HARVARD
Journal on Legislation

Volume 41, Number 1 Winter, 2004

Essays

1 Challenges for Democracy and Trade: The Case of the United States
Chantal Thomas

29 Restoring Investor Trust in Auditing Standards and Accounting Principles
Melvyn I. Weiss & Elizabeth A. Berney

Article & Comments

59 Accounting for Social Security and Its Reform
Howell E. Jackson

161 Liabilities, Debts, Revenues, and Expenditures: Accounting for the Actuarial Balance of Social Security
Robert L. Clark

173 Accrual Accounting for Social Security
Peter A. Diamond & Peter R. Orszag

187 Accounting for the Federal Budget and Its Reform
Elizabeth Garrett

199 The Virtues of Moving from Cash to Accrual Accounting for Social Security
Robert C. Pozen

209 Accrual Accounting and the Fiscal Gap
Daniel N. Shaviro

215 The Inadequacies of Accrual Accounting for Social Security
Kent Smetters

221 Reply
Howell E. Jackson

Articles

227 Severability, Inseverability, and the Rule of Law
Michael D. Shumsky

281 Resurrecting RICO: Removing Immunity for White-Collar Crime
Michael Goldsmith

Recent Developments

319 Payment for College Football Players in Nebraska
337 Universal National Service Act

Copyright © 2004 by the
President and Fellows of Harvard College

ESSAY

CHALLENGES FOR DEMOCRACY AND TRADE: THE CASE OF THE UNITED STATES

CHANTAL THOMAS*

Predominant political theory holds that legislators are protectionist regarding international trade because susceptibility to minority interest groups leads them to vote in ways that protect domestic industries at the expense of free trade. Because free trade is widely regarded as beneficial to the majority, the protectionist tendency of the legislature is believed to be a disservice to most Americans. These two theories have led to policies that restrict the role of the legislature in the formulation of trade policy, specifically, the creation of the fast track framework for trade policy legislation that exists today. This Essay challenges these two theories, offering evidence that fast track legislation and the theories supporting a reduced role for the legislature in trade policy may be based on widely held, but unjustified, beliefs.

President Bush is turning up the heat on Congress to pass his top-priority trade bill.

> [B]ehind the political maneuvering lie ideological divisions on the Hill over how to handle free trade . . . . Fast-track negotiating authority—under which Congress agrees to consider any new trade agreement on an up-or-down vote with no amendments—requires faith on the part of lawmakers that the president will protect their interests, and those of their constituents.[1]

Two assumptions justify the constraints placed on Congress, and the authority given to the President, in the framework of contemporary trade legislation in the United States.[2] First, legislatures are thought to be pro-

---

* Professor of Law, Fordham University School of Law. B.A., McGill University, 1992; J.D., Harvard Law School, 1995. Thanks to Karen Engle, Katherine Franke, Abner Greene, Rob Howse, Janet Halley, Gideon Parchomovsky, Peter Siegelman, Linda Sugin, Greg Thomson, and Benjamin Zipursky.

[1] Michael M. Phillips, *Bush Steps In to Break Logjam Stalling 'Fast Track' Trade Bill*, WALL ST. J., July 19, 2002, at A4.

[2] Under the "fast track" provisions of the Trade Act of 1974, Congress explicitly delegates to the President (and the United States Trade Representative) the authority to negotiate international trade agreements and devise implementing legislation for those agreements, and disallows congressional amendment of the negotiated agreements when they are presented for legislative approval. *See* Trade Act of 1974, Pub. L. No. 93-618, 88 Stat. 1978 (codified as amended at 19 U.S.C. §§ 2101–2495 (2000)). Under the fast track procedure, trade legislation occurs in two phases. First, Congress statutorily authorizes the President to negotiate an international trade agreement and articulates what the President's general negotiating objectives should be. 19 U.S.C. § 2112 (2000). Congress also authorizes the President to proclaim into law any tariff reductions negotiated as part of the agreement. *Id.*

tectionist as a consequence of the disproportionate influence exerted on them by minority interest groups.[3] Second, trade liberalization is assumed to be a better policy for the majority than trade protection.[4] As a consequence of both of these assumptions, constraining the legislative process in order to liberalize trade barriers through international agreement is deemed legitimate and desirable.

Employing recent empirical and policy work and well-established theoretical concepts, this Essay questions the validity of each of these assumptions. While a kernel of truth exists at the core of both, important qualifications and complexities in each concept are too often overlooked, with potentially dangerous consequences for the democratic legitimacy of trade policy.

First, as emerging political science and economics literature indicates, the account of legislatures as beholden to protectionist interests is outdated, failing to take into account shifts that complicate the traditional analysis.[5] Minority interest groups do continue to exert considerable influence over trade policy—but these interests often act in favor of protection as well as against protection, and they can influence not only the legislature but also the executive.[6] Indeed, anti-protection, pro-liberalization interests now generally seem to dominate the trade legislation process.[7] Trade legislation has become, therefore, an aggregation of this overarching directive and smaller concessions to pro-protection interests.

Neglect of these changed circumstances would be benign if, as is thought to be the case, trade liberalization always benefited the majority. If this were true, then the dominance of the anti-protection interests would happily coincide with the interests of the majority. But the account of trade liberalization as an unqualifiedly better policy for the majority neglects to take into account issues that may compete with the maximization of aggregate national wealth, such as wealth distribution, job security, and environmental conservation.[8] A closer look at the implications of the well-established economic concepts such as the principle of comparative advantage, the factor-price equilibrium theorem, and the Kaldor-Hicks definition of economic efficiency demonstrate that the potential benefits of trade liberalization for the majority are qualified.[9] Because of these qualifications, the interests of pro-liberalization minorities cannot

---

Second, Congress considers, within an expedited time frame and without the option of amendment, whether to accept or reject legislation proposed by the President to implement these agreements. 19 U.S.C. §§ 2191–2193 (2000). This second step applies only to "nontariff" concessions. *Id.*

[3] *See infra* Part I.

[4] *See infra* Part I.

[5] *See infra* Part II.A.

[6] *See infra* Part II.A.

[7] *See infra* Part II.A.

[8] *See infra* Part II.B.

[9] *See infra* Part II.B.

be assumed to coincide with the interests of the majority. The current framework for trade legislation may therefore need reevaluation in terms of its democratic legitimacy.

This Essay shows that much remains to be done before we can come up with a legislative framework for making trade policy that addresses the true complexity of the institutional and substantive challenges at stake. This Essay ends by contemplating what this discussion means for the larger questions of democracy in a globalized era, the legitimacy of international organizations, and the question of distributive justice (and democratic legitimacy) not only within but also among nations.

Part I introduces the conventional wisdom on trade policy that supports the procedural assumption that the legislature is institutionally corrupt and beholden to protectionist special interests, and the substantive assumption that trade liberalization benefits the majority. Part II raises questions about each of these assumptions. First, Part II challenges the assumption that legislatures are beholden to protectionist interests by arguing that protectionist interests no longer dominate the legislatures, that political ideology has shifted away from protectionism, and that the executive is, like Congress, influenced by minority interests. Part II next questions the assumption that trade liberalization always benefits the majority through close examination of tenets of economic policy and wealth measurements. Part III ends with a consideration of the implications of these questions, touching on both domestic and international issues. No easy answers present themselves, but this Essay urges a closer look at the questions.

I. An Introduction to the Conventional Wisdom

The premise that legislatures are necessarily protectionist finds its theoretical support in public choice literature on law and economics,[10] and especially in the idea of capture.[11] The theory of capture holds that legislators often choose policies that cater to minority interest groups over policies that benefit the majorities in their constituencies.[12] Legisla-

---

[10] For surveys of public choice literature, see generally Daniel A. Farber & Philip P. Frickey, Law & Public Choice (1991); Dennis C. Mueller, Public Choice II (1989).

[11] There are other dimensions of public choice theory that are not discussed here, such as "Arrow's Theorem," which concerns decision-making by legislatures when multiple alternatives are presented. *See generally* Kenneth Arrow, Social Choice and Individual Values (1951); *see also* Farber & Frickey, *supra* note 10, at 38–62.

[12] For a discussion of capture theory in legal scholarship, see generally Maxwell L. Stearns, Public Choice and Public Law: Readings and Commentary (1997); Daniel A. Farber & Philip P. Frickey, *The Jurisprudence of Public Choice*, 65 Tex. L. Rev. 873 (1987); John Shepard Wiley, *A Capture Theory of Antitrust Federalism*, 99 Harv. L. Rev. 713 (1986). Many have criticized the theory of capture as overly simplistic. *See* Arthur Maass, Congress and the Common Good 1–3 (1983); Joseph P. Kalt & Mark A. Zupan, *Capture and Ideology in the Economic Theory of Politics*, 74 Am. Econ. Rev. 279 (1984).

