated sex offenders had only a 3% rate of sexually reoffending, group of 224 in 1960 were castrated with a 3.5% sexual recidivism rate, in 1968 a study found that of 18 castrated sex offenders there was no recidivism, in 1989 a comparison study of 99 castrated and 35 non-castrated sex offenders found that 46% of non castrated offenders had sexually reoffended as opposed to only 3% of castrated offenders).

¹¹⁷ See Harrison, *supra* note 1, at 19. (discussing a sex offender in England who wanted to be castrated to offset reoffending. The prison would not allow it, but after going on a 42 day hunger strike and trying to do the operation himself, he was allowed to get the operation which he paid for out of his own pocket. He claims that he now has no sexual urges and feels that the procedure has stopped him from sexually abusing at least four or five children).

¹¹⁸ *Id.* at 19.

¹¹⁹ *Id.*; see *infra* note 229.

¹²⁰ Stone et al., *supra* note 27, at 93 (pointing out that side effects include changes in metabolic processes, loss of protein, changed pituitary function, weight fluctuation, loss of calcium in bones, hot flashes and perspiration, loss of body hair, depression, suicidal tendencies, and indifference to life).

castration to avoid longer sentences or chemical castration, no state *requires* the offender to undergo physical castration.¹²¹ Forcing someone to undergo a permanent and life altering surgical procedure challenges what is ethical and constitutional.¹²²

III. CHEMICAL CASTRATION: A NEW HOPE

An almost perfect solution to keep child molesters from reoffending would be a treatment with the effectiveness of surgical castration, but without the ethical dilemmas. This solution is chemical castration, which a number of states have already implemented.¹²³ Many people find any type of castration an unsettling punishment due to its improper uses throughout history.¹²⁴ However, chemical castration is one of the most effective and least restrictive ways to help treat child predators and keep children safe.

A. The Why & How of Chemical Castration

Chemical castration has been experimented with ever since the 1940's with the use of anti-androgens to lower the testosterone levels of sexually dangerous males.¹²⁵ About two decades later Dr. John Money became the first person in the United States to administer the anti-androgen drug Medroxyprogesterone Acetate (MPA) to sex offenders.¹²⁶ While other chemical agents have been administered to such offenders throughout the world, MPA is the most commonly administered drug in the U.S. for the purposes of chemical castration.¹²⁷ In fact, the MPA is administered through injections of the drug Depo Provera, which is used by many women as a hormonal

¹²¹ See *infra* notes 159, 160.

¹²² U.S. CONST. amend. VIII.

¹²³ See *infra* Part IV.B.

¹²⁴ See *Buck v. Bell*, 274 U.S. 200 (1927) (upholding laws that allowed for the sterilization of those declared to be imbeciles and mentally defective as it was thought to promote the general welfare of society. The offspring of these individuals would only fall into lives of crime and so this law prevents the continuous breeding of imbeciles); Mary Ann Farkas & Amy Stichman, *Sex Offender Laws: Can Treatment, Punishment, Incapacitation, and Public Safety Be Reconciled?*, 27 CRIM. JUST. REV. 256, 267 (2002) (describing that in Colonial times, black slaves could be castrated for merely having sex with white women. Also, in the eugenics movement of the Nineteenth Century, castration and sterilization was permitted in America for criminals and the mentally ill).

¹²⁵ Scott & Holmberg, *supra* note 12, at 502.

¹²⁶ *Id.*

¹²⁷ Prentky, *supra* note 113, at 341–42, 338 (cyproterone is commonly is used in Great Britain and Canada, but MPA is the “agent of choice” in the United States. Serotonin re-uptake inhibitors, LHRH agonists, and gonadotropin releasing hormone is also used for chemical castration purposes). See also Gijs & Gooren, *supra* note 109, at 274.

contraceptive.¹²⁸ The reason it has such different effects on men that undergo such treatment is that women only receive 150 mg of the drug every 3 months.¹²⁹ Sex offenders receive the drug every week, usually in higher doses.¹³⁰

MPA is able to reduce the occurrence of sexual imagery and lessen the offender's level of sexual desire.¹³¹ MPA is administered through weekly intramuscular injections of 100 to 1,000 milligrams of the drug, depending on the needs of the offender.¹³² The MPA inhibits the release of the follicle stimulating hormone and the lutenizing hormone from the brain's anterior pituitary gland.¹³³ Essentially, the drug causes the brain to believe the body has enough testosterone and so it does not allow the testicles to produce anymore.¹³⁴ The effect is a reduction in the levels of testosterone in the offender's blood down to that of a pre-pubescent male within one to two weeks; this low level of testosterone greatly lowers the offender's sex drive.¹³⁵ The offender is then temporarily impotent and, when the drug is in full effect, there is a reduction in orgasm, sperm production, sexual frustration, and the frequency and satisfaction of masturbation.¹³⁶ One of the most appealing aspects of chemical castration is that the offender is made calmer and more responsive to psycho-behavioral treatment.¹³⁷ The injections suppress the offender's sexual urge and desire as well as make it easier for the patient to concentrate on therapy, control his behavior, and prevent relapse.¹³⁸

The drug has the incredible effect of keeping the pedophile or child molester from offending, but scientists disagree as to how long the drug needs to be administered.¹³⁹ Some researchers have stated that the hormone therapy should be for a few months, others for up to five years, and some researchers argue

¹²⁸ *Depo-Provera*, DEPOPROVERA, http://media.pfizer.com/files/products/uspi_depo_provera_contraceptive.pdf (last visited Nov. 7, 2009).

¹²⁹ *Id.*; Carol Levine, *Depo-Provera and Contraceptive Risk: A Case Study of Values in Conflict*, 9 HASTINGS CENTER REP. 8, 9 (1979).

¹³⁰ Gijs & Gooren, *supra* note 109, at 275.

¹³¹ Peter J. Gimino, III, *Mandatory Chemical Castration for Perpetrators of Sex Offenses Against Children: Following California's Lead*, 25 PEPP. L. REV. 67, 74-75 (1997).

¹³² Gijs & Gooren, *supra* note 109, at 275 (stating that the dosage varies with the most common dosage being between 300 and 500 milligrams). *See also* Rosler & Witztum, *supra* note 25, at 47.

¹³³ Gimino, *supra* note 131, at 74.

¹³⁴ Harrison, *supra* note 1, at 20; Gimino, *supra* note 131, at 74.

¹³⁵ Harrison, *supra* note 1, at 20; Rosler & Witztum, *supra* note 25, at 47; Gimino, *supra* note 131, at 74.

¹³⁶ Harrison, *supra* note 1, at 20.

¹³⁷ *Id.*

¹³⁸ *Id.*

¹³⁹ Gijs & Gooren, *supra* note 109, at 275.

that the taking of these medications should never stop.¹⁴⁰ The main argument for continuing the treatment is the fact that the effects of the drugs, especially MPA, are reversible once weekly injections stop.¹⁴¹ It is said that in order to properly treat the offender, they must undergo behavioral therapy *as well as* the hormonal treatments in order to treat the problem as a whole.¹⁴² The idea is that proper treatment will treat the person and not just try to get rid of the offending behavior.¹⁴³ Either way, there is no doubt that the administration of these drugs is effective and has kept numerous offenders from sexually victimizing children.¹⁴⁴ As stated by one released offender undergoing hormone therapy treatments, “I realised [sic] that I could walk down the street, see boys I found sexually attractive, and not be possessed by thoughts about having sex with them . . . It took that edge off.”¹⁴⁵

B. Recent Legislation

Recognizing the overall societal benefit, a number of states have already enacted chemical castration statutes for certain sex offenders. California was the first state to implement a chemical castration statute in 1996 when it made the procedure a condition of probation for certain offenders.¹⁴⁶ Today, a number of states have followed California’s example including

¹⁴⁰ *Id.*

¹⁴¹ Harrison, *supra* note 1, at 21 (explaining that the depo provera will remain in the bloodstream for up to eight weeks, but the effects will “significantly fall within days”); Rosler & Witztum, *supra* note 25, at 46 (The effects are “dose dependant”); Farkas & Stichman, *supra* note 124, at 267 (stating that studies have shown that the effect occurred as long as the offender continued with the injections).

¹⁴² Harrison, *supra* note 1, at 20; Kear-Colwell & Boer, *supra* note 101, at 603.

¹⁴³ Kear-Colwell & Boer, *supra* note 101, at 603.

¹⁴⁴ Cloud, *supra* note 6 (citing a 1991 study of 400 people who underwent physical castration showed that only 1.2% had reoffended. In a study done by the St. Luke Institute following 121 pedophilic priests who underwent the treatment, only 3 relapsed); Stone et al., *supra* note 27, at 96–97 (stating that the treatment affects “both deviant and non deviant sexual behavior” as there is a general suppression of sexual drive, urges, and fantasies. There are statistically significant reductions in the frequency of paraphilic behaviors with the use of MPA with later studies confirming these benefits); Rosler & Witztum, *supra* note 25, at 45, 47 (reporting that chemical castration is a highly effective treatment for sex offenders since “MPA is effective in controlling [psycho sexual disorders], particularly pedophilia”); Prentky, *supra* note 113, at 338 (arguing that even in the very early studies of MPA used to treat sex offenders done in the 1960’s, MPA reduced a number of criminal sexual behaviors. The drug was found to have a strong effect on sexual behavior, sexual desire, and helped the offenders control their behavior, especially when implemented with therapy); Harrison, *supra* note 1, at 24. (citing that in a 1994 study, in a group of 629 men who underwent chemical castration there was only an 8% recidivism rate after 5 years. Research in the area of chemical castration is incredibly optimistic and some feel that the treatment will work for most sex offenders, especially when the individual taking the drug is also going through behavioral therapy).

¹⁴⁵ Harrison, *supra* note 1, at 25.

¹⁴⁶ CAL. PENAL CODE § 645 (Deering 2008); Gimino, *supra* note 131, at 70, 79.

Montana,¹⁴⁷ Iowa,¹⁴⁸ Wisconsin,¹⁴⁹ Louisiana,¹⁵⁰ Oregon,¹⁵¹ and Florida.¹⁵² In fact a number of states have gone even further than California in their adoption of such statutes. California requires offenders to undergo chemical castration if they are convicted of multiple offenses against a victim under the age of thirteen.¹⁵³ Louisiana, Wisconsin and Iowa also require that the offenses must be against a child for MPA injections to be implemented,¹⁵⁴ but Florida,¹⁵⁵ Oregon,¹⁵⁶ and Montana have no age requirement for the victim.¹⁵⁷ While some states leave it up to the court's discretion, California, Iowa, and Florida all make MPA injections a mandatory condition of release for those with multiple convictions.¹⁵⁸ In addition, Iowa, Louisiana, California, and Florida allow offenders to avoid MPA treatments if the offender voluntarily undergoes physical castration.¹⁵⁹ Plus, Texas *only* gives the offender the decision to undergo voluntary surgical castration as a condition of release and does not offer chemical castration as an option.¹⁶⁰ These states have all realized that more than just imprisonment is necessary to keep a sex offender from re-offending, and most of them realize that chemical castration is the most proper means to achieve that end.

C. Chemical Castration vs. Incarceration

Another reason why chemical castration is so appealing is that it allows the offender to be out of prison and comes at a cost that is exponentially lower than incarceration. In 1985, state governments spent \$9 billion on corrections.¹⁶¹ In 1993, state, local, and federal governments combined spent almost \$100 billion on the justice system and, in 1996, states spent about \$27

¹⁴⁷ MONT. CODE ANN. § 45-5-512 (2007).

¹⁴⁸ IOWA CODE ANN. § 903B.1 (West 2003).

¹⁴⁹ WIS. STAT. ANN. § 302.11 (West 2005).

¹⁵⁰ LA. REV. STAT. ANN. § 15:538 (West 2005).

¹⁵¹ OR. REV. STAT. ANN. § 144.625 (West 2005).

¹⁵² FLA. STAT. ANN. § 794.0235 (West 2007).

¹⁵³ CAL. PENAL CODE § 645 (Deering 2008).

¹⁵⁴ LA. REV. STAT. ANN. § 15:538 (2008) (offenses against a "minor" under 12 years of age); WIS. STAT. ANN. § 302.11 (West 2005) (required for "serious child sex offenders"); IOWA CODE ANN. § 903B.10 (West 2008).

¹⁵⁵ FLA. STAT. ANN. §§ 794.0235, 794.011 (West 2007).

¹⁵⁶ OR. REV. STAT. ANN. § 144.625 (West 2008).

¹⁵⁷ MONT. CODE ANN. § 45-5-512 (2007); MONT. CODE ANN. § 45-5-502 (2007); MONT. CODE ANN. § 45-5-503 (2007); MONT. CODE ANN. § 45-5-507 (2007).

¹⁵⁸ CAL. PENAL CODE § 645 (Deering 2008); FLA. STAT. ANN. § 794.0235 (West 2007); IOWA CODE ANN. § 903B.10 (West 2003).

¹⁵⁹ CAL. PENAL CODE § 645 (Deering 2008); LA. REV. STAT. ANN. § 15:538 (West 2005); FLA. STAT. ANN. § 794.0235 (West 2007); IOWA CODE ANN. § 903B.10 (West 2003).

¹⁶⁰ TEX. GOV'T CODE ANN. § 501.061 (Vernon 2004).

¹⁶¹ *Developments in the Law: Alternatives to Incarceration*, 111 HARV. L. REV. 1863, 1892 (1998).

billion on corrections.¹⁶² The average cost of housing and caring for an inmate is approximately \$20,000 per inmate per year.¹⁶³ Furthermore, the actual amount spent on incarceration is much higher being that costs for projects such as building new prisons are not included into agency budgets.¹⁶⁴ With these costs, keep in mind that between 1980 and 1994 alone, the population of imprisoned sex offenders grew approximately 330%.¹⁶⁵ Of these offenders, two-thirds sexually victimized minors and more than half of that population went after children under twelve years of age.¹⁶⁶

Even when such offenders are released from prison, many of them are involuntarily committed to hospitals and continue to cost the state money.¹⁶⁷ Ever since the 1930's there have been laws pertaining to "sexual psychopaths" or "sexually violent predators" that require that such offenders are institutionalized after they have served their time in prison or after being found not competent to stand trial.¹⁶⁸ Today, when an offender is placed in one of these institutions they can regularly petition for release.¹⁶⁹ Nevertheless, the offender is usually held until both a psychologist and the court agree that the individual's "personality disorder" or "brain abnormality" has "changed" so that he is "deemed safe" for release.¹⁷⁰ There are currently seventeen states that fund the involuntary commitment of such individuals.¹⁷¹ The cost of the programs varies between the states and the yearly cost ranges from \$500,000–\$45 million.¹⁷²

As of 2002, weekly injections of MPA cost \$160 per month.¹⁷³ While this is the cost of the drug alone and does not include additional costs involved with probation or further therapy, the difference in cost remains phenomenal as the average cost of one inmate's incarceration is over \$1,660 per month.¹⁷⁴ This difference is magnified when coupled with the fact that incarceration can be fifteen times more expensive per prisoner

¹⁶² *Id.*

¹⁶³ *Id.* at 1893.