The theory of capture challenges the model of the political process as a well-

tive susceptibility to capture by minority interests results from the "collective action problem" that majorities face: majorities, whose interests are often diffuse, are generally much less able to organize effectively than minority groups, whose interests are more concentrated.[13]

Trade policy serves as a canonical example of capture and collective action problems.[14] According to classical economic theory, the gains from trade exceed the losses,[15] but those gains are diffusely spread across the entire population, whereas the losses are concentrated on specific industries that cede market share to foreign competition.[16]

According to this explanation, while the maximization of consumption is the goal of free trade, it is also its weakness in the context of policy formation. Domestic producers who face losses as a result of import competition will come together and organize to pressure legislators for protective trade laws. Even though trade protection imposes costs on consumers that in the aggregate outweigh its benefits to producers, these costs are distributed very widely. The individual losers from trade pro-

---

functioning marketplace. That model, dubbed interest-group pluralism, is most widely associated with the early writings of Robert Dahl. *See generally* ROBERT A. DAHL, A PREFACE TO DEMOCRATIC THEORY (1956). According to interest-group pluralism, political decisions result from bargaining among interest groups, with lawmakers aggregating policy preferences in much the same way that economic markets ideally aggregate price preferences. *See id.* Public choice theory argues that the political marketplace is distorted.

[13] MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION: PUBLIC ACTION AND THE THEORY OF GROUPS 5–52 (1971). Olson summarizes this theory as the rule that "the larger the group the farther it will fall short of providing an optimal supply of a collective good, and very large groups normally will not, in the absence of coercion or separate, outside incentives, provide themselves with even minimal amounts of a collective good." *Id.* at 48.

[14] *See id.* at 5–52.

[15] The economic principle of comparative advantage holds that a nation benefits by opening its borders to foreign competition. Greater trade allows a nation to expand consumption possibilities beyond domestic production possibilities, allowing consumers to buy a wider variety of goods and services for lower prices. A canonical explanation of this principle is given in PAUL A. SAMUELSON & WILLIAM D. NORDHAUS, MACROECONOMICS 390–93 (16th ed. 1998). Gains from trade increase "real income" by allowing consumers to buy more with the same amount of money than they could before. *See id.* While those gains may be relatively modest on a per capita basis, they add up to a substantial benefit across the entire population. *See id.* (There are additional, longer-term gains from trade that take the form of increased efficiency as a result of the reallocation of domestic resources according to comparative advantage. *See id.*) For a discussion of the shortcomings of free trade theory that result from its focus on individuals-as-consumers, see *infra* Part II.B.

[16] The beneficiaries of trade liberalization are primarily consumers, while the costs of trade liberalization are primarily borne by import-sensitive producers. SAMUELSON & NORDHAUS, *supra* note 15, at 390–93. The principle of comparative advantage asserts that the losses imposed on consumers by tariffs—and presumably any other trade barrier—exceed the benefits to producers. It is interesting to note that the conclusions relating to tariffs are generalized to other forms of "trade barriers" with relatively little scrutiny in this treatment. *See id.* at 399 (asserting that the explanation applies "equally well to any other impediments to trade"). While the same principles can be relatively easily applied to import quotas, *see id.* at 391–92, applying the analysis to other forms of trade barriers is overly simplistic because these other trade barriers may generate benefits other than profits for protected producers. For example, a research subsidy or health regulation may generate benefits in addition to increased profits for producers.

tection will thus have much less incentive to organize against protective legislation than the losers from trade liberalization have to organize against liberalizing legislation.

Further, because trade policy affects a broad cross-section of economic activity,[17] protectionist pressures arising from various constituencies will generate "logrolling" among legislators, in which individual legislators exchange agreements to protect interests important to other legislators for others' agreements to protect interests important to them.[18] In sum, the theory of capture, as traditionally understood, predicts that trade policy controlled by legislatures will constitute the sum of distorted protections resulting from disproportionately influential producer groups.[19]

These theoretical conclusions are substantiated by historical incident, and particularly by the Tariff Act of 1930 (known as the Smoot-Hawley Act after its two main sponsors).[20] E. E. Schattschneider, in his study of the Smoot-Hawley Act (now cited as a "classic case study" of legislative dysfunction[21]), drew on evidence from congressional hearings on the Act to argue that "pressure politics" from interest groups had resulted in "dubious economic policy."[22] Indeed, the Smoot-Hawley Act is often said to have imposed the highest tariffs in U.S. history.[23] Coinciding with the Great Depression, the Smoot-Hawley Act also helped trigger retaliatory trade restrictions imposed by foreign trading partners of the United States.[24]

---

[17] Trade policy traditionally affected goods, but has more recently incorporated services and intellectual property rights. *See* General Agreement on Trade in Services, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1B, RESULTS OF THE URUGUAY ROUND OF MULTILATERAL TRADE NEGOTIATIONS: THE LEGAL TEXTS, 325, 33 I.L.M. 1168 (1994); Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, RESULTS OF THE URUGUAY ROUND OF MULTILATERAL TRADE NEGOTIATIONS: THE LEGAL TEXTS, 365, 33 I.L.M. 1197 (1994).

[18] Dennis Mueller's pithy description is, "You vote for my pet issue and I'll vote for yours." MUELLER, *supra* note 10, at 82.

[19] *See* Robert E. Hudec, *Comment*, *in* EFFICIENCY, EQUITY, LEGITIMACY: THE MULTILATERAL TRADING SYSTEM AT THE MILLENIUM (Roger B. Porter et al. eds., 2001).

[20] *See* Tariff Act of 1930 (Hawley-Smoot Tariff Act or Smoot-Hawley Act), ch. 497, 46 Stat. 590 (1930) (codified in scattered sections of 19 U.S.C.). The Act was sponsored by Senator Reed Smoot and Senator Joseph Roswell Hawley.

[21] FARBER & FRICKEY, *supra* note 10, at 17.

[22] *See* E. E. SCHATTSCHNEIDER, POLITICS, PRESSURES AND THE TARIFF: A STUDY OF FREE PRIVATE ENTERPRISE IN PRESSURE POLITICS, AS SHOWN IN THE 1929–1930 REVISION OF THE TARIFF 283–93 (1935). It is interesting to note, however, that Schattschneider himself was not an ardent free trade advocate, believing rather that "the benefits and costs of the protective tariff, viewed in their totality, are probably very nearly equal." *Id.* at 285.

[23] Harold H. Koh, *The Legal Markets of International Trade: A Perspective on the Proposed United States-Canada Free Trade Agreement*, 12 YALE J. INT'L. L. 193, 201 (1987). *But see* ALFRED E. ECKES, JR., OPENING AMERICA'S MARKET: U.S. FOREIGN TRADE POLICY SINCE 1776, at 106 (1995) and discussion *infra* note 45.

[24] *See* DOUGLAS A. IRWIN, FROM SMOOT-HAWLEY TO RECIPROCAL TRADE AGREEMENTS: CHANGING THE COURSE OF U.S. TRADE POLICY IN THE 1930S, at 16, 18–19 (Nat'l Bureau of Econ. Research, Working Paper No. 5895, 1997). *But see* ECKES, *supra* note 23, at 116–32.

It would be hard to overstate the importance of the Smoot-Hawley Act in shaping conventional wisdom on trade policy. Escalating worldwide protectionism provoked in part by the Act deepened and prolonged the economic depression.[25] Worldwide tariffs likely have kept consumer prices high, lowering real income, and prevented recovery among producers, because producers facing a contracted domestic market could no longer turn to foreign markets.[26] The ramifications of the global depression of the 1930s stretched far beyond the realm of fiscal misery: the depression contributed to the chaotic conditions in Weimar Germany that preceded the rise of Nazi fascism.[27]

Thus, the Smoot-Hawley Act serves as a stinging reminder of the depths of error to which legislatures are susceptible when in the throes of minority economic interests.[28] The Smoot-Hawley Act experience has served to reinforce the two lessons, institutional and substantive, that stand at the heart of the contemporary trade policy framework: legislatures are protectionist, and protectionism is disastrous trade policy. Very shortly after the passage of the Smoot-Hawley Act, the trade policy framework began to gravitate away from the legislature and toward the executive, resulting ultimately in the fast track framework that exists today.[29] Following World War II, the powerful nations joined together in an effort to foreclose any recurrence of the events of the 1930s.[30] Preventing uni-

---

[25] *See* Richard N. Cooper, *Trade Policy as Foreign Policy*, *in* U.S. TRADE POLICIES IN A CHANGING WORLD ECONOMY 291, 291 (1987). *But see* ECKES, *supra* note 23, at 139 (noting the lack of evidence that the Smoot-Hawley Act exacerbated the depression); Barry Eichengreen, *Did International Economic Forces Cause the Great Depression?*, 6 CONTEMP. POL'Y. ISSUES, Apr. 1998, at 90, 100.

[26] *See* IRWIN, supra note 24, at 22–33 (quoting President Franklin Delano Roosevelt's statement supporting trade liberalization in 1934 that "a full and permanent domestic recovery depends in part upon a revived and strengthened international trade").

[27] *See* JOHN JACKSON, THE WORLD TRADING SYSTEM: LAW AND POLICY OF INTERNATIONAL ECONOMIC RELATIONS 10 (1989) (noting the "thinking that pointed to the interwar economic problems as partial causes for the disastrous Second World War"); Cooper, *supra* note 25, at 291–92. *But see* ECKES, *supra* note 23, at 139 (noting that the available evidence does not confirm that "the 1930 tariff contributed significantly to the breakdown of international peace and the coming of World War II").

[28] An oft-quoted passage from hearings discussing the Reciprocal Trade Agreements Act of 1934 provides one example of this view:

> [O]ur experience in writing tariff legislation . . . has been discouraging. Trading between groups and sections is inevitable. Logrolling is inevitable, and in its most pernicious form. We do not write a national tariff law. We jam together, through various unholy alliances and combinations, a potpourri or hodgepodge of section and local tariff rates, which often add to our troubles and increase world misery.

78 CONG. REC. 10,379 (1934) (statement of Sen. Cooper), *quoted in* JOHN H. JACKSON ET AL., LEGAL PROBLEMS OF INTERNATIONAL ECONOMIC RELATIONS 139 (3d ed. 1995).

[29] *See* Hudec, *supra* note 19, at 297 ("As the story of the Smoot-Hawley reforms and the fast-track reforms show, many of the WTO's 'anti-democratic' characteristics are deliberate efforts by governments to limit certain forms of political access to the trade policy decisionmaking process.").

[30] SAMUELSON & NORDHAUS, *supra* note 15, at 400. For an extensive study of the post-

lateral economic protectionism became an urgent component of multilateral stability. The goal of the post–World War II international trade framework was to find ways to override domestic protectionist interests to negotiate trade liberalization that would allow for the expansion of the world economy and ensure both prosperity and peace.[31] Over time, the fast track system in the United States evolved as a means to that end.[32]

II. Updating the Convention Wisdom

The contemporary framework for trade legislation in the United States reflects the tenets that constraining Congress better allows for the negotiation of trade liberalization, and that trade liberalization benefits the majority. While these tenets are illustrated by the theoretical concepts and historical incidents discussed above, it is dangerous to believe that they are absolute.

This Part discusses two sets of complications for these tenets: institutional (Part II.A) and substantive (Part II.B). As an institutional matter, the scope and reach of economic minority interests in trade policy are no longer restricted to a protectionist focus on the legislature. Producer groups predominantly favor trade liberalization over trade protection. Trade legislation now reflects the interplay between these anti-protection interests and older, pro-protection interests. The tendency toward liberalization may be reinforced both by the influence of minority interests on the executive branch and by a shift in political ideology toward free trade. Amidst all of this new complexity, however, the preferences of the majority on trade policy in the United States remain obscured.

The continued institutional obscurity of majority preferences on trade in the United States is problematic because it is not clear, as a substantive matter, that trade liberalization will always benefit the majority. The premise that trade liberalization will benefit the majority is complicated by two sets of considerations. The first derives from well-established principles of welfare economics. Although classical economic theory predicts that trade will create gains for a majority of consumers, that prediction must be squared with implications elsewhere in economics that trade will generate losses for current workers in the United States. Although trade should maximize aggregate national wealth, the consumer gains from trade may be offset by the losses to labor—at the very least, the relationship between consumer gains and labor losses must be further investigated. The policy implications of this relationship are that it is un-

---

war negotiations leading to the International Trade Organization and the General Agreement on Tariffs and Trade, see generally Richard N. Gardner, Sterling-Dollar Diplomacy: The Origins and the Prospects of Our International Economic Order (1956).

[31] *See* Gardner, *supra* note 30, at 13–15.

[32] Hudec, *supra* note 19, at 295.

clear over the long term that the gains from trade, without a commitment to redistributive policy, will benefit a majority in the United States. That is, although trade should maximize aggregate national wealth, it will not necessarily maximize individual wealth.

The second set of considerations contests the tendency to see maximization of wealth as the only measure of social welfare. Because individuals may value goals that compete with wealth maximization, it may be too simplistic to conclude that, because trade liberalization theoretically will lead to aggregate national wealth maximization, it will benefit the majority.

A. Institutional Complications for the Assumption That Legislatures Are Beholden to Protectionist Minority Interests

The fast track legislative framework shifts authority over trade legislation away from Congress and toward the President.[33] Legislative constraint is deemed necessary in part because of the conventional wisdom that Congress remains dominated by protectionist special interests.[34]

As this Part shows, however, the influence of protectionist interest groups has waned markedly since the Smoot-Hawley era. The pro-liberalization policy that dominates now may well be just as much the product of special interests as was the protectionism that dominated the Smoot-Hawley era.

1. The Declining Importance of Protectionist Interests and the Increasing Importance of "Anti-Protection" Interests

While domestic industrial producers were largely protectionist in the early 1930s, when the Smoot-Hawley Act was enacted, economic interests have shifted away from protectionism in favor of trade liberalization.

Several events helped bring about this shift. First, as the U.S. economy grew in relative strength over the course of the twentieth century, domestic producers became less import-sensitive—that is, concerned about competition with foreign imports in the U.S. market. Instead they became and more export-oriented—that is, focused on opening up foreign markets to U.S. competition.[35]

---

[33] *See* 19 U.S.C. § 2112 (2000).

[34] Fast track allows legislators to avoid the political consequences of ignoring protectionist interests because it puts distance between their actions and the resulting trade bill. Protectionist interests oppose fast track, because without the possibility of congressional amendment, their interests are represented only by the President during trade negotiations. Therefore, the pressure of protectionist interests regarding a vote to delegate fast track authority would not be significantly different from, and if anything would be greater than, pressure brought to bear once the agreement has been negotiated and the government is subject to international pressure to implement it.

[35] *See* HELEN V. MILNER, RESISTING PROTECTIONISM: GLOBAL INDUSTRIES AND THE

Second, the rise of the international trade agreement as a trade policy tool created an incentive for export-oriented producers to organize as an interest group in domestic legislation, because such agreements link the trade practices of the United States with the trade practices of other member states.[36]

Third, the onset of globalization has meant that U.S.-based producer groups have become increasingly in favor of trade liberalization.[37] Multinational producers are not only indirectly concerned with keeping U.S. tariffs down (so as to persuade foreign governments to keep foreign tariffs down and preserve U.S. access to foreign markets), but are actually directly concerned with keeping U.S. tariffs down, because U.S.-based producers frequently import goods and services, either as inputs into domestic assembly and manufacture, or for sale in the domestic market.[38]

All of these dynamics mean that the dominant force among domestic producer groups is no longer protectionism, contrary to conventional wisdom.[39] To be sure, protectionist interests are far from gone, and are particularly strong in a few well-known sectors such as textiles and steel.[40] Trade legislation continues to allow for relatively high barriers for those industries.[41] But overall, legislators are willing to go along with a framework for trade legislation that they know will result in lower trade

---

POLITICS OF INTERNATIONAL TRADE 222–23, 225, 227–32 (1988). The importance of exports to U.S. producers dates at least to World War II, when the United States became a bona fide economic superpower and overtook war-torn Europe and Japan in industrial capacity.

[36] *See* IRWIN, *supra* note 24, at 1–3. While international trade agreements were rare prior to the Great Depression, their increasing importance thereafter permanently connected the interests of U.S. producers that wished to export to foreign markets and the trade barriers the U.S. government placed in front of foreign producers wishing to access the U.S. market. *Id.* at 2–3. As a consequence, export-oriented producers supported the reduction of U.S. trade barriers as a means of coaxing reciprocal reductions in foreign trade barriers and thereby increasing potential exports to foreign markets. *Id.* The Roosevelt Administration began an aggressive initiative to negotiate international trade agreements that featured the enactment of the Reciprocal Trade Agreements Act (RTAA) of 1934. *Id.* at 1. *See also* DOUGLAS A. IRWIN & RANDALL S. KROSZNER, INTERESTS, INSTITUTIONS, AND IDEOLOGY IN THE REPUBLICAN CONVERSION TO TRADE LIBERALIZATION, 1934–1945 (Nat'l Bureau of Econ. Research, Working Paper No. 6112, 1997). For a fascinating account of partisan politics between the Republican and Democratic parties, and their relationship to international trade policy, see *id.*

[37] "As globalization progresses, the pro-liberal forces became more concentrated . . . . The core political fact creating protectionism . . . had been reversed." Robert O. Keohane & Joseph S. Nye Jr., *The Club Model of Multilateral Cooperation*, *in* EFFICIENCY, EQUITY, LEGITIMACY, supra note 19, at 264, 268.

[38] *See* MILNER, *supra* note 35, at 222, 229–30.

[39] *See generally, e.g.*, I.M. DESTLER & JOHN S. ODELL, ANTI-PROTECTION: CHANGING FORCES IN UNITED STATES TRADE POLITICS (Inst. for Int'l Econ., Policy Analyses in International Economics No. 21, Sept. 1987); MILNER, supra note 35.