¹⁶⁴ *Id.* at 1893–94.

¹⁶⁵ Stone et al., *supra* note 27, at 87.

¹⁶⁶ *Id.*

¹⁶⁷ See Farkas & Stichman, *supra* note 124, at 258–59.

¹⁶⁸ *Id.*

¹⁶⁹ *Id.* at 259.

¹⁷⁰ *Id.* at 259–60.

¹⁷¹ WASH. ST. INST. FOR PUB. POL'Y, INVOLUNTARY COMMITMENT OF SEXUALLY VIOLENT PREDATORS: COMPARING STATE LAWS, Doc. No. 05-03-1101, at 1 (2005).

¹⁷² *Id.* at 2, 6.

¹⁷³ Harrison, *supra* note 1, at 20.

¹⁷⁴ *Id.*; see *supra* notes 163, 173 and accompanying text.

than for those on parole or probation.¹⁷⁵ In addition, Iowa and Louisiana require the offender to pay for his own treatment once released to lower costs even further.¹⁷⁶ Implementing chemical castration as a condition of release could allow prisoners to get out of prison sooner, help them control their urges, get jobs, and reduce prison crowding and spending. In fact, it has even been stated that “0.5 to 0.7% of gross domestic product is lost by incarcerating felons rather than placing them in jobs.”¹⁷⁷ Even the American Bar Association endorses probation over imprisonment, as cited in *State v. Christopher*, finding it to be:

a desirable disposition . . . [as it] maximizes the liberty of the individual while vindicating the authority of the court; it eases the reintegration of the offender into the community; it minimizes the hidden costs that imprisonment places on the family of the offender; and is it the most economic form of correctional supervision.¹⁷⁸

The *Christopher* court agreed that a good reason for granting probation was to “alleviate the overcrowding in our prisons by not incarcerating those people with whom the state can adequately deal in other ways.”¹⁷⁹ Requiring chemical castration for released child molesters and pedophiles is the most logical choice in to drive down costs and recidivism rates as well as help get some of these people out of prison to become fully functioning members of society.

IV. THE CONSTITUTION & WHY CHEMICAL CASTRATION IS PERFECTLY ACCEPTABLE

The United States Supreme Court has never directly assessed the constitutionality of chemical castration. While some find the administration of MPA to sex offenders to be a questionable procedure, chemical castration is not harsh enough to be classified as cruel and unusual punishment. Because it is so effective, the effects are reversible, and it deals with the safety of children—an unquestionably compelling state interest—this procedure is not in violation of the Constitution.¹⁸⁰

¹⁷⁵ *Developments in the Law*, *supra* note 161, at 1893.

¹⁷⁶ IOWA CODE ANN. § 903B.10(6) (West 2003); LA. REV. STAT. ANN. § 15:538(C)(5) (West 2005).

¹⁷⁷ *Developments in the Law*, *supra* note 161, at 1894.

¹⁷⁸ *State v. Christopher*, 652 P.2d 1031, 1033 (Ariz. 1982) (citing the ABA’s Standards for Criminal Justice).

¹⁷⁹ *Id.*

¹⁸⁰ *See supra* notes 2, 3.

A. Chemical Castration Is Not Cruel & Unusual Punishment

According to the Eighth Amendment of the United States Constitution, the federal government prohibits any punishment that can be categorized as “cruel and unusual”.¹⁸¹ In 1972, in *Furman v. Georgia*, the U.S. Supreme Court held that excessive punishments which are “greatly disproportioned to the offenses charged” constitute cruel and unusual punishment in violation of the Eighth Amendment.¹⁸² Therefore, any punishment that is grossly excessive to the crime charged is prohibited.¹⁸³

The only court in history that found chemical castration to be a disproportionate sentence was the Michigan Supreme Court in *People v. Gauntlett* in 1984.¹⁸⁴ In *Gauntlett*, MPA was going to be administered by injections of Depo-Provera to a man who was convicted of raping his fourteen-year old stepdaughter and molesting his twelve year old stepson.¹⁸⁵ The court believed that administering these injections was an “unlawful condition of probation.”¹⁸⁶ This is because the Food and Drug Administration (FDA) had not approved the drug when the decision was handed down in 1984.¹⁸⁷ At that time it “had not [yet] gained acceptance in the medical community as a safe and reliable medical procedure.”¹⁸⁸ Today, however, this argument has no merit because the drug was deemed safe and approved by the FDA in October of 1992.¹⁸⁹

Even though the drug is safe, there are still some unpleasant side effects of chemical castration.¹⁹⁰ These include, but are not limited to, loss in bone density, weight gain, loss of body hair, depression, and fatigue.¹⁹¹ The majority of these side effects are reversible and the most severe ones are rarely experienced.¹⁹²

¹⁸¹ U.S. CONST. amend. VIII.

¹⁸² *Furman v. Georgia*, 408 U.S. 238, 279–80 (1972) (Douglas, J., concurring).

¹⁸³ *See id.*

¹⁸⁴ *People v. Gauntlett*, 352 N.W.2d 310, 318 (Mich. Ct. App. 1984).

¹⁸⁵ *Id.* at 311, 313.

¹⁸⁶ *Id.* at 316–17.

¹⁸⁷ *Id.* at 315.

¹⁸⁸ *Id.* at 316.

¹⁸⁹ *U.S. Food and Drug Administration: All Approvals October 1992*, Fda.Gov, <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Reports.MonthlyApprovalsAll> (last visited September 14, 2009).

¹⁹⁰ Rosler & Witztum, *supra* note 25, at 47 (declaring that side effects include “weight gain, malaise, nightmares, headaches, muscular cramps, dyspepsia, gallstones, diabetes mellitus,” and a decrease in mineral bone density); Harrison, *supra* note 1, at 21 (reporting that side effects also include depression, a decrease in body hair, hypoglycemia, insomnia, and difficulty breathing).

¹⁹¹ *Id.*

¹⁹² Harrison, *supra*, note 1, at 21; Gijs & Gooren, *supra* note 109, at 276 (explaining that when the treatment was stopped, “the adverse side effects with the exception of diabetes, disappeared”).

This is one of the reasons that the dosage of MPA given to the offender is often kept as low as possible.¹⁹³ While it is necessary to put an end to deviant sexual behavior, the offender's safety is not disregarded.¹⁹⁴ Even though some offenders may receive the injections for the rest of their lives, the dosage may be lowered over time if deemed necessary or if the offender shows improvement.¹⁹⁵ Courts take these side effects into account in ordering individuals to undergo such chemical therapy.¹⁹⁶ If the courts simply forced individuals to undergo or continue treatment that would result in serious harm, then the procedure could be seen as cruel and unusual punishment.¹⁹⁷

In addition to protecting the individual's health, states with mandatory chemical castration statutes have procedural safeguards in place to protect the offenders from abuse.¹⁹⁸ Florida, for example, does not allow the mandatory administration of MPA injections to be carried out unless the procedural requirements of the state's chemical castration statute have been followed exactly.¹⁹⁹ If the procedure has not been strictly followed, even slightly, the court will not uphold that portion of the offender's sentence.²⁰⁰ The courts realize that chemical castration is a serious matter and do not nonchalantly impose such treatment on individuals.

¹⁹³ Gijs & Gooren, *supra* note 109, at 275.

¹⁹⁴ *Id.* at 275 (noting that doses are kept as low as possible to keep the side effects to a minimum).

¹⁹⁵ *Id.*

¹⁹⁶ *People v. Wheeler*, No. F051518, 2008 WL 2502521 at *9 (Cal. App. 5 Dist. June 24, 2008) (describing that Wheeler was taken off of MPA injections by his doctor because the drugs cause bone density loss which would have been dangerous for him as he already suffered from HIV); Florida only allows the MPA treatments to be administered to those who are determined to be appropriate candidates for the procedure by a medical expert. FLA. STAT. ANN. § 794.0235(2)(a) (West 2007); Oregon requires that screenings are done to determine that the individual is suited for such therapy upon release. OR. REV. STAT. ANN. § 144.625(2) (West 2003); Montana allows that the treatment only *may* be required. MONT. CODE ANN. § 45-5-512(1)-(2) (West 2007).

¹⁹⁷ See *Kennedy v. Louisiana*, 128 S. Ct. 2641 (2008); *Furman v. Georgia*, 408 U.S. 238, 280 (1972). An individual who is sentenced to MPA injections where the injections would actually cause him serious injury or death would be analogous to sentencing him to a slow death sentence which is already and therefore excessive for the crime of raping or molesting a child without taking the child's life. *Id.*

¹⁹⁸ WIS. STAT. ANN. § 302.11 (West 2005) (undergoing the procedure is a voluntary choice of the individual, but they will not be released on time if they refuse the treatment); Oregon only allows those who are most likely to benefit from it to undergo the procedure. OR. REV. STAT. ANN. § 144.625 (West 2003); Florida requires that the determination of whether the individual can receive the injections must be done within sixty days of imposing the sentence and a court order shall specify the duration of treatment. FLA. STAT. ANN. § 794.0235 (West 2007).

¹⁹⁹ FLA. STAT. ANN. § 794.0235 (West 2007); *Boone v. State*, 933 So.2d 1252, 1254 (Fla. Dist. Ct. App. 2006); *Houston v. State*, 852 So.2d 425, 428 (Fla. Dist. Ct. App. 2003).

²⁰⁰ *Boone*, 933 So.2d at 1254; *Houston*, 852 So.2d at 428.

Those against the treatment may argue that it is not wholly effective.²⁰¹ While findings so far have been impressive, there has not been any study with results showing that 100% of those who underwent the procedure did not re-offend.²⁰² Therefore, *some* people who receive these injections will not be affected by them.²⁰³ However, the drug is so effective that a vast majority of people who do receive the injections will benefit from them.²⁰⁴ As one offender stated, “the only way the streets will ever be safe is to put me on a course of injections where I can be controlled and I can be switched off.”²⁰⁵ Therefore, even if certain offenders do not respond positively to the treatment, it is still a treatment that works for a majority of offenders and will save many children from the horror of sexual molestation.²⁰⁶

In addition, a majority of jurisdictions that use MPA injections require offenders to receive the drug until it is deemed unnecessary.²⁰⁷ So if a person is wrongly convicted of an offense, it is unlikely that he will have to take the drugs for the rest of his life. A wrongly convicted individual who must undergo MPA injections as part of his parole or probation is taken out of prison: he might receive the injections for a while, but will be taken off the drug once it is realized that he does not need it. His sexual functioning will return to normal and any other side effects will reverse themselves over time.²⁰⁸ This is in direct opposition to a wrongly convicted individual who receives a life sentence or the death penalty. In most jurisdictions that administer MPA injections, the wrongly convicted child molester gets out of prison and undergoes injections that he will be taken off of eventually.

Chemical castration is not a disproportionate punishment for the crime of raping or molesting a child. This treatment is nowhere near as harsh as putting a person to death for the rape of a child or even requiring a person to undergo physical castration.²⁰⁹ In fact, administering weekly injections that reduce a person’s sex drive is more humane than a number of other treatments that have been utilized.²¹⁰ Additionally, one of

²⁰¹ See Farkas & Stichman, *supra* note 124, at 270 (“Not all sex offenders are suitable candidates.”).

²⁰² See *supra* note 144 and accompanying text.

²⁰³ *Id.*

²⁰⁴ *Id.*

²⁰⁵ Harrison, *supra* note 1, at 26.

²⁰⁶ See *supra* note 144 and accompanying text.

²⁰⁷ See *supra* notes 151–153, 155 and accompanying text.

²⁰⁸ See *supra* notes 141, 194.

²⁰⁹ See *Kennedy v. Louisiana*, 128 S. Ct. 2641, 2649 (see, e.g., *Louisiana*, *South Carolina*, *Montana*, *Georgia* and *Oklahoma*). See also *supra* notes 112, 118, 119 and accompanying text.

²¹⁰ See *supra* Part III.C. See also *supra* note 85 and accompanying text.

the main reasons the Supreme Court held that the death penalty was an excessive punishment for the rape of a child was because the crime did not involve the taking of a life.²¹¹ The only thing chemical castration takes from the offender is his sexual desire.²¹² Therefore, it is a perfectly proportionate penalty as it is this overwhelming desire that causes the offender to sexually attack innocent children.²¹³ Chemical castration is a serious penalty, but not a disproportionate one.

Chemical castration is the answer to the question of what to do with child molesters and pedophiles. It is unconstitutional to execute such individuals and it is incredibly expensive to keep them imprisoned or involuntarily committed for the rest of their lives.²¹⁴ There is a very high risk that such an offender will end up back in prison, and many treatments that have already been attempted do not have high enough success rates.²¹⁵ In *Trop v. Dulles*, the Supreme Court declared that the law must “draw its meaning from the evolving standards of decency that mark the progress of a maturing society.”²¹⁶ In today’s maturing society, legislation and failed treatments have led us to chemical castration. Therefore, chemical castration is, in fact, the answer. This treatment is what the “standards of decency,” from which the law is to draw its meaning, have evolved into regarding the treatment of child molesters and pedophiles. These crimes are reprehensible, but in today’s society it is necessary to turn to such treatment in order to actually solve the problem rather than simply dispose of the offenders.²¹⁷

B. The Fourteenth Amendment’s Right to Privacy

The Due Process Clause of the Fourteenth Amendment guarantees that no state shall “deprive any person of life, liberty or property without due process of law.”²¹⁸ In addition, the Supreme Court has interpreted the Fourteenth Amendment’s

²¹¹ *Kennedy*, 128 S. Ct. at 2649–50.

²¹² See *supra* note 131 and accompanying text.

²¹³ See *supra* note 35–38 and accompanying text.

²¹⁴ *Kennedy*, 128 S. Ct. 2641. See *supra* Part IV.C.

²¹⁵ See *supra* Parts II.B. & III.C.

²¹⁶ *Trop v. Dulles*, 356 U.S. 86, 101 (1958).

²¹⁷ *Kear-Colwell & Boer, supra* note 101, at 603 (“The aim of treatment is to address the pedophile’s problems in a more holistic manner, as opposed to simply treating the offending behavior as the only issue and ignoring the person who generates the behavior.”); *Cloud, supra* note 6 (stating that most people want to just lock up pedophiles, but most offenders are “tortured by these temptations and relieved we can do something for them”).

²¹⁸ U.S. CONST. amend. XIV § 1.

liberty interests to include an individual's right to privacy.²¹⁹ This makes it slightly problematic to sentence individuals to undergo chemical castration. The procedure infringes on their fundamental right to have a child, as well as the right to refuse medical treatment, both of which are protected under the Fourteenth Amendment's right to privacy.²²⁰ However, making chemical castration a mandatory condition of probation or parole does not unconstitutionally burden the rights of the offender. The state has a legitimate interest in the safety of its children and chemical castration is the least restrictive way to advance such an interest. Therefore, while a convicted child molester's constitutional rights are slightly burdened in sentencing him to undergo the procedure, they are not violated.