[40] *See* Richard N. Cooper, *Comment*, *in* EFFICIENCY, EQUITY, LEGITIMACY, *supra* note 19, at 207.

[41] Trade Act of 2002, Pub. L. No. 107-210, sec. 3103, § 204(b), 116 Stat. 933, 1024 (2002) (codified at 19 U.S.C.A. § 3203 (West. Supp. 2002)); Consolidated Appropriations-FY 2001, Pub. L. No. 106-554, § 146(a), 114 Stat. 2763, 252–53 (2000).

barriers. The "median legislator" is in favor of lower trade barriers as opposed to higher ones.[42]

2. The Shift in Political Ideology from Protectionism to Free Trade

Some studies have also suggested that political ideology influences legislators at least as much as the economic interests of particular groups in their constituencies and even of their constituencies as a whole.[43] Additional studies suggest that the general dominance of ideology as a predictor of voting patterns applies to trade legislation.[44] That is, legislators will often vote for trade legislation on the basis of whether they are generally supportive of, or suspicious of, free trade.

Political ideology has shifted from the protectionism of the 1930s to an embrace of free trade today. While the Smoot-Hawley Act found support at the time of its enactment in prevailing opinions on economic policy,[45] the "association of the Smoot-Hawley tariff with the Great Depres-

---

[42] *Cf.* HELEN V. MILNER, INTERESTS, INSTITUTIONS, AND INFORMATION: DOMESTIC POLITICS AND INTERNATIONAL RELATIONS 124 (1997).

[43] *See, e.g.*, Robert A. Bernstein & William W. Anthony, *The ABM Issue in the Senate, 1968–1970: The Importance of Ideology*, 68 AM. POL. SCI. REV. 1198 (1974) (showing ideology to be a better predictor of legislators' positions than benefits of anti-ballistic-missile-related spending to their districts); Kalt & Zupan, *supra* note 12, at 279 (arguing that the Senate in particular acts according to "altruistic-ideological" influences); Barton H. Thompson, Jr., *Judicial Takings*, 76 VA. L. REV. 1449, 1486–87 (1990). This view accords with protests by legislators that the pernicious dynamic described by public choice theorists is relatively rare in actual legislative practice. *See, e.g.*, Abner J. Mikva, *Foreword*, 74 VA. L. REV. 167, 167 (1988) ("Not even my five terms in the Illinois state legislature—that last vestige of democracy in the 'raw'—nor my five terms in the United States Congress, prepared me for the villains of the public choice literature.").

[44] *See, e.g.*, Judith Goldstein, *The Political Economy of Trade: Institutions of Protection*, 80 AM. POL. SCI. REV. 161, 173–81 (1986); James B. Kau & Paul H. Rubin, *Self-Interest, Ideology and Logrolling in Congressional Voting*, 22 J.L. & ECON. 365, 376–84 (1979).

[45] At the time of Smoot-Hawley, only professional economists seemed clearly opposed to the general wisdom of trade protection. For most of U.S. history prior to the 1930s, a policy of industry protection was preferred to free trade, on the grounds that such protection was necessary to ensure the successful industrialization of the United States, which began as a tiny and largely agrarian economy. *See generally* F. W. TAUSSIG, THE TARIFF HISTORY OF THE UNITED STATES (8th ed. 1966). United States tariff policy, for the entire history of the nation up to and including the Smoot-Hawley statute, aimed (in addition to generating government revenue) to protect the domestic United States economy, especially industry, from foreign competition, and to boost U.S. exports, especially agriculture, in which the United States had a comparative advantage from early on. *See* ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS 107–11 (Edwin Cannan, ed., 1976).

Not only was the Smoot-Hawley Act unremarkable in its protectionism, but the received wisdom that it established the "highest tariff schedule in American history" is mistaken as well, according to Alfred Eckes. Eckes, *supra* note 23, at 106. Eckes observes that the "*average duty on all imports*-free and dutiable" established by Smoot-Hawley (13.7%) was lower than it had been for almost a century prior to World War I. *See id.* Eckes points out that Smoot-Hawley's notorious 59.1% average tariff is based on the "average ad valorem equivalent rate on *dutiable imports* in 1932," but that "Congress did not enact such a high average tariff in 1930. Price declines during the depression produced that result,

sion and the collapse of trade . . . taught . . . a 'lesson' about the adverse consequences of high tariffs."[46] Since then, "the principle of free trade, not the principle of protection, has been the reigning ideology of trade policy in the country."[47]

Because prevailing trade ideology has moved away from protectionism in the years since World War II, contemporary legislators may be more likely to vote in line with free trade ideology than to seek to protect particular interests. This tendency would be promoted by the bulk and complexity of trade legislation—if it is difficult for a legislator to understand the specific costs and benefits of a trade bill, he may be more likely to vote on the basis of his general beliefs about trade.[48]

3. *The Influence of Minority Economic Interests on the Executive*

Another justification for constraining Congress is that the President is relatively more immune to rent-seeking by particular interest groups, given the national scope of his office and constituency.[49] To be sure, the initial act of delegation of trade legislative authority to the President in the mid-twentieth century quelled the pressure of protectionist groups.[50]

[since l]ike predecessor tariffs, Smoot-Hawley relied on a combination of specific and ad valorem duties." *Id.* Thus, the actual provable effects of the Smoot-Hawley Act may have ultimately been less important than its perceived effects, which propelled the dramatic changes in U.S. trade policy described above. *See* IRWIN, *supra* note 24, at 14–18.

[46] IRWIN & KROSZNER, *supra* note 36, at 10 (citing Robert Pastor, *The Cry-and-Sign Syndrome: Congress and Trade Policy*, in MAKING ECONOMIC POLICY IN CONGRESS 161 (Allan Schick ed., 1983)). Irwin and Kroszner ultimately find the ideology-as-causation argument unconvincing, however. *See id.* at 25.

[47] Pastor, *supra* note 46, at 184.

[48] On the difficulties legislators face in applying effective scrutiny to trade bills, see Robert Howse, How to Begin to Think About the 'Democratic Deficit' at the WTO 6–9 (Oct. 22, 2002) (unpublished manuscript, *available at* http://faculty.law.umich.edu/rhowse/Drafts_and_Publications/howse7.pdf).

[49] Jonathan Turley, *Dualistic Values in the Age of International Legisprudence*, 44 HASTINGS L.J. 185, 246 (1993). According to this theory, particular groups are less likely to wield disproportionate influence on the President, because the President is subject to pressures from a much larger and more diverse number of groups than are individual members of Congress. Moreover, as a unitary office, the executive cannot be susceptible to the logrolling dynamic associated with the legislative process. Thus, a process that shifts lawmaking authority away from the legislature and toward the executive arguably reduces distortions caused by private-sector producer groups and by the structural propensity toward logrolling, and allows for lawmaking that is more in tune with the public will. *See, e.g.*, Frank H. Easterbrook, *The State of Madison's Vision of the State: A Public Choice Perspective*, 107 HARV. L. REV. 1328, 1341 (1994); Mark Seidenfeld, *A Civic Republican Justification for the Bureacratic State*, 105 HARV. L. REV. 1511, 1568–69 (1992). For a comparison of presidential and congressional susceptibility to capture, see Richard Pierce, *Institutional Aspects of Tort Reform*, 73 CAL. L. REV. 917, 935–37 (1985). This argument has formed an important complement to the logrolling idea in the context of trade legislation. *See, e.g.*, Paul B. Stephan III, *Barbarians Inside the Gate: Public Choice Theory and International Economic Law*, 10 AM. U. J. INT'L L. & POL'Y 745, 754–59 (1995).

[50] *See* IRWIN & KROSZNER, *supra* note 36, at 4–5.

Over time, however, the ability of interest groups to influence policy in the executive branch has probably increased. The presidential administration has developed into a complex network of agencies, which may be as susceptible to interest groups as legislators.[51] The responsiveness of the President to the preferences of the general electorate may be inhibited to the extent that the President is dependent on administrative agencies for information-gathering and policy-formulation.[52] At first blush, this caveat may seem to apply less to trade policy than to other policy sectors. While an agency (the Office of the United States Trade Representative (USTR)) is charged with managing trade policy formulation and negotiation for the executive branch, the USTR generally does not act with finality, but rather acts subject to the direct approval of the President.[53] Thus, one might conclude that because the President enjoys considerable control over trade policy formulation, trade policy is relatively free from capture. However, the USTR relies on a broad cross-section of administrative agencies for advice and information in the context of formulating and negotiating trade policy, and of drafting implementing legislation.[54] This dependence on other administrative agencies arguably indirectly hampers the independence of the President and USTR from special interests and increases the likelihood of disproportionate industry influence.

Moreover, there are some indications the President may be less directly susceptible to *some* interest groups and more susceptible to others, although further inquiry into the direct influence of interest groups on the President is needed.[55] For example, multinational corporations, by virtue of their size, may be more likely than smaller producers to influence presidential policy-making.[56] Foreign interests may also exert relatively

---

[51] For a discussion of the scholarship on agency capture and broader studies of public choice, see Thomas W. Merrill, *Capture Theory and the Courts: 1967–1983*, 72 CHI.-KENT L. REV. 1039, 1050–52 (1997).

[52] *See, e.g.*, Christopher C. DeMuth & Douglas H. Ginsburg, *White House Review of Agency Rulemaking*, 99 HARV. L. REV. 1075 (1986); Richard J. Pierce, Jr., *The Role of the Judiciary in Implementing an Agency Theory of Government*, 64 N.Y.U. L. REV. 1239, 1281 (1989); Peter L. Strauss & Cass R. Sunstein, *The Role of the President and OMB in Informal Rulemaking*, 38 ADMIN. L. REV. 181, 187 (1986).

[53] *See* Pub. Citizen v. United States Trade Representative, 5 F.3d 549, 551 (D.C. Cir. 1993).

[54] *See* JACKSON, *supra* note 28, at 150–51. The USTR's dependence on other agencies is partially formalized in the statutory requirement that it chair an "interagency trade organization" charged with making recommendations and giving advice to, and performing other trade-related functions for, the President. *See* 19 U.S.C. § 2171 (c)(1)(I) (2000); 19 U.S.C. § 1872 (a)(3) (2000).

[55] Although there are some discussions of interest-group dynamics in the executive-branch context, *see, e.g.*, RICHARD E. NEUSTADT, PRESIDENTIAL POWER: THE POLITICS OF LEADERSHIP (1980), this literature pales in comparison to the volumes devoted to the study of interest-group influence on legislatures and administrative agencies.

[56] *See, e.g.*, Thomas M. Murray, *The U.S.-French Dispute Over GATT Treatment of Audiovisual Products and the Limits of Public Choice Theory: How an Efficient Market Solution was "Rent-Seeking,"* 21 MD. J. INT'L L. & TRADE 203, 214 (1997); Turley, *supra* note 49, at 239–59. "Past Presidents have defined foreign policy in terms of multinationals' interests

greater influence on the President.[57] These interests, incidentally, may be more likely to press for liberalization, as opposed to the protectionism traditionally expected to be favored by other minority economic interests.

Finally, even protectionist producers—the rent-seeking group whose influence the current framework was originally crafted to reduce—have developed ways of protecting their interests. Because trade legislation must be ratified by Congress, those legislators in constituencies dominated by a specific import-sensitive industries often negotiate special deals in exchange for their approval.[58]

4. *The Ineffectiveness of Voter Influence on Trade Policy*

Whether voters disagree or agree with trade liberalization, their views are unlikely to be registered effectively in the political process as a result of at least two dynamics. The first follows from the theory of capture that has animated much of the foregoing discussion—that benefits or harms to the majority are often too diffusely distributed to stimulate influential political activity. From this perspective, it is difficult to determine by assessing interest-group activity what the majority of voters think about trade liberalization.

The second dynamic arises more generally out of the American political landscape. Majorities can, after all, at times overcome collective action problems and make their voices heard—but the conditions must be right in order for this to happen. First, the issue must be of great enough importance to the "average" voter. Second, the voter must perceive a real

---

and have even engaged in combat over the property claims of multinationals based in the United States. Military forces were repeatedly sent to Cuba, the Dominican Republic, Honduras, Nicaragua, Panama and other countries to protect and advance United States business interests." Turley, *supra* note 49, at 251–53 & nn.333–41.

[57] Joel R. Paul, *The Geopolitical Constitution: Executive Expediency and Executive Agreements*, 86 CAL. L. REV. 671, 680–81 (1998). The susceptibility of the executive branch to foreign interests need not be pecuniary in nature—to some extent it is the "virtual monopoly power over foreign relations [that] has tempted presidents to . . . make foreign commitments." Paul, *supra*, at 679. For example, while in office, each President has concluded hundreds of international agreements, only a handful of which are generally known. *See* T.I.A.S. Index. Thus, "[l]eaving foreign policy to the sole discretion of the President has invited foreign regimes to pressure, cajole, and ultimately capture the executive." Paul, *supra*, at 680. Indeed, suspicion of foreign influence was a motivating factor behind locating many international policy-making powers in the Congress to begin with. *See* FEDERALIST NO. 68, Nos. 69, No. 73 (Alexander Hamilton); *see also* Paul, *supra*, at 733–37. In addition, though, there is some indication that foreign interests may exert direct pecuniary pressure on the President. *See generally* Note, Jessica S. Horrocks, *Campaigns, Contributions and Citizenship: The First Amendment Right of Resident Aliens to Finance Federal Elections*, 38 B.C. L. Rev. 771 (1997); Note, *"Foreign" Campaign Contributions and the First Amendment*, 110 Harv. L. Rev. 1886, 1890 (1997).

[58] During the legislative process for the NAFTA implementation bills, for example, the Clinton Administration cut "deals" with domestic industries in exchange for their withdrawal of legislative pressure, both directly and through interested legislators. *See* MILNER, *supra* note 42, at 112.

alternative.[59] In American politics, neither of these conditions appears to apply to trade policy.

Trade policy does gain some visibility from time to time in national political campaigns. Ross Perot's campaign in the 1996 election rested significantly on his position against trade liberalization, which attracted support from both the Democratic and Republican parties.[60] Perot's loss in that election (and Ralph Nader's in the subsequent one), however, arguably confirm the bipolar constraint on the American political system.

Within this bipolar constraint, voters who wish to vote according to trade policy are often constricted by the lack of a real choice.[61] To the extent such a choice has been provided, voters have tended to side with the candidate more skeptical of free trade, as in Bill Clinton's first election win based partially on his promises to build labor and environmental conditions into NAFTA.[62] In most cases, the absence of a clear choice persists, however, because both the Republican Party and Democratic Party are internally divided on trade.[63]

Moreover, intuitively it seems unlikely that American voters would switch allegiance between the two major political parties on the grounds of trade policy. Voters in the Democratic party displeased by a pro-trade stance due to environmental, consumer, or labor concerns are unlikely to support the Republican Party. Voters in the Republican Party who may oppose trade liberalization on populist, socially conservative groups grounds may also be unlikely to embrace the Democratic Party. Given the lack of consistently viable alternatives in the trade policy positions of the major political parties and candidates, voters may be relatively unable to exercise the franchise to direct politicians to respond to majority will on trade policy positions. As a result, trade politics in the United States landscape would appear to remain primarily the domain of minority economic interest groups.

---

[59] For an excellent study of electoral politics and trade policy, see generally DANIEL VERDIER, DEMOCRACY AND INTERNATIONAL TRADE: BRITAIN, FRANCE, AND THE UNITED STATES, 1860–1990 (1994). Verdier describes these two measures as "salience" and "divisiveness." *See id.* at 9–25.

[60] *See, e.g.*, Gerald F. Seib, *Alienated Army: Perot Voters Roam Political Landscape, Backing Varied Causes*, WALL ST. J., Aug. 12, 1994, at A1.

[61] *See* VERDIER, *supra* note 59, at 9–25. According to Verdier, an issue with low "divisiveness" is unlikely to become the subject of an election. *Id.* at 12. Political parties have little incentive to highlight the issue for voters, because they do not have a clear case to make to the voter as to why she should endorse their party over others based on that issue. *Id.*

[62] See Hannah L. Meils, Note, *A Lesson from NAFTA*, 78 INDIANA L.J. 877, 889 (2003).

[63] The Republican Party contains populists, like Pat Buchanan, who oppose trade, as well as proponents of free trade like George H. W. Bush and George W. Bush. The Democratic Party includes those who staunchly oppose trade agreements, such as current presidential candidate Dennis Kucinich, and those who staunchly endorse it, such as Senator Joe Lieberman.

5. Conclusion

Protectionist economic interests are less significant now than they were in the Smoot-Hawley era, and pro-liberalization economic interests are far more significant. Moreover, protectionist political ideology is less legitimate than it was during that era. Consequently, the decrease in the influence of protectionism, although in part a consequence of procedural constraints on Congress that buffer the effect of minority interests, stems also from socioeconomic shifts that changed the policy position of those minority interests without decreasing their influence.

When taken together with the growing influence of special interests on the executive branch and the obstacles to direct voter influence on trade policy, the contemporary trade legislative framework seems far from satisfactory as a means for forming trade policy that reflects majority interests. Political institutions continue to be highly imperfect barometers of majority will on trade policy.

B. Substantive Complications of the Assumption That Trade Liberalization Always Benefits the Majority

By demonstrating the unreliability of public institutions, public choice theory "suggests that the public interest will best be served by transferring decisional authority away from political institutions altogether . . . [to the] 'market,' understood to mean a social arrangement whereby outcomes are reached through the decentralized action of individual actors presumed to be rationally pursuing their own interests."[64]

The critiques from public choice theory of the legislative process described in Part II.A would seem to support trade liberalization as a policy: If the market is more efficient than the government in serving the societal interest, and trade liberalization is market-expanding, then trade liberalization serves society's interests.