1. The Right to Refuse Medical Treatment

The Fourteenth Amendment's right to privacy includes the right of a competent person to refuse treatment as a liberty interest protected under the Due Process Clause.²²¹ In assessing whether a regulation unconstitutionally infringes upon this right, it is necessary to balance the individual's liberty interest against the interest of the state.²²² For example, "if a patient cannot be confined without endangering other[s] . . . and yet he refuses medication that would curb his dangerous tendencies, this would be one factor to weigh in overriding his decision to refuse."²²³

In *Washington v. Harper*, the Supreme Court officially declared that the right to refuse medical treatment is a liberty interest protected under the right to privacy through the Due Process Clause of the Fourteenth Amendment.²²⁴ While it is not a fundamental right that can only be burdened by a compelling state interest, the Court held in *Harper* that a higher standard than rational basis review should be implemented.²²⁵ The

²¹⁹ *Roe v. Wade*, 410 U.S. 113, 152 (1973) ("The Constitution does not explicitly mention any right of privacy . . . [but] the Court has recognized that a right of personal privacy, or a guarantee of certain areas or zones of privacy, does exist under the Constitution.").

²²⁰ Edward A. Fitzgerald, *Chemical Castration: MPA Treatment of the Sexual Offender*, 18 AM. J. CRIM. L. 1, 39–52 (1990); Jason O. Runckel, *Abuse it and Lose it, a Look at California's Mandatory Chemical Castration Law*, 28 PAC. L.J. 547, 562–67 (1997); *Washington v. Harper*, 494 U.S. 210, 221–22 (1990) (stating that the right to refuse medical treatment is constitutionally protected); *Griswold v. Conn.*, 381 U.S. 479 (1965).

²²¹ *Cruzan v. Dir., Miss. Dep't of Mental Health*, 497 U.S. 261, 278 (1990).

²²² *Id.* at 279.

²²³ *Rennie v. Klein*, 462 F. Supp. 1131, 1145 (D. N.J. 1978).

²²⁴ *Harper*, 494 U.S. at 221–22.

²²⁵ See *id.* at 223; Fitzgerald, *supra* note 220, at 48; Runckel, *supra* note 220, at 576.

Supreme Court declared here that in determining if an incarcerated individual could be forcibly medicated, it is necessary that such regulations be “reasonably related to the penological interests.”²²⁶ If the offender is fully competent, then the Fourteenth Amendment gives the offender the right to refuse the MPA injections.²²⁷ Incidentally, while incarcerated, a child molester does not pose a threat to those around him or himself.²²⁸ It is only once he is back in society and living in the community that he represents a threat to others.²²⁹ It is upon the offender’s release that he poses a very serious threat to the community.²³⁰ In this case, the State’s interest in protecting children from sexual attacks justifies the mandatory imposition of the treatment for these offenders and there is no doubt that such a treatment is rationally related to the “penological interest.”²³¹

2. The Right to Have a Child

In *Skinner v. Oklahoma*, the Supreme Court first recognized that everyone has the right to have children.²³² The court stated that procreation is “one of the most basic civil rights of man” which is “fundamental to the very existence and survival of the race.”²³³ Years later in *Griswold v. Connecticut*, the Court declared that there is a right to privacy for married individuals that is protected under the Fourteenth Amendment.²³⁴ This decision invalidated a state law that illegalized contraceptives as it overly invaded the privacy rights of married couples and infringed upon their fundamental right to decide whether or not to have a child.²³⁵ This right to privacy was extended to all individuals in *Eisenstadt v. Baird*.²³⁶ In *Eisenstadt*, the Court ruled that a state law which allowed the distribution of contraceptives to married couples to prevent pregnancy, but not to unmarried individuals for the same reason, violated the Equal Protection Clause of the Fourteenth Amendment.²³⁷ The Court stated that a married couple is not an entity on its own, but is merely “an association of two individuals.”²³⁸ The Court went on

²²⁶ *Harper*, 494 U.S. at 223.

²²⁷ *Id.* at 221.

²²⁸ Fitzgerald, *supra* note 220, at 49.

²²⁹ *Id.*

²³⁰ *Id.*

²³¹ *Id.* at 50.

²³² *Skinner v. Oklahoma*, 316 U.S. 535 (1942).

²³³ *Id.* at 541.

²³⁴ *Griswold v. Conn.*, 381 U.S. 479 (1965).

²³⁵ *Id.*

²³⁶ *Eisenstadt v. Baird*, 405 U.S. 438 (1972).

²³⁷ *Id.*

²³⁸ *Id.* at 453.

to state that “if the right of privacy means anything, it is the right of the individual, married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a person as the decision whether to bear or beget a child.”²³⁹

This right to privacy was again extended in 1973 in *Roe v. Wade*.²⁴⁰ In *Roe*, the Court ruled that the fundamental right to privacy was “broad enough to encompass a woman’s decision whether or not to terminate her pregnancy.”²⁴¹ While a woman’s fundamental right to privacy included the right to choose, the state could constrain this right as long as it had a compelling state interest allowing the constraint to survive a strict scrutiny standard of review.²⁴² The Court held that a woman was not entitled to terminate her pregnancy at any time or for whatever reason she wished.²⁴³ At some point the state’s interest in the protection of fetal life may “become dominant.”²⁴⁴ The state’s interest reaches a compelling level when it is protecting the unborn child at the stage of viability where the fetus is capable of surviving outside of the mother’s womb.²⁴⁵ Here the Court gave the states a way to try to implement regulations on a woman’s constitutionally protected right to an abortion.²⁴⁶

Decades later, the Court used the same reasoning in *Gonzales v. Carhart* in upholding Congress’ 2003 ban on partial birth abortion as constitutional.²⁴⁷ This decision effectively banned women from having abortions in the third trimester of their pregnancy.²⁴⁸ In ruling on *Carhart*, the Court heavily considered the health and life of the unborn child.²⁴⁹ In

²³⁹ *Id.*

²⁴⁰ *Roe v. Wade*, 410 U.S. 113 (1973).

²⁴¹ *Id.* at 153.

²⁴² *Id.* at 155.

²⁴³ *Id.* at 153.

²⁴⁴ *Id.* at 155.

²⁴⁵ *Id.* at 163.

²⁴⁶ *Id.* at 153–55.

²⁴⁷ *Gonzales v. Carhart*, 127 S. Ct. 1610 (2007). *See generally* 18 U.S.C. § 1531 (2006) (banning abortions done at the end of the second and into the third trimester where the fetus has reached the age of viability. The abortion is performed delivering most of the fetus and then performing an overt act to kill the fetus).

²⁴⁸ A woman cannot undergo partial birth abortion at any time unless her life is in danger. 18 U.S.C. § 1531; Telephone Interview with Hotline Operator, Birth Choice Health Clinics (Jan. 10, 2009) (arguing that a partial birth abortion is the only way to perform an abortion once a woman is in her third trimester due to the size of the fetus and the necessary dilation of the woman’s cervix. Today it is only permitted when the life of the mother is in danger. However, in these situations it is more common to induce labor).

²⁴⁹ *Carhart*, 127 S. Ct. at 1626 (keeping intact the decision in *Roe v. Wade* that a woman has the right to an abortion within the first two trimesters as there are other available abortion procedures).

implementing the ban, Congress found partial birth abortion to be a “gruesome and inhumane procedure that is never medically necessary and should be prohibited.”²⁵⁰ The Court found that this type of a procedure had the “power to devalue human life” and requires specific regulations due to the moral and ethical concerns that justify such additional provisions.²⁵¹ The Court supported Congress’ creation of a bright line rule to distinguish abortion from infanticide as it found the two become very similar as the pregnancy goes farther along.²⁵² So while a woman still has a constitutionally protected right to an abortion, her rights are sometimes outweighed by the life of the unborn child.²⁵³ In dealing with such late term abortions, this decision was not about the individual bodily autonomy of the mother or her right to decide whether or not to bear children.²⁵⁴ This decision was, in large part, about the life of the unborn child.²⁵⁵

The rationale in *Carhart* is analogous to the justification for administering MPA injections to convicted child molesters and pedophiles. In dealing with the treatment of child sexual predators, we are not so concerned with the rights or bodily autonomy of the offender.²⁵⁶ While the offender’s rights and needs must be considered, the main point of treating these offenders is to end the sexual victimization of children.²⁵⁷ Therefore, the administration of the treatment is very much for the child’s safety.²⁵⁸ A woman has the right to have an abortion in the earlier stages of her pregnancy.²⁵⁹ Once she enters into the last few months of her pregnancy, the state has stripped her of her right to choose whether or not to have a child.²⁶⁰ It is at that moment when the child’s life becomes dominant over her right to choose.²⁶¹ Similarly, the child predator has the right to fantasize and entertain thoughts about children. It is not until he has actually acted on his thoughts and desires and attacks a child that he would be made to undergo treatment. It is at that moment of conviction when he has crossed the line and the state

²⁵⁰ *Id.* at 1624.

²⁵¹ *Id.* at 1633.

²⁵² *Id.* at 1633–34, 1622 (stating that, at this point, the fetus is so developed that, while it is partially outside the womb before it is terminated, the fetus can move its legs, kick its feet, and open and close its hands).

²⁵³ *Id.*

²⁵⁴ *Id.*

²⁵⁵ *Id.*

²⁵⁶ Harrison, *supra* note 1, at 26.

²⁵⁷ Kear-Colwell & Boer, *supra* note 101, at 601.

²⁵⁸ *Id.*

²⁵⁹ *Carhart*, 127 S. Ct. 1610.

²⁶⁰ *Id.*

²⁶¹ *Id.*

may alter his rights because the life, sexual safety, and bodily autonomy of a child dominate over his right to privacy.

One might argue that administering MPA injections to these offenders would have to meet a higher standard than any restrictions imposed on a woman's right to choose. In *Planned Parenthood v. Casey* the Supreme Court ruled that a woman's right to an abortion is a liberty interest, but not a fundamental right.²⁶² Therefore, instead of a compelling state interest as required when a fundamental right is in jeopardy,²⁶³ the regulation must simply be rationally related to a legitimate state interest and refrain from placing obstacles in a woman's path to obtaining an abortion.²⁶⁴ In contrast, laws requiring convicted sex offenders to undergo chemical castration would have to pass the higher standard of strict scrutiny review.²⁶⁵ In order for a regulation to pass strict scrutiny standards, it must be supported by a compelling state interest and it must be the least restrictive measure possible to achieve this goal.²⁶⁶ Children who go through the horror of molestation often suffer later in life because of the difficulties in coping with such a traumatic experience.²⁶⁷ Keeping children safe from sexual predators is an unquestionably high priority, as evidenced by legislation within the last several decades.²⁶⁸ There is no question that protecting children from sexual predators is a compelling state interest.

In addition, there is almost no question that administering injections of MPA is the least restrictive measure available to advance the states' interest of protecting children. MPA does not necessarily strip an individual of his right to decide whether or not to have a child.²⁶⁹ It simply diminishes the offender's sexual desire.²⁷⁰ If it is determined that the offender can cease treatment, then his testosterone levels will return and the physical effects will reverse.²⁷¹ In addition, while it is more difficult, it is still possible for a person undergoing MPA treatments to have erections and even ejaculate with the help of a partner.²⁷² It is also possible to adjust the dosage, if necessary,

²⁶² *Planned Parenthood of Southeastern Pennsylvania v. Casey*, 505 U.S. 833, 839–40 (1992).

²⁶³ *Id.* at 840.

²⁶⁴ *Id.*

²⁶⁵ Runckel, *supra* note 220, at 563.

²⁶⁶ *Id.* at 569.

²⁶⁷ *See supra* note 41 and accompanying text.

²⁶⁸ *See* statutes cited *supra* notes 146–152 and accompanying text. *See also supra* Part III.B.

²⁶⁹ *See supra* note 141 and accompanying text.

²⁷⁰ *See supra* note 135, 136 and accompanying text.

²⁷¹ *See supra* note 141 and accompanying text.

²⁷² Gimino, *supra* note 131, at 92.

to avoid total impotence and combat unpleasant side effects.²⁷³ Even if the offender is unable to do these things, he can have sperm samples frozen so that he may still have a child. Just because offenders undergoing MPA treatments experience low testosterone levels and a sense of “erotic apathy” does not mean that they no longer have the ability or option to bear children.²⁷⁴

Additionally, MPA injections actually allow offenders to be released from incarceration and even become fully functioning members of society.²⁷⁵ One offender undergoing MPA injections stated that he no longer had “that major sex urge within” him.²⁷⁶ He is now trying to take part in more things because he has “more hope” that he is not going to “get into . . . trouble.”²⁷⁷ When comparing MPA injections to all other implemented or experimented treatments used on offenders,²⁷⁸ chemical castration is clearly the least restrictive means to achieve the compelling state interest of protecting children from sexual attack. Therefore, regulations which require convicted child molesters and pedophiles to undergo chemical castration do not violate the Fourteenth Amendment.

CONCLUSION

In sum, chemical castration is the nation’s next step in dealing with pedophilia and child molestation through punitive and rehabilitative measures. The procedure is, no doubt, a deterrent for some and an immensely therapeutic process for others in learning to control and cope with their deviant sexual desires. Recent legislation evidences the nation’s fear of sexual predators, especially those who pursue children.²⁷⁹ Incarceration and other forms of treatment have done little to actually solve the problem. Those that were most effective went too far, whereas other therapies were often ineffective and borderline barbaric. Chemical castration, on the other hand, is effective, safe, and humane.²⁸⁰

²⁷³ *Id.*

²⁷⁴ Farkas & Stichman, *supra* note 124, at 268 (citation omitted).

²⁷⁵ See statutes cited *supra* notes 146–152 (mandating the use of chemical castration as a condition of release, probation, or parole); *People v. Collins*, 1 Cal. Rptr. 3d. 641, 646–47 (Cal. App. 4 Dist. 2003) (reporting that in states like California, a sexual predator can be released from involuntary commitment when they are no longer present danger of reoffending. This can be done with MPA injections).

²⁷⁶ Harrison, *supra* note 1, at 26.

²⁷⁷ *Id.*

²⁷⁸ See *supra* Part II B. & C.

²⁷⁹ See statutes cited *supra* notes 146–152 and accompanying text. See also *supra* Part III.B.

²⁸⁰ See *supra* note 144, 189 and accompanying text.

Chemical castration is also incredibly cost effective. The cost of the treatment is miniscule when compared to the expense of incarceration or involuntarily commitment in a state hospital. Lastly, and most important, chemical castration *is* constitutional. It is not an overly excessive punishment for the offense charged, especially in comparison with other punishments and treatments that have been utilized for such offenders. It does not violate the right to have a child or the right to refuse treatment due to the compelling interests of the state and the nature of the drug. Chemical castration is the treatment to which we have been led after decades of recidivism and failed treatment for child molesters and pedophiles. This drug can give offenders a second chance at life, and give children less to be afraid of in the world.