Even beyond the realm of public choice, the view that trade liberalization is highly beneficial is pervasive. It stems in part from classical economics, and in part from the lessons of World War II. In the last two decades, this view has hardened into the tenet that trade liberalization is *always* the better policy. As Joseph Stiglitz has observed, the emphasis on trade liberalization has become part of an orthodoxy of policies dubbed the "Washington Consensus."[65]

---

[64] Merrill, *supra* note 51, at 1054; *see also* James Gwartney & Richard E. Wagner, *The Public Choice Revolution*, 23 INTERCOLLEGIATE REV. 17, 17 (1988). Public choice theory, in other words, suggests that a relatively decentralized mechanism (the market) results in an equilibrium of aggregated individual preferences that better serves society than does a relatively centralized mechanism (the government) that inaccurately identifies and implements "public choices."

[65] JOSEPH STIGLITZ, GLOBALIZATION AND ITS DISCONTENTS 67 (2002).

This rigid emphasis on trade liberalization, however, may fail to recognize a deep-rooted tension between market economics and the concept of democracy.[66] The market does not always distribute wealth evenly. As this Part will demonstrate using established principles of economics, pro-market outcomes may not necessarily be favored by a majority.

Before demonstrating the ambiguous implications for democracy of the pro-market position, a review of the social justification for pro-market policies may be in order. Within economics, the starting point for a theory of social welfare is the definition of the "social welfare function": the function of the *preferences* of the society's members.[67] Whereas some theories of the good embrace general principles that "trump" individual preferences, such as justice, fairness, liberty, or equality, welfare economics assumes "that the individual—and no one else—is the best judge of his own well-being."[68]

The problem with this eminently democratic model is that it is impossible to determine with perfect certainty the preferences of all members of society. As a result, it is impossible to specify with perfect accuracy the social welfare function.[69] Consequently, it seems, even a preference-regarding theory of social welfare must adopt assumptions that give "prior" shape to the social welfare function.

Within economics, the maximization of wealth often serves as a substitute for a richer definition of social welfare.[70] "Welfare economics" tries to determine the extent to which economic arrangements promote a definition of social welfare that is believed to approximate as closely as possible the actual preferences of all of society's members.[71]

---

[66] *See* Amy Chua, *The Paradox of Free Market Democracy*, 41 HARV. INT'L L.J. 287 (2000); Chantal Thomas, *Does the "Good Governance Policy" of the International Financial Institutions Privilege Markets at the Expense of Democracy?*, 14 CONN. J. INT'L L. 551, 558 (1999).

[67] Aggregate welfare is a function of the aggregate of individual utilities. An individual's "utility" is attained when that individual reaches the maximum level of satisfaction available to her given her choices. Thus, each individual's utility is itself a function of that individual's preferences. Social welfare, by extension, is a function of the preferences of all individuals in the society. *See, e.g.*, MUELLER, *supra* note 10, at 373–83. Note that the terms "welfare," "utility," and "efficiency" are essentially interchangeable in this discussion.

[68] E. J. Mishan, *A Survey of Welfare Economics*, 70 ECON. J. 197, 199 (1960). For theories of the social good based on justice, fairness, liberty, and equality, see generally JOHN RAWLS, A THEORY OF JUSTICE (1971).

[69] *See* Mishan, *supra* note 68, at 200.

[70] *See* MUELLER, *supra* note 10, at 373.

[71] *See* Mishan, *supra* note 68, at 200 ("[H]ypotheses about welfare can be cast into the form, if W then R; by which we mean that if we accept certain assumptions about the nature of the economic universe, and if we accept some particular definition of an improvement in social welfare, then we can draw an inference R—namely that when certain [economic] conditions are met social welfare is increased.") The dangers of circularity in this definition when social welfare is purely economic should be apparent: a given economic arrangement promotes social welfare, but social welfare is defined as a given economic arrangement.

If social welfare is defined as that condition under which society's wealth is maximized, and if the liberalized market can be better at producing greater aggregate wealth than the restricted market, then a policy that would liberalize the market would promote social welfare. On this view, every trade-liberalizing policy promotes social welfare.

However, this view of utility may too readily assume that people prefer the national maximization of wealth over alternative socioeconomic arrangements with different implications.[72] First, national wealth maximization may not equal individual wealth maximization. Second, people may hold values that compete even with individual wealth maximization. These competing values might be economic, such as greater economic security, or non-economic. Such conflicting preferences might mean that a majority may have good reason to hesitate at the prospect of dramatic trade liberalization, as opinion polls seem to suggest it does, without greater reassurance by governments that these other preferences will be addressed.

1. National Wealth Maximization Versus Individual Wealth Maximization

To show that national wealth is maximized says very little about the way in which that wealth is distributed. The economic theory of comparative advantage, which shows that nations gain from freer trade, is a *micro*economic principle, treating nations as individuals.[73] Yet a social welfare function defined as the total of all individuals' preferences should incorporate the question of distribution. If individuals want to maximize their wealth, a trade-liberalizing measure must be wealth-increasing for at least a majority in order to be preferred by a majority. Yet both theoretical argument and empirical evidence, as described below, suggest that the lowering of trade barriers, without more, may not increase median real income.[74]

---

[72] Herbert Hovenkamp's description is particularly apt:

> [E]conomists have settled on revealed market preferences as the most robust measure of the social utility that public laws and programs produce. These means have in time overtaken the economists' original ends, so that for many of those engaged in economic analysis of law, social welfare comprises no more than can be assessed by economic critera, particularly the Kaldor-Hicks measure of allocative efficiency.

Herbert Hovenkamp, *Legislation, Well-Being, and Public Choice*, 57 U. CHI. L. REV. 63, 63 (1990); *see also* Amartya Sen, *Rationality and Social Choice*, 85 AM. ECON. REV. 1, 15 (1995).

[73] *See* DOMINICK SALVATORE, INTERNATIONAL ECONOMICS 7 (6th ed. 1998).

[74] The pertinent measure here is the median real income, as opposed to the more commonly used mean. A mean does not indicate the distribution of income, whereas a median does. For example, the mean of the three incomes of $225,000, $50,000, $25,000 is

Comparative advantage identifies trade as beneficial because it increases consumption possibilities—that is, it increases the supply and thus lowers the price of goods. Free trade therefore benefits the individual as consumer: individual "real income" increases as a result of increased consumption possibilities.[75]

The benefit of these increases in real income should not be underestimated. The ability to purchase more basic goods for the same number of dollars constitutes a significant individual and social benefit and a significant gain in real income.

This gain, however, should be measured against the potential losses to be incurred by trade. One such loss is predicted by a basic theorem of international economics, the factor-price equalization theorem (the "FPE theorem"). The FPE theorem stipulates that international trade will bring about equalization in relative *and* absolute returns to "factors of production" (land, labor, and capital) across nations.[76] In the United States, labor is expensive relative to other nations and capital is cheap.[77] Therefore, according to the FPE theorem, free trade will cause the cost of labor in the United States—that is real wages—to decline to reach equilibrium with real wages abroad.[78] At the same time, the cost of capital in the United States will rise to reach equilibrium with the cost of capital abroad.[79]

The FPE theorem suggests that in a country such as the United States, the gains from trade will favor the "owners of capital" (investors, producers, employers) over the "owners of labor" (workers).[80] (This dynamic helps to explain the prevalent support for free trade by United States industry and the prevalent resistance to it by United States organized labor.[81]) One consequence of such a dynamic could be an increase in socioeconomic inequality.

$100,000, whereas the median is $50,000. As demonstrated by the example, a median, unlike a mean, is not skewed by outliers.

[75] *See supra* note 15.

[76] *See, e.g.*, SALVATORE, *supra* note 73, at 124. For the original proofs of the factor-price equalization theorem (the "FPE theorem"), see Paul A. Samuelson, *International Trade and the Equalisation of Factor Prices*, 58 ECON. J. 163, 165–84 (1948); Paul A. Samuelson, *International Factor-Price Equalisation Once Again*, 59 ECON. J. 181, 181–97 (1949).

[77] *See* Wassily Leontief, *Domestic Production and Foreign Trade: The American Capital Position Re-examined*, 7 ECONOMIA INTERNAZIONALE 3–32 (1954), *reprinted in* READINGS IN INTERNATIONAL ECONOMICS 503 (Richard E. Caves & Harry G. Johnson eds., 1968).

[78] SALVATORE, *supra* note 73, at 124.

[79] *Id.*

[80] *See* RONALD ROGOWSKI, COMMERCE AND COALITIONS: HOW TRADE AFFECTS DOMESTIC POLITICAL ALIGNMENTS 119, 120 (1989).