Pro [Whose?] Choice: How the Growing Recognition of a Fetus' Right to Life Takes the Constitutionality out of *Roe*

Rachel Warren*

If this suggestion of personhood is established, the [case for abortion rights] of course, collapses, for the fetus' right to life is then guaranteed specifically by the [Fourteenth] Amendment.¹

INTRODUCTION

Roe v. Wade's² granting of constitutional protection for the right to terminate an unwanted pregnancy was, in many ways, just the beginning of the true abortion rights debate.³ The rights of the unborn are a main issue in this debate—mainly whether they have rights as a human person that could protect their lives from ending before birth. While the Court has since declined to officially and fully address the issue, there is a growing recognition of the personhood of the unborn child both from a medical⁴ and social⁵ standpoint. One cannot help but wonder what impact the expanding recognition of fetal rights will have on the already shaky future of *Roe*.

This Comment will argue that recent trends in legislation and modern scientific development call for the fulfillment of *Roe*'s own acknowledgement that the right to life will “collapse” the right to an abortion. Part I lays out the history of abortion

* J.D. Candidate 2010, Chapman University School of Law. B.S. Music Industry, University of Southern California. I would like to thank Professor Celestine McConville for her invaluable guidance throughout the writing process, and Professor Ronald Rotunda and Steve Aden for their thoughtful advice. I would also like to thank my husband, whose endless encouragement and unconditional support will forever amaze me.

¹ *Roe v. Wade*, 410 U.S. 113, 156–57 (1973).

² *Id.* at 113.

³ Amy Lotierzo, *The Unborn Child, a Forgotten Interest: Reexamining Roe in Light of Increased Recognition of Fetal Rights*, 79 TEMP. L. REV. 279, 279 (2006) (pointing out that it was the landmark decision that spurred the “increasing recognition and expansion of the rights of unborn children in the various areas of the law”).

⁴ See Robert L. Stenger, *Embryos, Fetuses, and Babies: Treated as Persons and Treated with Respect*, 2 J. HEALTH & BIOMEDICAL L. 33, 38 (2006) (showing how modern advances in the medical and scientific fields have led to conclusory evidence that human life is present from first existence).

⁵ This Comment uses state and federal laws that give rights to the unborn to gauge the current societal values. See *infra* note 81.

jurisprudence, from the cases predating *Roe*, the *Roe* decision itself, and finally the cases following *Roe*. Part II establishes the widespread acceptance of fetal humanity and rights in both science and law, as well as the likelihood that abortion jurisprudence is ripe for upheaval. Part III then calls attention to the unavoidable competing interests of mother and child which the Court must address, as well as preliminary solutions anticipated by others. Part IV concludes that, when weighing these conflicting interests, the right to life must prevail over the right to an abortion.⁶

I. HISTORY: *ROE V. WADE* AND ITS PROGENY

A. The Road to *Roe*

Historically, the road leading to the Court's landmark decisions in *Roe v. Wade*⁷ and *Doe v. Bolton*⁸ was built upon the Fourteenth Amendment's right of privacy.⁹ Eight years before *Roe* officially constitutionalized the practice of abortion, the Court extended the right of privacy to the use of contraceptives by married couples in *Griswold v. Connecticut*.¹⁰ This decision marked the first time the Court expanded this right to reproductive decisions.¹¹ The Court defended the right of privacy within the context of a private activity between spouses within their home, but never addressed practices independent of marriage or outside the home.¹²

The leap from scrutinizing contraceptive restrictions to abortion laws came six years later when the Court addressed a statute that criminalized abortions unless it was "necessary for

⁶ It is important to understand at the outset that this Comment discusses the conflicting interests of the right to life versus the right of liberty associated with abortion, as opposed to the conflicting rights to live arising from potentially-fatal pregnancies. Because the right to abortion is currently based on the liberty-over-life reasoning, the latter "self-defense" argument is a topic best reserved for another day. It should also be noted that a statute with an exception to save the woman's life was precisely the type of law struck down in *Roe*. See *infra* note 189.

⁷ 410 U.S. 113 (1973).

⁸ 410 U.S. 179 (1973).

⁹ Martin Rhonheimer, *Fundamental Rights, Moral Law, and the Legal Defense of Life in a Constitutional Democracy*, 43 AM. J. JURIS. 135, 157 (1998). The concept of a constitutional right of privacy was first introduced in a Harvard Law Review article by Samuel Warren and Louis Brandeis, where it was suggested that "each individual had the right to choose to share or not to share with others information about his or her 'private life, habits, acts, and relations.'" Dorothy J. Glancy, *The Invention of the Right of Privacy*, 21 ARIZ. L. REV. 1, 2 (1979) (quoting Samuel Warren and Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890)).

¹⁰ *Griswold v. Connecticut*, 381 U.S. 479, 485–86 (1965).

¹¹ *Id.*

¹² *Id.* The Court called the invasion of privacy "repulsive to the notions of privacy surrounding the marriage relationship." *Id.* at 486.

the preservation of the mother's life or health."¹³ In *United States v. Vuitch*, the Court held that a District of Columbia statute was not vague and broadened permissible abortions by expanding the term "health" to include both physical and psychological well-being.¹⁴ The majority did not delve into a privacy rights analysis for this particular statute.¹⁵ However, Justice Douglas' dissent suggested that a "compelling personal interest in marital privacy" also included "the limitation of family size," thus giving the *Griswold* holding a new possible application.¹⁶

While a decision in *Roe* was pending, the Court briefly turned back to contraceptive restrictions in *Eisenstadt v. Baird* and expanded *Griswold* to include the use of contraceptives by unmarried individuals.¹⁷ *Eisenstadt* was also significant for abortion rights, as the Court laid groundwork for extending the right of privacy from simply *preventing* a pregnancy to *terminating* a pregnancy by asserting that "[i]f the right of privacy means anything, it is the right of the individual, married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a person as the decision whether to bear or beget a child."¹⁸

Building on these principles, the Court delivered its decisions in *Roe v. Wade* and *Doe v. Bolton* in 1973. In this pair of cases handed down on the same day, the Court granted constitutional protection for abortions and rejected the notion of an unborn child's status as a person under the Fourteenth Amendment, and hence, its right to life.¹⁹

B. The Landmark Decisions: Constitutional Right of Abortion Solidified

Roe v. Wade and *Doe v. Bolton* arose out of state statutes that, similar to the one upheld in *Vuitch* and found in the majority of the states, prohibited abortions except in situations

¹³ *United States v. Vuitch*, 402 U.S. 62, 68 (1971).

¹⁴ *Id.* at 72; Karen J. Lewis, *Abortion: Judicial Control*, in *ABORTION-MURDER OR MERCY?: ANALYSIS AND BIBLIOGRAPHY* 1, 2 (Francois B. Gerard ed., 2001).

¹⁵ *Vuitch*, 402 U.S. at 73.

¹⁶ *Id.* at 78 (Douglas, J., dissenting).

¹⁷ *Eisenstadt v. Baird*, 405 U.S. 438, 454–55 (1972). The Court acknowledged that the *Griswold* holding was based on the state having no business to enter into the sacred marital relationship, but nevertheless deemed this right to be inherent in individuals as well. *Id.* at 453.

¹⁸ *Id.* (emphasis omitted).

¹⁹ *Roe v. Wade*, 410 U.S. 113, 154 (1973); *Doe v. Bolton*, 410 U.S. 179, 201 (1973).

where it was necessary to save the mother's life.²⁰ In a 7–2 vote, the *Roe* Court invalidated the statute and held that the “right of privacy, whether it be founded in the Fourteenth Amendment’s concept of personal liberty . . . [or] in the Ninth Amendment’s reservation of rights to the people, is broad enough to encompass a woman’s decision whether or not to terminate her pregnancy.”²¹ Thus, the constitutional right to an abortion was established.

But what about the constitutional right to life? The competing interest of fetal life turned on the Court’s decision as to whether the fetus was a person and therefore also had protection under the Fourteenth Amendment.²² One might expect that the right to live becomes inherent once that life begins. However, Justice Blackmun specifically declined to address the issue, writing that the absence of a consensus in scientific and sociological ideology meant that the Court “need not resolve the difficult question of when life begins.”²³ Instead, the Court simply held that the use of the word “person” within the Fourteenth Amendment referred to only post-birth.²⁴ As such, any interest in the protection of the fetus could not be considered compelling enough to legitimize restrictions through the entire pregnancy.²⁵

²⁰ *Roe*, 410 U.S. at 117–118; *Doe*, 410 U.S. at 182–83 n.4. *Roe* was a challenge to a Texas statute that made it a crime to “procure an abortion” at any stage of pregnancy, subject to up to five years in prison with an exception only to save the mother’s life. *Roe*, 410 U.S. at 117 n.1. The class action lawsuit was spearheaded by a pregnant single woman, Jane Roe, along with her physician, James Hubert Hallford, who asserted that the statute was unconstitutionally vague and infringed upon the woman’s individual right of privacy under the First, Fourth, Fifth, Ninth and Fourteenth Amendments. *Id.* at 120. *Doe* was a challenge to a similar Georgia statute, distinguishable from *Roe* in that it provided for more exceptions—serious threats to the mother’s health, likely grave birth defects, and pregnancies resulting from rape—and proscribed a sentence of up to ten years. *Doe*, 410 U.S. at 183.

²¹ *Roe*, 410 U.S. at 153.

²² *See id.* at 156–57.

²³ *Id.* at 159. One scholar has pointed out that just a few pages later, Justice Blackmun referred to the unborn child as “the *potentiality* of human life,” thereby indicating that a judgment call was in fact made as to whether the unborn was “human.” Michelle Haynes, *Inner Turmoil: Redefining the Individual and the Conflict of Rights Between Woman and Fetus Created by the Prenatal Protection Act*, 11 TEX. WESLEYAN L. REV. 131, 135 (2004) (quoting *Roe*, 410 U.S. at 162 (emphasis added)).

²⁴ Charles I. Lugosi, *Respecting Human Life in 21st Century America: A Moral Perspective to Extend Civil Rights to the Unborn from Creation to Natural Death*, 48 ST. LOUIS U. L. J. 425, 431 (2004). Rather than rely on information provided by the Defense regarding fetal development, the Court declined to find a clear definition of “person” within the Constitution. Instead, the Court focused on the Constitution’s use of the word “citizen” which referenced a postnatal state. Despite the fact that “person,” not “citizen,” was used in the Fourteenth Amendment, and despite the fact that state statutes expressly defined unborn children as human beings, the Court held that the unborn child had no rights under the Fourteenth Amendment. *Id.*; *Roe v. Wade*, 410 U.S. 113, 157–58 (1973).

²⁵ *Roe*, 410 U.S. at 162.

While the unborn child could not assert any rights, the Court recognized that this newly-established “fundamental right” to an abortion was qualified and therefore subject to possible restrictions that served a compelling state interest.²⁶ The Court again refused to use the beginning of life as a threshold and instead held that any state interest in preserving the life of the unborn would not be sustained until the fetus reached the point of what Justice Blackmun called “viability”—when the child could “live outside the mother’s womb and ultimately function as a contributing member of society.”²⁷

Although science at that time was unclear as to when “viability” began, the Court set standards based on “present medical knowledge” of the trimester progress of the pregnancy and declared that viability was reached only after the first trimester.²⁸ Within the first trimester, no state interference was permissible; any decision to abort the pregnancy was left to the mother and her physician.²⁹ In the second trimester, the state may begin to assert regulations, but only if reasonably related to the mother’s physical well-being.³⁰ Even into the third trimester when the state interest becomes compelling, any regulations must contain an exception “to preserve the life or health of the mother.”³¹

Significantly, however, the Court left the door open to revisit the issue upon more scientific evidence as to the personhood of the fetus, stating that “[w]hen those trained in the respective disciplines of medicine, philosophy, and theology are unable to arrive at any consensus, the judiciary, at this point in the development of man’s knowledge, is not in a position to speculate as to the answer.”³² Moreover, the Court stated that if the personhood of the fetus were established, the case for abortion

²⁶ *Id.* at 154; Lugosi, *supra* note 24, at 431–32. Because the unborn child had just been deemed a non-person, the compelling state interest was in the protection of “potential life.” See Michael S. Robbins, *The Fetal Protection Act: Redefining “Person” for the Purposes of Arkansas’ Criminal Homicide Statutes*, 54 ARK. L. REV. 75, 85 (2001).

²⁷ Robbins, *supra* note 26, at 85; Lewis, *supra* note 14, at 4.

²⁸ *Roe*, 410 U.S. at 163.

²⁹ *Id.*; Lugosi, *supra* note 24, at 432.

³⁰ *Roe*, 410 U.S. at 163; Haynes, *supra* note 23, at 135–36.

³¹ Lugosi, *supra* note 24, at 432 (quoting *Roe v. Wade*, 410 U.S. 113, 163–64). As will be shown in the concurrent opinion of *Doe v. Bolton*, to preserve the “health” of the mother means “all factors, physical, emotional, psychological, familial, and the woman’s age.” *Doe v. Bolton*, 410 U.S. 179, 192 (1973). See *infra* note 35 for further analysis of the combined holdings.

³² *Roe*, 410 U.S. at 159. The modern consensus of these areas is discussed in Part II, *infra*.

rights “of course, collapses, for *the fetus’ right to life is then guaranteed* specifically by the [Fourteenth] Amendment.”³³

In *Doe v. Bolton*, the Court built upon *Roe*’s newly-founded principles and expanded *Vuitch* to define “health” as including “all factors—physical, emotional, psychological, familial, and the woman’s age—relevant to the well-being of the patient.”³⁴ The significance of this holding is that states were now severely limited in establishing any compelling interest in restricting *Roe*’s constitutional right.³⁵

C. The Aftermath: Regret and Calls for a Re-Visitation

In the decades that followed *Roe* and *Doe*, the Court continued to expand abortion rights based on the unquestioned principles of the right of privacy and non-person fetal character. These expansions included striking down requirements of spousal or parental notification,³⁶ scrutinizing attempts to restrict public funding for non-therapeutic abortions³⁷ and narrowing the requirements for informed consent.³⁸ Within a few years, the Court again addressed the viability issue, reiterating

³³ *Id.* at 156–57 (emphasis added). The fulfillment of this acknowledgement or “promise” by the *Roe* Court is discussed further in Part IV, *infra*.

³⁴ *Doe*, 410 U.S. 179, 192.

³⁵ JOSEPH W. DELLAPENNA, DISPELLING THE MYTHS OF ABORTION HISTORY 695 (2006). Many states had restricted abortions to that which was necessary to preserve the health of the mother, with the intent that abortions be permitted only in life-or-death situations. With *Doe* giving “health” such an expansive definition, such statutes were now unconstitutional. Instead, a woman was entitled to what scholars have called “abortion on demand.” FRANCIS J. BECKWITH, DEFENDING LIFE: A MORAL AND LEGAL CASE AGAINST ABORTION CHOICE 20 (2007).

³⁶ See *Planned Parenthood v. Danforth* 428 U.S. 52, 69, 74–75 (1976) (holding that a married woman need not secure her husband’s consent to get an abortion as this would constitute an unconstitutional veto power of a third party; likewise, a minor need not secure consent of a parent to get an abortion); *Bellotti v. Baird*, 443 U.S. 622, 643 (1979) (affirming *Danforth*’s invalidation of blanket parental consent requirements); *City of Akron v. Akron Center for Reproductive Health*, 462 U.S. 416, 441–42 (1983) (affirming *Danforth* and *Bellotti*). But see *Planned Parenthood v. Ashcroft*, 462 U.S. 476, 492–94 (1983) (approving a parental consent statute because it contained a provision for judicial bypass).

³⁷ *Beal v. Doe*, 432 U.S. 438, 443–47 (1977); *Maher v. Roe*, 432 U.S. 464, 470 (1977); see also *Harris v. McRae*, 448 U.S. 297, 302, 322 (1980) (upholding the Hyde Amendment, which placed a limit on federal funding for abortions through HHS programs such as Medicaid to cases where an abortion would be necessary to save the woman’s life, when pregnancy has resulted from rape or incest, or situations where the pregnancy caused the woman “severe and long-lasting physical health damage”).

³⁸ See *City of Akron*, 462 U.S. at 450 (invalidating a 24-hour waiting period after consent); *Thornburgh v. American College of Obstetricians & Gynecologists*, 476 U.S. 747, 759–60 (1986) (striking down an informed consent statute that provided for disclosure of fetal development, possible psychological side effects and alternatives to the procedure on the grounds that this information might discourage the patient).

that the health of the woman would always prevail over the life of the unborn child.³⁹

Even in affirming *Roe* and *Doe*, the holding and reasoning began to be questioned by the Court itself. A decade after *Roe*, medical technology had advanced enough to detect “viability” as early as twenty weeks.⁴⁰ As a result, in *City of Akron v. Akron Center for Reproductive Health*,⁴¹ three Justices—O’Connor, White and Rehnquist—voted to discard the use of *Roe* as a precedent. Justice O’Connor criticized the trimester framework as “completely unworkable” because the stages of pregnancy would always “differ according to the level of medical technology available.”⁴²

Another significant criticism of *Roe* arose in *Thornburgh v. American College of Obstetricians & Gynecologists*.⁴³ Chief Justice Burger, who had concurred in *Roe*, now filed a dissenting opinion in which he called for its reexamination.⁴⁴ Chief Justice Burger expressed his regret that the concerns listed in the *Roe* dissents—endorsement of “abortion on demand” and the invalidation of any interest to protect fetal life—had now become an unwanted reality.⁴⁵ Justice White attacked *Roe*’s illogical viability standard, pointing out that “the State’s interest, if compelling after viability, is equally compelling before viability.”⁴⁶ Justice Stevens, though concurring in the majority opinion, took the opportunity to explain further that the right to

³⁹ *Colautti v. Franklin*, 439 U.S. 379, 386–87 (1979). In *Colautti*, the Court considered whether to uphold a statute that required physicians to “preserve the life and health of the fetus [as though it were] intended to be born and not aborted” when the fetus is viable or if there is “sufficient reason to believe that the fetus may be viable.” *Id.* at 380–81 n.1. The Court found the statute was unconstitutionally vague because the term “may be viable” was not distinguishable from *Roe*’s definition of “viability.” *Id.* at 390. It also reaffirmed that the health of the woman must always prevail over the life of the fetus. *Id.* at 400–01.

⁴⁰ BECKWITH, *supra* note 35, at 20. *Roe* had declared viability occurring only after twenty-four to twenty-eight weeks. *Id.*

⁴¹ *City of Akron*, 462 U.S. 416, 418. Here, the majority acted characteristically in striking down an ordinance that required all second-trimester abortions to be performed in a hospital because it failed to serve the state’s interest of protecting the woman’s health. *Id.* at 449–52.

⁴² *Id.* at 452, 454 (O’Connor, J., dissenting). Justice O’Connor went so far as to admit that if viability kept getting pushed back all the way to conception, the right to have an abortion would disappear. BECKWITH, *supra* note 35, at 20.

⁴³ *Thornburgh v. American College of Obstetricians & Gynecologists*, 476 U.S. 747 (1986). Here, the Court invalidated a statute requiring physicians to disclose alternative procedures, fetal development and possible psychological side-effects when obtaining informed consent and called for a second physician to be present in situations where the fetus might survive. *Id.* at 759–64.

⁴⁴ *Id.* at 785 (Burger, C.J., dissenting).

⁴⁵ *Id.* at 783–84 (Burger, C.J., dissenting).

⁴⁶ *Id.* at 795 (White, J., dissenting).

life would trump the right to abortion in all states if the personhood of a fetus were to be recognized, stating:

[I]ndeed, if there is not such a difference [between a child and a fetus], the permissibility of terminating the life of a fetus could scarcely be left to the will of the state legislatures. And if distinctions may be drawn between a fetus and a human being in terms of the state interest in their protection—even though the fetus represents one of “those who will be citizens”—it seems to me quite odd to argue that distinctions may not also be drawn between the state interest in protecting the freshly fertilized egg and the state interest in protecting the 9-month-gestated, fully sentient fetus on the eve of birth. Recognition of this distinction is supported not only by logic, but also by history and by our shared experiences.⁴⁷

This logical reasoning seems to have resonated when the Court found itself addressing the viability issue again only a few years later in *Webster v. Reproductive Health Services*.⁴⁸ The Court adopted the *City of Akron* and *Thornburgh* dissents in admitting that its constitutional construction of *Roe*’s trimester system was “unsound in principle and unworkable” and thus *stare decisis* should not be invoked to uphold it.⁴⁹

The Court stopped short, however, of overruling *Roe*.⁵⁰ In admonishing the majority for failing to do so, Justice Scalia pointed out that *Roe* itself precluded any state from attempting to enact an identical statute, and therefore *Roe* would only ever

⁴⁷ *Id.* at 779 (Stevens, J., concurring) (internal footnotes omitted).

⁴⁸ *Webster v. Reproductive Health Services*, 492 U.S. 490 (1989). Here, the preamble to the challenged Missouri statute provided that the State recognized human life from conception and therefore all unborn children had rights as full citizens. MO. REV. STAT. § 1.205 (2000). Though this was contrary to its characterization of the fetus in *Roe*, the Court declined to determine the constitutionality of the preamble on the grounds that it did not in itself place regulations on abortion. *Webster*, 492 U.S. at 506. The Court asserted that it would place “no limitation on the authority of a State to make a value judgment favoring childbirth over abortion.” *Id.* (quoting *Maier v. Roe*, 432 U.S. 464, 474 (1977)).

⁴⁹ *Webster*, 492 U.S. at 518. According to the Court, the “rigid *Roe* framework” of trimesters and viability was inconsistent with the Constitution’s general terms and principles. The resulting effect of *Roe* was therefore a “web of legal rules that have become increasingly intricate, resembling a code of regulations rather than a body of constitutional doctrine.” *Id.* The Court also echoed Justice White’s dissent in *Thornburgh* by stating “we do not see why the State’s interest in protecting potential human life should come into existence only at the point of viability, and that there should therefore be a rigid line allowing state regulation after viability but prohibiting it before viability.” *Id.* at 519.

⁵⁰ BECKWITH, *supra* note 35, at 33. The reason cited by the Court was that the differing facts of the two cases made a complete reversal inappropriate: in *Roe*, the Court was considering a Texas statute that criminalized all abortions, while in the present case Missouri had merely determined that potential human life must be safeguarded at viability. *Webster*, 492 U.S. at 521.

“be disassembled doorjamb by doorjamb, and never entirely brought down, no matter how wrong it may be.”⁵¹

One of these disassembled “doorjamb” came with *Planned Parenthood v. Casey*.⁵² The plurality opinion rejected *Webster*’s outlook on *stare decisis* and announced that *Roe* was not “unworkable” despite its “engendered disapproval” and thus upheld *Roe*’s viability standard.⁵³ Yet, its refusal to overturn *Roe* was based solely on “precedential force” rather than the “soundness of *Roe*’s resolution of the issue;” the *Casey* opinion never affirmed that *Roe* was a correct constitutional interpretation.⁵⁴ Such an omission subtly, but quite tellingly, distances the Court from *Roe*’s reasoning and provides evidence of its recognition of the landmark decision’s flaws.

Most significant to abortion jurisprudence applicability, the *Casey* Court did not label abortion as a “fundamental” right.⁵⁵ Instead, the Court acknowledged that *Roe*’s progeny had vastly ignored the state’s legitimate interest in protecting potential life because the rigid trimester framework made it nearly impossible for any regulation to be imposed within the first twelve weeks of pregnancy.⁵⁶ The Court attempted to repair this flaw by replacing *Roe*’s strict scrutiny standard with an “undue burden” test.⁵⁷ This new standard permits states to enact regulations with the “purpose or effect” to protect potential life insofar as it does not place a “substantial obstacle in the path of a woman’s choice.”⁵⁸ As a result, state regulations are more likely to pass scrutiny.⁵⁹

⁵¹ *Webster*, 492 U.S. at 537 (Scalia, J., concurring).

⁵² *Planned Parenthood v. Casey*, 505 U.S. 833 (1992). The Court evaluated the constitutionality of a Pennsylvania Act with the usual categories of restrictions: informed consent, parental approval and spousal notification. 18 PA. CONS. STAT. §§ 3203–3209 (2000). The Court upheld four of the five provisions in the Act and struck down the spousal notification requirement. BECKWITH, *supra* note 35, at 33.

⁵³ *Casey*, 505 U.S. at 860–61. Interestingly, this joint opinion included Justice Kennedy, who had just three years prior joined Chief Justice Rehnquist’s *Webster* opinion that asserted exactly the opposite. *Webster*, 492 U.S. at 496. After outlining the ramifications of overturning such a landmark decision, the *Casey* Court admitted that its decision was in part due to the fear that such an overruling would cause a loss of confidence in the Judiciary. *Casey*, 505 U.S. at 867.

⁵⁴ *Casey*, 505 U.S. at 871; PAUL BENJAMIN LINTON, ABORTION UNDER STATE CONSTITUTIONS: A STATE-BY-STATE ANALYSIS 6–7 n.13 (2008).

⁵⁵ LINTON, *supra* note 54, at 6 n.13.

⁵⁶ *Casey*, 505 U.S. at 872. The Court pointed to *Thornburg* and *City of Akron* as examples. *Id.*

⁵⁷ BECKWITH, *supra* note 35, at 33–34; Lewis, *supra* note 14, at 14.

⁵⁸ *Casey*, 505 U.S. at 877. See also *Mazurek v. Armstrong*, 520 U.S. 968, 972 (1997) (modifying the two-pronged *Casey* test of “purpose or effect” to just “effect”).

⁵⁹ Lewis, *supra* note 14, at 14. The Court further elaborated that a state’s interest in protecting the unborn meant that it could enact rules to ensure that the woman was fully informed of her options. Under this standard, a state may even enforce measures to

Casey's first major application by the Court was in *Stenberg v. Carhart*.⁶⁰ Here, the Court struck down a partial-birth abortion ban on the grounds that it placed an undue burden in the woman's path to getting the abortion.⁶¹ Because this procedure, known as Dilation and Extraction ("D&X"),⁶² could be performed pre-viability,⁶³ and because the language of the statute could be confused for a permissible pre and post-viability procedure called Dilation and Evacuation ("D&E"),⁶⁴ the Court found that the statute posed an undue, and thus unconstitutional, burden.⁶⁵ The Court ignored the obvious viability issue, even though the child would have survived out of the womb *had the doctor not killed her* before delivery was complete.⁶⁶

In response to *Stenberg*, Congress passed the Partial-Birth Abortion Ban Act of 2003, which remedied the flaws in the Nebraska statute by clearly describing the procedure so as not to be confused with D&E.⁶⁷ Congress also provided an exception to

persuade a woman to choose childbirth over abortion. This includes philosophic and social arguments, even during the first trimester. *Casey*, 505 U.S. 872–73, 878.

⁶⁰ *Stenberg v. Carhart*, 530 U.S. 914 (2000).

⁶¹ *Id.* at 929–30. The Nebraska statute at issue restricted the practice of the partial delivery of "a living unborn child before killing the unborn child and completing delivery." NEB. REV. STAT. ANN. § 28-326(9) (LexisNexis 2008). The Court also invalidated the statute because it made an exception only to *save the life* of the mother, rather than the *Casey*-required exception to *promote the health* of the mother. *Stenberg*, 530 U.S. at 930.

⁶² The D&X procedure, also known as "intact D&E," is accomplished by the physician first dilating the woman's cervix and then using forceps to rotate the child to the breach position and pull the legs, body and arms through the cervix and vagina. Once the head is visible and lodged in the cervix, the physician forces scissors into the base of the skull and makes a large opening, in which he inserts a suction catheter to suck out the child's brains and collapse the skull. Once all contents of the brain are sucked out, the physician completes delivery by pulling out the emptied head. The partially-born child has been observed clasp her fingers and kicking her feet until the physician begins to use the suction tube, at which point the body goes limp. *Gonzales v. Carhart*, 550 U.S. 124, 137–39 (2007).

⁶³ Even the Plaintiff physician admitted that the D&X procedure would never be attempted before the sixteenth week but instead was actually intended for later-term pregnancies. *Stenberg*, 530 U.S. at 927–28.

⁶⁴ The D&E procedure consists of the physician first dilating the woman's cervix and then inserting forceps, which are used to grab the unborn child and pull it through the cervix and vagina. In the process, the friction of the forceps causes the child to be torn apart, and the physician removes the body piece by piece until completely removed. It typically takes the physician 10–15 passes to remove all pieces of the dismembered body. *Gonzales*, 550 U.S. at 135–36.

⁶⁵ *Stenberg*, 530 U.S. at 938–40.

⁶⁶ The separate dissenting opinions of Chief Justice Rehnquist and Justices Scalia, Kennedy and Thomas, however, did not let the issue slide. Each found that the very description of the procedure itself lent to an undisputable finding of protectable life and suggest how close the Court had been to overturning *Roe*. *Stenberg*, 530 U.S. at 952 (Rehnquist, C.J., dissenting); *Id.* at 953–56 (Scalia, J., dissenting); *Id.* at 956–79 (Kennedy, J., dissenting); *Id.* at 980–1020 (Thomas, J., dissenting).

⁶⁷ 18 U.S.C. § 1531 (2006).

save the mother's life⁶⁸ and, unlike the state legislature, successfully cited findings that the procedure would never be medically necessary to preserve the woman's health.⁶⁹ Thus, when challenged in *Gonzales v. Carhart*, the Court was obliged to find the undue burden in *Stenberg* remedied and upheld the ban.⁷⁰ In addition to making strides towards protecting the unborn child, *Gonzales* also disassembled another *Roe* "doorjamb" by accepting the absence of a health exception.⁷¹ The Court found the exception to save the mother's life sufficient, even though it essentially requires the woman to go to court first.⁷² This is a far cry from *Roe*'s establishment of abortion on demand.⁷³ While concurring in the majority, Justice Thomas—joined by Justice Scalia—reiterated that "the Court's abortion jurisprudence, including *Casey* and *Roe* . . . has no basis in the Constitution."⁷⁴

It is apparent that the abortion issue is as ripe as it has ever been. *Casey* has re-opened the door to fetal rights, and *Roe* itself admitted that the legal status of abortion rights should change upon a better understanding of "when life begins."⁷⁵ Thus, an understanding of today's scientific progress and the state's acceptance of the unborn child's personhood has a profound impact on the abortion issue.⁷⁶ With the trend moving toward recognition of fetal humanity, a fundamental issue arises which cannot be ignored in an inevitable reevaluation of *Roe*: does a woman's right to privacy truly outweigh a child's right to live?⁷⁷

⁶⁸ 18 U.S.C. § 1531(a) (2006) (stating that the prohibition on the procedure "does not apply to a partial-birth abortion that is necessary to save the life of a mother whose life is endangered").