[81] *See id.* For a comparative discussion of the effects of the relationship between "factor inputs" and trade politics in different regions of the world, see generally *id.*

Empirical evidence indicates income inequality has in fact increased in the post-World War II era.[82] The U.S. Bureau of the Census reports that the rate of change in the "Gini coefficient," an accepted measure of income inequality, has more than doubled, in the direction of greater inequality, since 1947.[83]

Recent reports of, for example, the dramatic increase in corporate executive compensation relative to rank-and-file compensation support an inference that the gap between the very top of the income spectrum and the rest of it has widened even further.[84]

This inequality may result from a decline in median real income,[85] as some argue, or from a slower rate of income growth at the median level

---

82

Percentage of National Income

Income Group	1970	1980	1990	1995
Lowest fifth	4.1	4.3	3.9	3.7
Second fifth	10.8	10.3	9.6	9.1
Middle fifth	17.4	16.9	15.9	15.2
Fourth fifth	24.5	24.9	24.0	23.3
Highest fifth	43.3	43.7	46.6	48.7

ARTHUR F. JONES & DANIEL H. WEINBERG, THE CHANGING SHAPE OF THE NATION'S INCOME DISTRIBUTION 4 tbl.2 (Bureau of the Census, Current Population Reports, No. P60-204, June 2000).

[83] *See id.* at 2 fig.1 (showing that the rate of change stood at 5% in 1947 and at more than 10% in 1990).

[84] At congressional hearings on excessive compensation of corporate executives last year, labor representative Damon Silvers asserted that CEO pay had risen from "25 times" that of the average worker in 1964 to "500 times" today. *See CEO Compensation in the Post-Enron Era: Hearing Before the Senate Comm. on Commerce, Sci., and Transp.*, 108th Cong. (2003) (statement of Damon A. Silvers, Associate General Counsel, American Federation of Labor and Congress of Industrial Organizations). A more conservative estimate, provided at the same hearing by a Harvard Business School professor, holds that "there has been about a 7-fold increase in the inflation-adjusted median level of CEO pay since 1980, which far outstrips the increases seen by rank-and-file workers." *See id.* (statement of Brian Hall, Associate Professor, Harvard Business School). These dramatic figures, coupled with a rash of scandals regarding executive compensation, helped fuel federal legislative reforms. *See* Carol Hymowitz & Joann S. Lublin, *Corporate Reform: The First Year: Boardrooms Under Renovation—Scandals Prompted Changes, But Critics Say More Are Needed To Prevent Another Enron*, WALL ST. J., July 22, 2003, at B1.

[85] While the evidence as to whether median real income has suffered an absolute decline is equivocal, it appears clear that real income to labor has declined relative to real income to capital, supporting the first half of the FPE theorem. *See* Lawrence Mishel, *Rising Tides, Sinking Wages*, AM. PROSPECT, Fall 1995, at 61 (analyzing Bureau of Labor Statistics data and observing that falling real wages have resulted in part from the fact that "the rate of return to capital—profits, interest, dividends—has surged to postwar records and thereby dampened wage growth"). *But see* Michael J. Boskin, *Prisoners of Faulty*

than at the upper echelon. All agree, however, that the income inequality is there. The question is, can any of this increased inequality be attributed to greater trade?

Certainly, something as complex as socioeconomic inequality cannot be attributed to a single factor, even one as important as trade policy. To argue that the increase in inequality is the result of liberalized trade would be to overlook a vast multitude of other issues.[86] But the FPE theorem does predict that, in the United States, where labor is relatively expensive compared to other parts of the world, wages will decrease to reach an equilibrium. While the predictive power of the FPE theorem in real life is compromised by the fact that factors are not in actuality perfectly mobile, it does appear at least plausible that this dynamic is partially in play in the United States.

Politicians often argue that free trade will lead to higher-paying jobs.[87] If the FPE theorem is correct, wage increases in export sectors are due to labor immobility—that is, there is a demand for workers in new types of jobs for which there is not currently an adequate labor supply. If this immobility is short-term (because workers retrain or new workers migrate in), eventual labor supply increases in those sectors will narrow the wage premium. If labor immobility is long-term—because, for example, the new jobs require professional skills that are not widely held—wages may increase but will not be shared across total labor supply.

It turns out, then, that free trade can, by itself, be expected to lead to lower overall wages in the United States because of U.S. "factor endowments." (In the much longer term, it is presumably hoped that efficiency gains would increase economic opportunities across the board.) This vi-

---

Statistics, WALL ST. J., Dec. 5, 1996, at A20 (arguing that the Department of Labor statistics had over-valued inflation rates and therefore under-valued real income); Christina Duff, *Proposed Changes in Inflation Measure Criticized by Bureau of Labor Statistics*, WALL ST. J., Dec. 20, 1996, at A2. The increase in income inequality can be viewed as additional evidence of the decline of income to labor relative to income to capital, because earners in higher-income brackets receive a greater proportion of income to capital than earners in lower-income brackets. It is important to note that these observations do not compel a particular position on trade liberalization, because redistributive policies might correct any inequality (or one might believe that short-term inequality leads to long-term gain for all on a supply-side view of economic output growth). The point here is that without redistributive policy, one cannot expect that voters with perfect information would endorse trade liberalization if the above conditions apply.

[86] Greater workplace "efficiencies" due to technology gains, and infrastructural dynamics leading to a relatively unskilled workforce, could be just two such factors. *See* INT'L MONETARY FUND, WORLD ECONOMIC OUTLOOK chs. 3–4 (May 1997), *available at* http://www.imf.org/external/pubs/WEOMAY/Weocon.htm.

[87] In support of his bid to renew fast-track trading authority, President Bush said, "By promoting open trade, we expand export markets and create high-paying jobs for Americans." *House Votes Wide Trade Powers for Bush—"Fast Track" Authority Wins Passage in 215-214 Tally; Senate Passage Expected*, WALL ST. J., Dec. 7, 2001, at A3. President Clinton made similar arguments during his Administration in attempts to justify lowering trade barriers. *See, e.g.*, Remarks to the Chamber of Commerce in Little Rock, Arkansas, 1999 Pub. Papers 2252, 2258 (Dec. 10, 1999).

sion is consistent with a world in which upper-echelon income increases in the United States, but median income stays the same or decreases. If this is so, how can free trade be justified as a policy that is good for the majority of U.S. residents?

Redistribution is a first, and important, response. Government policies can redistribute the increased wealth resulting from trade gains from the upper echelon to the rest of the income spectrum, so that all "can benefit from international trade."[88] Better consumption is a second response. Even if wages decline due to increased trade, real income of wage laborers could increase because prices, also due to increased trade, decrease more than wages.

While these solutions seem sensible, they also show why the median individual in the United States might be skeptical of the promised gains from trade policy. That median person has to believe that the government *will* in fact undertake sufficient redistribution policies. She also has to believe that her cheaper shopping bills (to dramatize the situation a bit) will make up for her contracting wages. Finally, she has to believe that the increased profits that accrue to "owners of capital" will be reinvested in job-creating ways in the United States ("trickle-down" economics).

The above discussion shows how trade policy, like other dimensions of economic policy, can satisfy an economist's objective of maximized economic efficiency and wealth production at the same time that it fails to address the equity concerns of a democratic society. The "Kaldor-Hicks" definition of optimality in economics requires that those made better off could compensate those made worse off—the compensation principle.[89] In other words, the goal of economics is often to identify which

---

[88] *See* SALVATORE, *supra* note 73, at 127; *see also* J. E. MEADE, THE THEORY OF INTERNATIONAL ECONOMIC POLICY 2 (1955). Some commentators, however, argue that the greater the mobility of factors across barriers, the more likely regulatory competition is to discourage precisely such redistribution.

[89] For original expositions of the Kaldor-Hicks principle, see Nicholas Kaldor, *Welfare Propositions of Economics and Interpersonal Comparisons of Utility*, 49 ECON. J. 549, 550 (1939); J. R. Hicks, *The Foundations of Welfare Economics*, 49 ECON. J. 696, 700–01 (1939). In his treatise *Economic Analysis of Law*, Richard Posner makes an attempt to provide an "ethical" justification for being satisfied with the mere possibility of compensation: "What, in short, is the ethical basis of the Kaldor-Hicks concept, corresponding to the utilitarian, or preference-regarding, ethical basis of Pareto superiority? One answer is that the things that wealth makes possible—not only or mainly luxury goods, but leisure, comfort, modern medicine, and opportunities for self-expression and self-realization—are major ingredients of most people's happiness, so that wealth maximization is instrumental to utility maximization." RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 15–16 (4th ed. 1992). Yet Posner's response is adequate only in the sense that it discusses how to define individual utility; it cannot address social or aggregate utility, because a Kaldor-Hicks measure of social utility can "minimize" the wealth of some or even most individuals. *See* Hicks, *supra*, at 711–12. At the social level, the Kaldor-Hicks definition of optimal wealth is preferable to a Pareto definition of optimal wealth only if there is some mechanism, market or otherwise, for wealth/resource redistribution. Pareto optimal wealth or efficiency is defined as the situation in which no entity's wealth (or utility) can be increased without a decrease in another entity's wealth. PAUL A. SAMUELSON & WILLIAM D.

policies produce the most gain, and to leave to government the question of how best to distribute that gain or whether to sacrifice some of that gain. What is economically optimal in the aggregate may not be—on its own, without a redistributive scheme—preferred by a majority.