⁶⁹ Partial-Birth Abortion Ban Act of 2003, Pub. L. No. 108–105, § 2(13), 117 Stat. 1201 (2003).

⁷⁰ *Gonzales v. Carhart*, 550 U.S. 124, 168 (2007).

⁷¹ *Id.* at 161–65.

⁷² 18 U.S.C. § 1531(d)(1) (2006); *Gonzales*, 550 U.S. at 168.

⁷³ See *supra* note 35 and accompanying text.

⁷⁴ *Gonzales*, 550 U.S. at 169 (Thomas, J., concurring).

⁷⁵ *Roe v. Wade*, 410 U.S. 113, 159 (1973); Lugosi, *supra* note 24, at 437 (arguing that *Casey* "suggest[s] that it is time to think 'outside the box' and directly answer two questions: whether, as a matter of law, the unborn are living human beings and whether the law should confer constitutional personhood on unborn human beings from the time of conception until the time of natural death").

⁷⁶ The scientific data and trends in state laws have been pointing toward a widespread recognition that the unborn child is not only a living human being, but also a citizen entitled to protection. The analysis of this data will be discussed immediately proceeding in Part II, *infra*.

⁷⁷ This issue, its resolution among scholars and this author's solution will be discussed in Parts III and IV, *infra*.

II. CURRENT DEVELOPMENTS: THE REJECTION OF *ROE* AND ACCEPTANCE OF FETAL PERSONHOOD

Within the *Roe* opinion itself, Justice Blackmun's majority admitted that its rejection of the personhood of—and the constitutional protection for—the life of the fetus was related to the fact that “those trained in the respective disciplines of medicine, philosophy, and theology are unable to arrive at any consensus.”⁷⁸ While there may not have been such a consensus at that time, the passing of thirty-six years has allowed the expansion of medical technology and social understanding to close this gap. Today, there is a trend toward recognizing the unborn as both a *human* as well as a *person* afforded lawful protection.⁷⁹ Modern scientific advances in the understanding of human development show that the fetus is indeed a human being from her first day of existence.⁸⁰ Societal values as reflected in current state laws⁸¹ show trends towards fetal personhood and citizenship—thirty-six states consider the killing of the fetus a form of homicide.⁸² This is an increase from twenty-seven states just five years ago.⁸³ Further, even several Supreme Court Justices who dissented in *Stenger* have expressed recognition of fetal humanity.⁸⁴ These developments, coupled with the ongoing cumulative criticism of *Roe*'s reasoning,⁸⁵ call for a reevaluation of the abortion rights issue with new consideration given to the life of the unborn.

A. The Widespread Criticisms of Wrongful Reasoning

An argument for fetal personhood would be moot if *Roe* and its progeny were universally hailed as an accurate constitutional

⁷⁸ *Roe*, 410 U.S. at 159.

⁷⁹ The recognition of both humanity and personhood is important in a reevaluation of *Roe*. Scientific proof of the moment the unborn becomes a human being is key; however, it means nothing if the Constitution does not protect this human as a “person” under its law. This is where trends in both state and federal law become essential, as each have given protection to the unborn child as a person and a citizen.

⁸⁰ *Stenger*, *supra* note 4, at 38.

⁸¹ While “philosophy” by its very nature may never be conclusive, societal value trends are the best gauge of the philosophical climate of the times. This Comment uses current trends in state laws to measure the values of society as a whole, and as will be discussed, *infra* Subsection C, the results are quite telling.

⁸² Americans United for Life, *Fetal Homicide: Model Legislation & Policy Guide for the 2009 Legislative Year*, available at http://www.aul.org/xm_client/client_documents/LegislativeGuides/AUL_Fetal-Homicide-2009-LG.pdf (outlining the states with such provisions and specifying when in the gestation period these rights are realized).

⁸³ M. Todd Parker, *A Changing of the Guard: The Propriety of Appointing Guardians for Fetuses*, 48 ST. LOUIS U. L.J. 1419, 1428 (2004).

⁸⁴ See *Stenberg v. Carhart*, 530 U.S. 914, 952 (Rehnquist, C.J., dissenting); *Id.* at 953–56 (Scalia, J., dissenting); *Id.* at 956–79 (Kennedy, J., dissenting); *Id.* at 980–1020 (Thomas, J., dissenting).

⁸⁵ BECKWITH, *supra* note 35, at 23.

application. As discussed above, however, the opinion sparked regret and calls for an overturning by the Justices themselves.⁸⁶ Congress attempted to take matters into its own hands and tried to pass constitutional amendments to overrule and reverse the case.⁸⁷ While both governmental attempts have yet to be successful, scholarly criticism continues to unravel the constitutional flaws. Scholars have overwhelmingly shown that the history of abortion law laid out by Justice Blackmun in *Roe* is inherently flawed.⁸⁸ The grounding for *Roe* is said to be “untrustworthy and essentially worthless.”⁸⁹ Scholars are nearly unanimous in recognizing that the primary purpose of anti-abortion laws was to provide protection for the unborn.⁹⁰ For *Roe* to conclude otherwise was “fundamentally erroneous.”⁹¹ Moreover, notable constitutional law scholars have asserted that *Roe* has no constitutional foundation.⁹²

Indeed, Justice Blackmun’s conclusion that a fetus is not a constitutional “person” was based on his assertion that the State failed to cite a case where the fetus was given such status.⁹³ However, there was in fact a federal case holding just that—a

⁸⁶ See *supra* Part I.C for a recap of *Roe*’s immediate aftermath.

⁸⁷ Karen J. Lewis, et. al., *Abortion: Legislative Response*, in ABORTION—MURDER OR MERCY?: ANALYSIS AND BIBLIOGRAPHY 17, 24 (Francois B. Gerard ed., 2001). The 94th Congress alone introduced almost eighty such amendments, although each encountered difficulty reaching the floor of the Senate and the House. The 98th Congress, however, debated and voted on a constitutional amendment that stated “[a] right to abortion is not secured by this Constitution.” This amendment was defeated with a vote of 50–49. *Id.* at 24–25.

⁸⁸ BECKWITH, *supra* note 35, at 23; Lotierzo, *supra* note 3, at 280; Gregory J. Roden, *Roe v. Wade and the Common Law: Denying the Blessings of Liberty to Our Posterity*, 35 UWLA L. REV. 212, 256 (2003) (demonstrating that there is no basis in pre-*Roe* common law that would allow the Court to bind the states in 1973 to what it believed was the English common law at the founding of the Country).

⁸⁹ BECKWITH, *supra* note 35, at 23.

⁹⁰ See, e.g., James S. Witherspoon, *Reexamining Roe: Nineteenth Century Abortion Statutes and the Fourteenth Amendment*, 17 ST. MARY’S L. J. 29, 70 (1985) (“[T]he primary purpose of the nineteenth-century antiabortion statutes was to protect the lives of unborn children is clearly shown by the terms of the statutes themselves.”); Clarke D. Forsythe & Stephen B. Presser, *Restoring Self-Government on Abortion: A Federalism Amendment*, 10 TEX. REV. L. & POL. 301, 310 (2006) (noting that there are sixty-four cases from forty states establishing that the objectives of such state laws were for the protection of the unborn child); BECKWITH, *supra* note 35, at 23.

⁹¹ BECKWITH, *supra* note 35, at 24 (quoting James S. Witherspoon, *Reexamining Roe: Nineteenth Century Abortion Statutes and the Fourteenth Amendment*, 17 ST. MARY’S L. J. 29, 70 (1985)). In Justice Blackmun’s defense, it has been suggested that his reasoning was flawed because he relied on articles written by a scholar whose work was not discredited until after the Court’s opinion. *Id.*

⁹² Forsythe & Presser, *supra* note 90, at 314–15. Such scholars include Alexander Bickel, Archibald Cox, John Hart Ely, Philip Kurland, Richard Epstein, Mary Ann Glendon, Gerald Gunther, Robert Nagel, Michael Perry, and Harry Wellington. *Id.*

⁹³ BECKWITH, *supra* note 35, at 26–27.

case cited in the *Roe* opinion itself.⁹⁴ The case was *Steinberg v. Brown*, where the federal district court in the Northern District of Ohio had declined to extend the *Griswold* reasoning to abortion rights because at that point, “the preliminaries have ended, and a new life has begun. Once human life has commenced, the constitutional protections found in the Fifth and Fourteenth Amendments impose upon the state the duty of safeguarding it.”⁹⁵ Moreover, there were more than fifty cases on the books, both federal and state, recognizing that the unborn child had explicit or implicit personhood.⁹⁶ The Court had, and still has, no constitutional prohibition to declare a fetus a person.⁹⁷

Thus, *Roe* and its progeny sit on shaky ground. If a consensus on fetal humanity can be established, then the Court must stay true to Justice Blackmun’s admission and not only revisit, but also overturn, the infamous opinion.

B. Scientific Evidence of Fetal Humanity

To grasp the true significance of scientific developments, one must first become familiar with the state of the medical understandings before *Roe*. Historically, the commonly accepted signifier of human life was known as “quickening,”⁹⁸ the point at which the mother could feel the baby move.⁹⁹ The belief was that once quickening was detectible, the unborn must have received a soul.¹⁰⁰ The only reason quickening was ever a standard was because this was the point at which people could determine that the fetus was alive—and because it was alive, it was considered human.¹⁰¹ As more biological facts of human development were discovered, the quickening theory was dismissed.¹⁰² Instead, scientific technology has led to the medical conclusion that human life starts at conception.¹⁰³

⁹⁴ *Id.* Justice Blackmun cited *Steinberg v. Brown* as an example of a District Court upholding an abortion restriction that was not void for vagueness or overbroad. *Roe v. Wade*, 410 U.S. 113, 154–55 (1973).

⁹⁵ *Steinberg v. Brown*, 321 F. Supp. 741, 746–747 (N.D. Ohio 1970).

⁹⁶ Gregory J. Roden, *Prenatal Tort Law and the Personhood of the Unborn Child: A Separate Legal Existence*, 16 ST. THOMAS L. REV. 207, 270 (2003).

⁹⁷ Lugosi, *supra* note 24, at 438.

⁹⁸ BECKWITH, *supra* note 35, at 24.

⁹⁹ BabiesOnline.com, Baby, Pregnancy, and Parenting Information: What is Quickening?, <http://www.babiesonline.com/articles/pregnancy/quickening.asp> (last visited September 29, 2009). It is now known that quickening is unreliable because, although the baby actually starts moving by the eighth or ninth week, it could take the woman up to 26 weeks before she can detect it. *Id.*

¹⁰⁰ BECKWITH, *supra* note 35, at 24.

¹⁰¹ *Id.* at 25.

¹⁰² *Id.*

¹⁰³ *Id.*

1. The Beginning of Life

Functions that have traditionally been used to define “life” have been discovered to occur well within the first trimester, before the mother may even know that she’s pregnant.¹⁰⁴ The baby’s heartbeat is detectable by the eighteenth day.¹⁰⁵ Brain cells are developed within the child’s first two weeks, while actual brain activity has been monitored by the fifth week.¹⁰⁶ There is evidence that an unborn child can feel pain within eight weeks—meaning that the child can, in fact, feel pain during even the earliest abortions.¹⁰⁷

2. Life that is Human

Today, the scientific and medical community has widely recognized the existence of human life from the very moment of conception. Many contemporary human embryologists maintain that a new human being comes to be at syngamy (fertilization), which is the point at which the maternal and paternal chromosomes merge to form a set.¹⁰⁸ Indeed, research shows that even after initial fertilization, “[n]o substantial changes take place . . . [the child] *is the same individual organism as the adult into whom it later develops.*”¹⁰⁹ This means that once the chromosomes have merged in syngamy, there is no further genetic information needed to “make” the unborn child into an individual human being.¹¹⁰ The only things she needs for growth and development are oxygen, food, water and healthy interaction with her environment.¹¹¹

¹⁰⁴ It is interesting to note that *Casey* defined the beginning of life at respiratory function, which seems quite arbitrary when considering that loss of such function has never signified death. Charles I. Lugosi, *Conforming to the Rule of Law: When Person and Human Being Finally Mean the Same Thing in Fourteenth Amendment Jurisprudence*, 22 ISSUES L. & MED. 119, 125–126 (2007).

¹⁰⁵ *See id.*

¹⁰⁶ *Id.* at 125 n.28.

¹⁰⁷ Kerby Anderson, *Arguments Against Abortion*, LIFEWAY, Feb. 13, 2004, <http://www.lifeway.com/> (search for “Kerby Anderson”; then follow “Arguments Against Abortion” hyperlink.) This finding directly rebuts assertions in previous cases that abortion is permissible because the fetus can feel no pain. *See, e.g., Webster v. Reproductive Health Services*, 492 U.S. 490, 569 (1989) (Stevens, J., concurring in part and dissenting in part) (“There can be no interest in protecting the newly fertilized egg from physical pain or mental anguish, because the capacity for such suffering does not yet exist.”).

¹⁰⁸ BECKWITH, *supra* note 35, at 66.

¹⁰⁹ Lugosi, *supra* note 104, at 123 (quoting Angelo Serra and Robert Colombo, *Identity and Status of the Human Embryo: The Contribution of Biology*, in IDENTITY AND STATUTE OF HUMAN EMBRYO: PROCEEDINGS OF THIRD ASSEMBLY OF THE PONTIFICAL ACADEMY FOR LIFE 128, 159 (Juan De Dios Vial Correa & Elio Sgreccia eds., 1998)).

¹¹⁰ BECKWITH, *supra* note 35, at 67; LAURENCE H. TRIBE, ABORTION: THE CLASH OF THE ABSOLUTES 117 (1990).

¹¹¹ BECKWITH, *supra* note 35, at 67–68; TRIBE, *supra* note 110, at 117.

The evidence of human life continues to mount in the early stages of development. Within five or six days of conception, the child has entered the “blastocyst” stage in which she has fully unique human DNA.¹¹² A study of the blastocyst stage shows that the cells are capable of becoming only the specialized cells in the human body.¹¹³ While it may be tempting to simply brush off the presence of human DNA as merely “potential life,” biologists have refuted this position.¹¹⁴ Instead, the modern understanding of biologists is that the discovery of human DNA after just a few days irrefutably demonstrates that the child, no matter how small, “is *human life*; it is not potential life or potentially human life.”¹¹⁵

Thus, “[i]t is scientifically correct to say that an individual human life begins at conception, when egg and sperm join.”¹¹⁶ This is no theory: scientific proof of a separate embryonic personhood is an indisputable and fundamental truth—a human being exists from conception.¹¹⁷ The official labels of “blastocyst” or “fetus” matter not—the terms are descriptive of a stage of human life development, similar to the labels of “infant” or “adolescent.”

Science can only go so far in the argument for constitutional protection for the unborn. While it can establish *humanity*, the law itself must recognize a *personhood* deserving of protection. As will be shown next, an increase in scientific biological knowledge of the development of the unborn has historically led to more and more restrictive abortion prohibitions.¹¹⁸

C. Societal Recognitions of Fetal Personhood

Historically, common law explicitly recognized that the killing of an unborn child was homicide.¹¹⁹ The crime was, however, generally placed within the realm of manslaughter rather than of murder.¹²⁰ The reason for this downgrading is that prior to modern scientific developments, the popular theory

¹¹² Stenger, *supra* note 4, at 37.

¹¹³ James A. Thomson, *Human Embryonic Stem Cells*, in *THE HUMAN EMBRYONIC STEM CELL DEBATE: SCIENCE, ETHICS, AND PUBLIC POLICY* 15, 17 (Suzanne Holland et al. eds., 2001).

¹¹⁴ Stenger, *supra* note 4, at 38.

¹¹⁵ *Id.* (emphasis added). This is the contemporary accepted understanding of biologists. *Id.*

¹¹⁶ BECKWITH, *supra* note 35, at 68.

¹¹⁷ Lugosi, *supra* note 104, at 123 (“Human embryology is so advanced there is no doubt that a new human being is created at the time of conception.”).

¹¹⁸ BECKWITH, *supra* note 35, at 24.

¹¹⁹ Roger J. Magnuson & Joshua M. Lederman, *Aristotle, Abortion, and Fetal Rights*, 33 WM. MITCHELL L. REV. 767, 777 (2007).

¹²⁰ *Id.* at 778.

of quickening was applied.¹²¹ Contemporary laws have, however, reflected an expansion in the rights of the unborn, including the classification of fetal homicide as murder.¹²² If *Roe* was right, one might expect the opposite trend. The growing rate speaks for itself. In 2004, twenty-seven states deemed the killing of a fetus to be criminal homicide.¹²³ By the following year, that number had risen to thirty-one.¹²⁴ Today, thirty-six states have such laws.¹²⁵ At this rate, every state could have such a law in less than a decade. Such fetal endangerment prosecutions reflect a desire to protect the unborn.¹²⁶ This trend has manifested itself in both state and federal laws.

1. State Criminal Laws

After *Roe* held that a pre-viable unborn child is not a person under the Constitution, the opposite occurred in state criminal law.¹²⁷ In refusing to take guidance from the Supreme Court's viability standard for human recognition, twenty-four states have extended legal protection for the life of the unborn, regardless of the stage of pregnancy.¹²⁸ Arizona acted first, amending its criminal law provisions to include protection for the fetus without any viability requirements.¹²⁹ Twenty-three other states have since followed, the majority of which acting after *Casey's* re-opening of the door to the state's interest.¹³⁰ Thirteen more grant protection once the child has reached a specified stage of development of anywhere from seven weeks to viability.¹³¹

California's murder statute is a good example of side-stepping *Roe's* findings to recognize the killing of a fetal

¹²¹ *Id.* at 777. See *supra* Subsection B for an explanation of the quickening theory.

¹²² Lotierzo, *supra* note 3, at 285–86.

¹²³ Parker, *supra* note 83, at 1428.

¹²⁴ Lotierzo, *supra* note 3, at 284.

¹²⁵ Americans United for Life, *supra* note 82 (listing the thirty-six states with criminal laws prohibiting fetal homicide).

¹²⁶ TRIBE, *supra* note 110, at 235.

¹²⁷ Magnuson & Lederman, *supra* note 119, at 779.

¹²⁸ *Id.*

¹²⁹ ARIZ. REV. STAT. ANN. § 13-1103(A)(5), (B) (LexisNexis 2008). The statute deems manslaughter committed when a person causes “the death of an unborn child by any physical injury to the mother” during “any stage of its development.” *Id.* at (B). To comply with *Roe*, it makes exceptions for abortion procedures. *Id.* at (B)(1).

¹³⁰ In addition to Arizona, the other states granting fetal protection under criminal laws at all stages of the pregnancy include Idaho, Illinois, Kentucky, Louisiana, Minnesota, Mississippi, Missouri, Nebraska, North Dakota, Ohio, Oklahoma, Pennsylvania, South Dakota, Texas, Utah, Virginia, and Wisconsin. Lotierzo, *supra* note 3, at 284–85 n.52.

¹³¹ *Id.* These states include Arkansas, California, Florida, Georgia, Indiana, Massachusetts, Michigan, Nevada, New York, Rhode Island, South Carolina, Tennessee and Washington. *Id.*

person.¹³² The State's fetal protection against criminal acts extends to the unborn child of at least seven weeks old.¹³³ The original statute had no reference to the unborn, but was added to correct a California Supreme Court ruling that did not allow a defendant to be found guilty of murder for killing a fetus.¹³⁴ While California's courts initially followed *Roe's* viability requirement and defined fetal murder as occurring only when the child would be able to survive on its own,¹³⁵ in 1994 this position was abandoned with a recognition that *Roe's* principals were "simply inconsistent" with fetal homicide statutes.¹³⁶

Such a trend is truly a modern movement with society "moving briskly toward the recognition of the personhood of the unborn."¹³⁷ As biologists established a better understanding of human science, it is no wonder that criminal liability has followed suit.¹³⁸ The states' respect for the humanity of the fetus has transcended both criminal and tort law: thirty states have laws in place that will immediately make abortion illegal should the Court reverse *Roe*.¹³⁹

¹³² CAL. PENAL CODE § 187 (West 2008).

¹³³ *People v. Davis*, 872 P.2d 591, 599 (Cal. 1994).

¹³⁴ Monica Mendes, *A Low Threshold of Guilt: Interpreting California's Fetal Murder Statute in People v. Taylor*, 39 LOY. L.A. L. REV. 1447, 1449 (2006).

¹³⁵ Katharine B. Folger, *When Does Life Begin . . . or End? The California Supreme Court Redefines Fetal Murder in People v. Davis*, 29 U.S.F. L. REV. 237, 247 (1994); Mendes, *supra* note 134, at 1449–50.

¹³⁶ *Davis*, 872 P.2d at 597; Mendes, *supra* note 134, at 1449, 1451.

¹³⁷ Magnuson & Lederman, *supra* note 119, at 777. Some commentators assert that fetal homicide laws are in harmony with abortion because they merely preserve the woman's exclusive choice to continue with or terminate her pregnancy. See, e.g., Carolyn B. Ramsey, *Restructuring the Debate over Fetal Homicide Laws*, 67 OHIO ST. L.J. 721, 725 (2006); cf. Mary Beth Hickcox-Howard, *The Case for Pro-Choice Participation in Drafting Fetal Homicide Laws*, 17 TEX. J. WOMEN & L. 317, 319–22 (2008) (explaining that fetal homicide laws are flawed and need to be re-drafted to avoid conflicts with abortion rights). But this Comment focuses on the significance of the status these laws give to the unborn child: homicide is the killing of a *person*. If the states wanted to preserve the non-personhood of a fetus, the offense would have been no more than criminal assault or the like. In fact, pushing for this lower-level type of crime as a substitute is the very thing that Pro-Choice groups often lobby for when opposing the passage of a new fetal homicide law. *Id.* at 319. Thus, fetal homicide laws are crucial because what follows is the reasoning that since homicide is the killing of a *person*, the unborn child should always be considered a *person* under the law.

¹³⁸ It should be noted that tort law, in addition to criminal law, has increasingly recognized fetal personhood. This is evidenced by the disregarding of the traditional "born alive" rule, which allowed for civil remedies only when the child was born alive and could thus bring the action herself. Magnuson & Lederman, *supra* note 119, at 778. Most states have abandoned this rule because it made tortfeasors liable only if the child dies rather than if the child was only injured by the same act. Beth Driscoll Osowski, *The Need for Logic and Consistency in Fetal Rights*, 68 N.D. L. REV. 171, 175 (1992). Instead, courts have recognized that there is a "trend in state courts toward greater legal rights for the unborn." *Id.* at 183.

¹³⁹ Associated Press, *Many States Would Ban Abortion if Roe Overturned*, CHARLESTON DAILY MAIL, Oct. 6, 2004, at 11A.

2. Federal Laws Following Suit

Despite failures to legislatively overturn *Roe*,¹⁴⁰ Congress successfully passed the Partial-Birth Abortion Ban Act of 2003 (“the Ban”).¹⁴¹ The Ban was passed in response to *Stenberg v. Carhart*’s striking down of Nebraska’s ban on the D&X procedure.¹⁴² The Ban criminalized abortions that deliberately and intentionally begin to deliver the child until a significant portion of the live infant is born and then “performing an overt act that the [physician] knows will kill the partially delivered living fetus.”¹⁴³ Where the Court in *Stenberg* had declined to address the live birth issue, Congress recognized that to allow such a procedure would be to confuse the “medical, legal, and ethical duties of physicians to preserve and promote life, as the physician acts directly against the physical life of a child, whom he or she had just delivered, all but the head, out of the womb, in order to end that life.”¹⁴⁴

In calling the procedure “gruesome and inhumane,” Congress explained that a purpose for the Ban was to protect unborn children, as future citizens, from experiencing “the pain associated with piercing his or her skull and sucking out his or her brain.”¹⁴⁵ Here, there is a congressional admission that the fetus is a citizen in need of protection, that having “potential life” and being a “future citizen” means that one is entitled to current rights. Also significant is that Congress outlawed the procedure due to the pain inflicted on the unborn child. If pain is the threshold for human life and protection, the fetus would be a human at least at the point in which she has sensory abilities.¹⁴⁶

The second major federal fetal rights law is the Unborn Victims of Violence Act of 2004 (“the Act”).¹⁴⁷ Also known as “Laci and Conner’s Law,”¹⁴⁸ the Act makes it a separate federal

¹⁴⁰ Congress’s attempts to overturn *Roe* through a constitutional amendment are discussed *supra* note 87 and accompanying text.

¹⁴¹ 18 U.S.C. § 1531 (2006).

¹⁴² *Stenberg v. Carhart*, 530 U.S. 914, 938–40 (2000).

¹⁴³ 18 U.S.C. § 1531(b)(1)(A) (2006); Lotierzo, *supra* note 3, at 279.

¹⁴⁴ Partial-Birth Abortion Ban Act of 2003, Pub. L. No. 108–105, §§ 2(1), 2(14)(J), 117 Stat. 1201, 1206 (2003).

¹⁴⁵ *Id.* at 2(14)(L), (M).

¹⁴⁶ Recall from *supra* Subsection B that an unborn child can feel pain even in the first trimester.

¹⁴⁷ 18 U.S.C. § 1841 (2006).

¹⁴⁸ The law is named after Laci Peterson and her unborn son, Conner, the victims in a 2003 California case in which a husband was convicted of killing his eight-months pregnant wife. Scott Peterson was convicted of double-murder and given the death penalty. Though the Act had been pending in Congress, the highly-publicized case gave it the significant support necessary to pass. Magnuson & Lederman, *supra* note 119, at 780–81.

crime to cause death or bodily harm to an unborn child.¹⁴⁹ Thus, the murder of a pregnant woman constitutes a double homicide. Moreover, the law recognizes two victims, even if the mother survives the offense.¹⁵⁰ Significant to the fetal rights movement is the fact that the Act defines “a child in utero” as “a member of the species homo sapiens, at *any stage of development*, who is carried in the womb.”¹⁵¹ This gives the unborn child protection under federal law as a separate person from the moment of known pregnancy—viable or not. This “advances the theory that unborn children should be afforded legal rights and protections under the Constitution.”¹⁵²

It is not surprising that scholars immediately noted the conflict between this legislation and the reasoning in abortion jurisprudence.¹⁵³ And yet, the Act reflects the social climate of the country: the legislative history notes that eighty-four percent of Americans believe that separate homicide charges are necessary for an unborn child.¹⁵⁴ Congress cited its purpose for the Act was to “respond to [the] overwhelming desire of the American public to provide, under Federal law, that an individual who injures or kills an unborn child . . . will be charged with a separate offense.”¹⁵⁵

D. The Time is Right to Correct the Wrong

Science and legislatures are not the only entities recognizing the personhood of the fetus: the Court is getting close as well. As notable constitutional law scholar, Erwin Chemerinsky, has noted, “[t]here is no area of constitutional law that cannot be changed by one or two appointments to the [C]ourt.”¹⁵⁶ The anti-*Roe* trend was certainly accelerated by Justices Brennan and Marshall resigning and being replaced by Justices Souter and

¹⁴⁹ 18 U.S.C. § 1841(a)(1) (2006); Lotierzo, *supra* note 3, at 281–82. The Act only applies to harm inflicted to the unborn child during the commission of a federal crime; it exempts the practice of abortion from liability. *Id.* at 283.

¹⁵⁰ Lotierzo, *supra* note 3, at 303.

¹⁵¹ 18 U.S.C. § 1841(d) (2006) (emphasis added).

¹⁵² Lotierzo, *supra* note 3, at 303.

¹⁵³ See, e.g., Amanda K. Bruchs, *Clash of Competing Interests: Can the Unborn Victims of Violence Act and Over Thirty Years of Settled Abortion Law Co-Exist Peacefully?*, 55 SYRACUSE L. REV. 133, 156–57 (2004) (proposing a modification to the legislation that would change “child in utero” to be only viable fetuses and make the offense only an enhanced crime against the woman rather than a separate crime).

¹⁵⁴ Lotierzo, *supra* note 3, at 304 (citing to H.R. REP. NO. 108-420(I), at 5). The Congressional findings were based on a poll by Princeton Survey Research Associates and indicated that more than half of the public supported separate charges regardless of how far along the pregnancy was. *Id.* at n.249.

¹⁵⁵ *Id.* at 304 (quoting H.R. REP. NO. 108-420(I), at 5).

¹⁵⁶ Jill Schachner Chanen, *Senate Is Hurdle for Next Supreme Court Pick*, 3 ABA J. E-REPORT, Oct. 29, 2004.

Thomas.¹⁵⁷ Three of the Justices appointed by Presidents Reagan and George H.W. Bush made up the plurality opinion in *Casey*.¹⁵⁸ Though the Court upheld abortion rights, this case marked a significant turning point as the Court upheld an abortion restriction similar to those it had struck down less than a decade earlier.¹⁵⁹ More recently, the additions of Chief Justice Roberts and Justice Alito seem to have significantly quickened the pace towards fetal rights, with Justice Alito supplying the crucial fifth vote in *Gonzales v. Carhart*¹⁶⁰ and Chief Justice Roberts procuring the first unanimous vote in the Court's abortion jurisprudence in *Ayotte v. Planned Parenthood*.¹⁶¹ Indeed, with these appointments, the Court is currently only one vote shy of completely overturning *Roe*.¹⁶²

With the Supreme Court "in flux,"¹⁶³ a reconsideration of *Roe* is an unavoidable possibility. When that day comes, the Court will have to take into account all of the current scientific and sociological developments.¹⁶⁴ It would not be the first time the Court has reversed itself to correct a false assumption.¹⁶⁵ This is especially true when one is reminded that at one time, women

¹⁵⁷ MARK A. GRABER, *RETHINKING ABORTION: EQUAL CHOICE, THE CONSTITUTION, AND REPRODUCTIVE POLITICS* 127 (1996). Justices Brennan and Marshall were part of the *Roe* and *Doe* majorities, while Justices Souter and Thomas lean towards pro-life. *Id.*

¹⁵⁸ BECKWITH, *supra* note 35, at 31–32; GRABER, *supra* note 157, at 127.

¹⁵⁹ GRABER, *supra* note 157, at 127. The law upheld in *Casey*, as discussed *supra* Part I.C, required informed consent and a mandatory waiting period; such laws had been previously struck down by *City of Akron* and *Thornburgh*. See *supra* note 38.

¹⁶⁰ *Gonzales v. Carhart*, 550 U.S. 124, 130 (2007). See also Christopher E. Smith, et al., *The Roberts Court and Criminal Justice at the Dawn of the 2008 Term*, 3 CHARLESTON L. REV. 265, 287 (2009).

¹⁶¹ *Ayotte v. Planned Parenthood of Northern New England*, 546 U.S. 320, 321 (2006); John F. Basaik, Jr., *The Roberts Court and the Future of Substantive Due Process: the Demise of "Split-the-Difference" Jurisprudence?*, 28 WHITTIER L. REV. 861, 901–902 (2007). *Ayotte* is significant because the Court unanimously declined the chance to strike down a state law that lacked a health restriction, and instead decided the case on the basis of remedy. *Id.*

¹⁶² Scott A. Moss, *The Courts Under President Obama*, 86 DENV. U. L. REV. 727, 731–32 (2009) (pointing out that because of the Justices likely to retire within the next four years, the most that President Obama's appointments will do is preserve the status quo).

¹⁶³ Megan Fitzpatrick, *Fetal Personhood After the Unborn Victims of Violence Act*, 58 RUTGERS L. REV. 553, 566 (2006).

¹⁶⁴ It is true that no Supreme Court Justice has officially taken the position that the fetus is a constitutional "person." TRIBE, *supra* note 110, at 125. However, this Comment takes the position that the day is inevitably coming when the Court must acknowledge the science of humanity and the recognition of fetal personhood as a genuine third party in the abortion debate. As this section has shown, the Court itself is becoming more and more open to the idea of admitting its faulty reasoning and correcting its mistake.

¹⁶⁵ See, e.g., *Brown v. Board of Education*, 349 U.S. 294, 301 (1955) (overturning a progeny of earlier rulings dating back to *Plessy v. Ferguson* in 1896 which had allowed for the segregation of public schools due to a false assumption about the extent of equal protection). At least one scholar has suggested that *Roe* should be overturned on equal protection grounds just as *Brown* was. Lugosi, *supra* note 104, at 120, 287.

and slaves were also considered less than a “person” for the purpose of constitutional protection.¹⁶⁶

If the time is right, then this begs the question—now what? If *Roe* is reversed and constitutional protection is granted to unborn children, how does this affect a woman’s right to terminate an unwanted pregnancy? The next two sections address the conflict of interests and explore compromises before proposing a solution that is consistent with constitutional jurisprudence.

III. THE PROBLEM: RIGHT TO LIFE V. RIGHT TO PRIVACY

A. Stating the Obvious: Recognizing the Conflicting Interests

The law of non-contradictions necessitates a conclusion that “an unborn child cannot be a person and a non-person at the same time and in the same respect.”¹⁶⁷ Recognizing that the unborn child is a person and yet unequal under the law is simply unjust.¹⁶⁸ Instead, an unborn child who is granted recognition as a person is entitled to legal protection.¹⁶⁹ Thus, if the fetus is finally given a classification as a constitutional person, then any legislature that permits abortion is allowing others to deprive the unborn child of life without due process of law and without equal protection—a completely unconstitutional allowance.¹⁷⁰

While the right to privacy has long been considered a sacred right,¹⁷¹ it is not an absolute right and can be limited by the conflicting right of the unborn.¹⁷² No matter where one stands on the abortion issue, simple logic dictates that it is wrong to continue to allow the absolute right of abortion in light of the “accelerating of protections for the rights of the unborn.”¹⁷³ Rather, there is a conflicting interest between the right to an abortion and the right to life that must be resolved.

¹⁶⁶ TRIBE, *supra* note 110, at 119–20.

¹⁶⁷ Magnuson & Lederman, *supra* note 119, at 786.

¹⁶⁸ See Lugosi, *supra* note 24, at 438.

¹⁶⁹ See Heather M. White, *Unborn Child: Can You Be Protected?*, 22 U. RICH L. REV. 285, 288 (1988). It has also been pointed out that when the fetus is recognized as a human being, the furthering of the unborn child’s rights is necessary in protecting the rights of born children. Charlene Quint Kalebic, *Children, the Unprotected Minority: A Call for the Reexamination of Children’s Rights in Light of Stenberg v. Carhart*, 15 REGENT U. L. REV. 223, 228 (2003).

¹⁷⁰ TRIBE, *supra* note 110, at 115.

¹⁷¹ *Union Pac. Ry. v. Botsford*, 141 U.S. 250, 251 (1891) (noting that the right of privacy is more “carefully guarded” than any other right).

¹⁷² White, *supra* note 169, at 292.

¹⁷³ Magnuson & Lederman, *supra* note 119, at 785.

B. Preliminary Solutions: Why They are Not Enough

While many scholars have come to a consensus that the unborn child has a right to life, few have taken it a step further and offered a solution to the conflicting interests.¹⁷⁴ Some have recognized that the unborn child is entitled to protection from tortious acts under the law and yet refrain from extending this conclusion to protect the child from death by abortion.¹⁷⁵ However, a law that protects the fetus from abuse or neglect¹⁷⁶ would be inconsistent if it then allows the even more grievous tortious act of inflicting involuntary death.

Other scholars argue that the conflicting interests should be solved by leaving it up to the states to decide whether the fetus is a person.¹⁷⁷ However, this is not *solving* the underlying issue—it is merely delegating the decision.¹⁷⁸ Moreover, even if it were left to the states, more issues would arise as a result of inconsistent laws. Consider the obvious problem that will arise when citizens of an anti-abortion state cross into a pro-abortion state in order to get the procedure.¹⁷⁹ If the fetus is a protected citizen of one state, would this status change simply because her mother traveled into another state?¹⁸⁰ What if the child is conceived in a state that deems the child a person, yet his mother is a resident of a state (or later moves to a state) where no personhood is recognized until birth? Federalism simply cannot be relied on in this context because the truth as to the beginning of human life is not discretionary. History dictates that personhood is not something that can change based on geographic location: the Thirteenth Amendment was passed precisely to remove the state's power of deeming a class of citizens non-persons simply because of their race.¹⁸¹ In the same way, there should be no

¹⁷⁴ See, e.g., *id.* at 785–86 (pointing out the illogic of abortion rights but refraining from calling for a solution); Lugosi, *supra* note 24, at 437 (arguing that the fetus has constitutional protection but leaves it up to the states to pass laws that outlaw abortion to spur a reversal of *Roe*).

¹⁷⁵ See, e.g., White, *supra* note 169, at 288 (arguing that the fetus may have protection under the law from abuse and neglect, as long as this does not “overste[p] the parameters of the Supreme Court’s decision in *Roe v. Wade*”).

¹⁷⁶ *Id.*

¹⁷⁷ TRIBE, *supra* note 110, at 126.

¹⁷⁸ Furthermore, *Roe* specifically stated that should personhood be established, then the unborn child’s right to life would be protected by the Fourteenth Amendment, thus preventing *any* state from making the decision to allow abortions. *Roe v. Wade*, 410 U.S. 113, 156–57 (1973). For a further discussion, see Part IV, *infra*.

¹⁷⁹ *Id.* at 127.

¹⁸⁰ *Id.*

¹⁸¹ See Alexander Tsesis, *Principled Governance: the American Creed and Congressional Authority*, 41 CONN. L. REV. 679, 706–708 (2009) (explaining that the legislative intent behind the Thirteenth Amendment was to guarantee that the states would not be able to undermine equality under the law).

state-by-state determination of personhood based on the arbitrary factors of age, size and economic status. There needs to be consistency among all of the states to prevent such disputes, as they subject the existence of human life to varying degrees of interpretation when it is instead a question of fact.

At least one scholar has suggested that deeming a fetus a person is problematic because it treats women as incubators, inferring that this alone should preserve the right to an abortion.¹⁸² This, however, has nothing to do with addressing the conflict between the interests. Even if the woman is an “incubator,” this argument does not adequately address the right to life versus the right to be free from being an “incubator.”¹⁸³ Instead, it is merely a restatement of the conflict between a woman’s liberty interest of being free from carrying an unwanted child and that child’s interest in living. But hardship on a woman’s liberty interest has no bearing on the actual physical personhood of the child she is carrying; while it may make some more comfortable with the idea of abortion, it does not diminish or resolve the underlying issue.

Another scholar has suggested that the conflict of interests is three-fold: state, child and mother.¹⁸⁴ To resolve these interests, a balancing test is proposed in which the “interests of unborn children should be weighed alongside those of pregnant women and the state in legislative enactments and judicial review of abortion laws.”¹⁸⁵ Because this scholar concludes that this would be a case-by-case analysis, there is no suggestion of how to apply this balancing test.¹⁸⁶ A case-by-case analysis is unworkable. In every case, the fundamental interests are the same: interest in protecting future citizens versus interest in saving one’s life versus interest in obtaining an abortion. Without offering a solution of how to weigh the competing interests, all that is present is, again, an acknowledgement of the problem.

Thus, scholars have generally left an actual solution still wanting. Below is a proposal that is more consistent in applicability and resolving of the underlying issue.

¹⁸² TRIBE, *supra* note 110, at 130.

¹⁸³ Moreover, it is worth noting that in most cases, the woman either consciously or negligently became this “incubator.”

¹⁸⁴ Lotierzo, *supra* note 3, at 311–12.

¹⁸⁵ *Id.* at 281.

¹⁸⁶ *Id.* at 312.

IV. SOLUTION: RIGHT TO LIFE MUST PREVAIL

The solution to the conflicting interests comes from what one might find an unlikely source: *Roe* itself. The opinion suggests that the right to life would prevail over the right to privacy.¹⁸⁷ Recall *Roe*'s own admission that once "personhood is established, the [case for abortion rights] of course, collapses, for the fetus' right to life would then be guaranteed specifically by the [Fourteenth] Amendment."¹⁸⁸ Here, Justice Blackmun does not state that when fetal personhood is accepted, the right to life becomes inherent only at a certain point in the pregnancy. Nor does he state that when fetal personhood is accepted, a balancing test must be employed. To the contrary, the fetus' life is constitutionally protected and the entire case for abortion rights *collapses*. This means that abortion's right of privacy cannot be sustained against the right of life. Simply put, the right of a human being to live, in any analysis, must always trump the liberty right to terminate that life.¹⁸⁹

Merely because the unborn child lacks the capacity to assert her rights and equal protection under the law is not an excuse to ignore and deny those rights.¹⁹⁰ Clearly, the unborn child would benefit from exercising her Fourteenth Amendment right to live, her right to prevent involuntary death.¹⁹¹ The unborn child has the right to choose life for herself. The fact that she cannot express it for herself does not diminish this right. Normally, the child's guardian would represent this right, but here this is the very person with the conflicting interest. Thus, it is necessary for the state to step in and protect the child's rights.¹⁹²

Abortion supporters will likely rebut this proposition by asserting that the right of privacy gives the woman autonomy over her own body without government interference; that a woman will be subjected as an "incubator" if not given the choice

¹⁸⁷ *Roe v. Wade*, 410 U.S. 113, 156–57 (1973); Lotierzo, *supra* note 3, at 300.

¹⁸⁸ *Roe*, 410 U.S. at 156–57.

¹⁸⁹ It is important to note here that abortions necessary to save the mother's life are purposely left out of this analysis. This is because when the mother's life is actually threatened, the conflicting interests change to the right to life versus the right to life, rather than the right to *liberty* versus the right to life. This Comment focuses on the latter because abortion jurisprudence has allowed for abortions essentially on demand, regardless of the condition of the mother. See *supra* note 6.

¹⁹⁰ Rhonheimer, *supra* note 9, at 161.

¹⁹¹ *Id.*

¹⁹² This reasoning has been applied in European countries such as Germany that protect the life of the unborn except in situations where the mother's life is truly at stake. The German court not only held that criminalizing abortion was permitted, but constitutionally required in order to protect the lives of the unborn. Stephen Gardbaum, *State and Comparative Constitutional Law Perspectives on a Possible Post-Roe World*, 51 ST. LOUIS U. L.J. 685, 691 (2007).

whether to carry her child to term.¹⁹³ The proposed solution is not meant to discriminate against women's rights, but rather to promote the rights of the unborn human being who has no choice but to be carried by her mother for the first nine months of her existence.

Moreover, the right of privacy assumes that the actions one is taking affect only oneself. As soon as the actions have a negative impact on another person, the actions transcend from private to social.¹⁹⁴ The unborn child is only lacking in independence because her age requires her mother's womb to assure survival. This is no different from a post-birth child who depends on her parent's provisions of food and clothing to survive. For that matter, it is no different from any disabled or incapacitated adult.

A fundamental truth of human rights is that the value of human life cannot be measured by age, size or desirability. To deny an unborn child equal protection as a person simply because she is small and young is scientifically incorrect. To permit a mother to end the life of her developing baby simply because she does not want the inconvenience of a pregnancy or parenthood is constitutionally wrong. *Roe* itself called for the recognition of fetal personhood in light of an expansion of medical and social developments,¹⁹⁵ and once recognized, this person's life must be protected above another's lesser right. While it may not be the easiest solution to swallow and would take a humble Court ready to correct more than three decades of embarrassing missteps, it is the right thing to do.

CONCLUSION

Immediate and constant criticisms of *Roe* and its progeny, together with the Court's own admission that the issue must be readdressed when more information is known, make reconsideration inevitable. The Court cannot forever hide from the scientific truths that have been discovered, nor the trends in legislation. Thus, the recognition of fetal personhood is no longer an "if" but a "when." And when that day comes, the Court must

¹⁹³ Some scholars have anticipated that the time of recognizing fetal personhood is drawing near and have made this argument. See, e.g., *TRIBE*, *supra* note 110, at 130.

¹⁹⁴ See, e.g., *Lugosi*, *supra* note 24, at 438 ("[T]here is a limit to personal liberty when its exercise is incompatible with not just the liberty of another, but the life of another person.").

¹⁹⁵ Recall that Justice Blackmun asserted that the reason for failing to determine fetal personhood was because "those trained in the respective disciplines of medicine, philosophy, and theology are unable to arrive at any consensus." *Roe v. Wade*, 410 U.S. 113, 159 (1973); see *supra* Part I.B. Now that there is such a consensus, the Court must be compelled to recognize personhood as well. See *supra* Part II.

2009]

Pro [Whose?] Choice

247

render a solution to the competing interests between mother and child. To stay consistent with both constitutional and human rights principals, the Court should stay true to its admission in *Roe* and let the right to abortions “collapse.”