2. *Wealth-Maximizing Individuals Versus Risk-Averse Individuals*

If it is true that welfare is a function of total individual preferences, and that individuals are wealth-maximizing, then perhaps free trade can still accord with welfare (without more) if individuals prefer the relatively small chance of a relatively large increase in wealth presented by the trade liberalization model. Many economists, however, assume individuals to be risk averse.[90] If risk preference is incorporated into our welfare equation, then, it is even less likely that a majority of individuals will prefer trade liberalization to the status quo.

3. *Wealth-Maximizing Individuals Versus Individuals with Non-Wealth Preferences*

While it may be true that an individual would prefer a system that would increase economic gain at the expense of environmental protection, it may also be true that she would not; by extension, society might not.[91] People may rank preferences such as environmental protection, the international promotion of human rights or democracy, a higher level of health regulation, and so on above consumption gains.[92] What is neces-

---

NORDHAUS, ECONOMICS 149 (14th ed. 1992). For discussions of the compensation principle and its application to common law issues, see Peter Linzer, *On the Amorality of Contract Remedies*, 81 COLUM. L. REV. 111, 114 (1981); Edward Yorio, *In Defense of Money Damages for Breach of Contract*, 82 COLUM. L. REV. 1365, 1395 & n.154 (1982).

[90] *See, e.g.*, ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 45 (2d ed. 1997) (defining risk aversion as a state in which an individual "considers the utility of a *certain* prospect of money income to be higher than the expected utility of an *uncertain* prospect of equal expected monetary value") (emphasis added); A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS 53 (2d ed. 1989) ("[There is a] generally more realistic assumption that parties are *risk averse* (at least with respect to large risks).") (emphasis in original); POSNER, *supra* note 89, at 12 ("Risk aversion is not a universal phenomenon . . . [b]ut economists believe . . . that most people are risk averse most of the time.").

[91] *See* Amartya Sen, *Rationality and Social Choice*, 85 AM. ECON. REV. 1, 15 (1995) ("The public choice tradition has tended to rely a good deal on the presumption that people behave in a rather narrowly self-centered way—as *homo economicus* in particular . . . ."); *see also* DAVID COLLARD, ALTRUISM AND ECONOMY: A STUDY IN NON-SELFISH ECONOMICS (1978); AMITAI ETZIONI, THE MORAL DIMENSION: TOWARD A NEW ECONOMICS (1988); Robert H. Frank, *If Homo Economicus Could Choose His Own Utility Function, Would He Want One With a Conscience?*, 77 AM. ECON. REV. 593 (1987); Eileen Gauna, *The Environmental Justice Misfit: Public Participation and the Paradigm Paradox*, 17 STAN. ENVTL. L.J. 3, 45–46 (1998); Richard L. Hasen, *Clipping Coupons for Democracy: An Egalitarian/Public Choice Defense of Campaign Finance Vouchers*, 84 CAL. L. REV. 1 (1996).

[92] *See* David M. Driesen, *The Societal Cost of Environmental Regulation: Beyond Ad-*

sary is a renewed appreciation within economic policy, as Amartya Sen has demonstrated, for the "interconnectedness" of human fulfilment.[93] Ironically, the recognition of non-economic components to utility was part of the original definition of social welfare in welfare economics.[94]

More recently, work has been done regarding "public" goods produced by regulation that resists a strictly market-derived definition of utility.[95] This work has direct relevance for international trade in the new global era because trade liberalization, at least as it has been conceived to this point, suggests a trade-off between market maximization and forms of regulation of public goods that could be perceived as trade barriers.

4. *Opinion Poll Data*

If opinion polls are correct, a majority of voters rank issues such as employment security and consumer and environmental protection above increased purchasing power resulting from liberalized trade.[96] Of course,

---

ministrative Cost-Benefit Analysis, 24 ECOLOGY L.Q. 545 (1997); *see also* Hasen, *supra* note 91; *cf.* Richard Posner, *The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication*, 8 HOFSTRA L. REV. 487, 488–97 (1980).

[93] AMARTYA SEN, DEVELOPMENT AS FREEDOM 5 (1999) ("What people can positively achieve is influenced by economic opportunities, political liberties, social powers, and the enabling conditions of good health basic education, and the encouragement and cultivation of initiatives.").

[94] *See* Herbert Hovenkamp, *The First Great Law & Economics Movement*, 42 STAN. L. REV. 993, 993 (1990). More recently, Professors Farber and Frickey have carried on this tradition. *See* FARBER & FRICKEY, *supra* note 10, at 34 ("[C]ost-benefit analysis—or more specifically, the underlying standard of economic efficiency—cannot be applied until a prior decision is made about how to distribute social entitlements.").

[95] *See generally, e.g.*, Bruce Chapman, *Between Markets and Politics: A Social Choice Theoretic Appreciation of the Charitable Sector*, 6 GEO. MASON L. REV. 821 (1998). Another aspect of new work in economics criticizes standard economic assumptions regarding preferences not for the assumption of wealth-maximizing preferences per se, but rather for the assumption that such preferences will be rational. For a good discussion of this literature see Eyal Zamir, *The Efficiency of Paternalism*, 84 VA. L. REV. 229, 249–50 & nn.58–59 (1998). Also, much recent work has been devoted to the problem of incommensurability. *See generally, e.g.*, Richard Craswell, *Incommmensurability, Welfare Economics, and the Law*, 146 U. PA. L. REV. 1419 (1998).

[96] Opinion polls consistently reflect widespread misgivings about trade liberalization among the general public. *See, e.g.*, Pippa Norris, *Global Governance and Cosmopolitan Citizens*, *in* GOVERNANCE IN A GLOBALIZING WORLD 155, 170 (Joseph S. Nye & John D. Donahue eds., 2000) (reporting that opinion surveys showed that "only one-third of the public favored free trade over protectionism, with most people preferring trade limitations to maintain jobs, and this pattern was found even in postindustrial societies during the relatively affluent mid-1990s"); Helene Cooper, *Shift Into Reverse: Ban on Mexican Trucks In U.S. Interior Shows Rise of Protectionism*, WALL ST. J., Feb. 5, 1996, at A1 (citing recent Wall Street Journal/NBC News poll finding that sixty-two percent of the respondents thought trade agreements "endanger jobs"); Keith H. Hammonds, *BW/Harris Poll: Freer Trade Gets an Unfriendly Reception*, BUS. WK., Sept. 22, 1997, at 34 (finding that fifty-six percent of those surveyed believed that "expanded trade leads to a decrease in the number of U.S. jobs"); Paul Magnusson et al., *Economy Now for the Slow Track*, BUS. WK., Nov. 24, 1997, at 36 (discussing attempts by the White House and other advocates of fast track to address "growing public fears that free trade leads inevitably to economic and social disruption"); *see also* Helene Cooper, *Experts' View of Nafta's Economic Impact: It's a*

opinion polls are subject to all sorts of methodological objections, and one of the most important is the implications of the lack of information or misinformation supplied to those polled.[97] This Essay does not contest the difficulties raised by imperfect information and information asymmetries that can lead voters to reach erroneous conclusions on policy in general and on economic policy in particular.[98] The above discussion, however, suggests that individuals may have a "rational" (utility-maximizing) basis on which to reject trade liberalization.

5. Conclusion

Although trade liberalization will produce gains in real income through increased consumption possibilities, it may also incur other losses that must be reconciled with those gains, in order to determine whether it is a policy that will directly benefit the majority. The main such losses in the United States would be losses to labor in the form of decreased wages as "factors" move across borders to reach equilibrium. Even if such losses are only relative—that is, they take the form of a smaller gain than the gains accruing to capital or short term—they complicate the portrait of trade as an unqualifiedly beneficial policy. These complications are compounded by other values that might compete with trade liberalization, such as job security or environmental conservation.

In sum, the conventional wisdom that trade liberalization necessarily benefits the majority must be qualified. It may "benefit" (maximize the utility of) the majority *if*, for example, government also adopts sufficient

---

Wash, WALL ST. J., June 17, 1997, at A20 ("[R]ecent U.S. polls show a majority believe trade agreements like Nafta destroy jobs at home."); Greg Hitt, *To California Vinterns, Promised a Rose Garden, Fast-Track Bill Is Wreathed in Grapes of Wrath*, WALL ST. J., Oct. 6, 1997, at A24 (citing recent Wall Street Journal/NBC poll finding that only "23% of those surveyed said Nafta has had a positive impact on the nation's economy"); Gerald F. Seib, *So Are We All Free Traders? Well, Not Quite*, WALL ST. J., June 11, 1997, at A24 (citing recent Wall Street Journal/NBC News poll which found that "by a 43% to 28% margin, Americans tend to think Nafta has had a negative impact on America"). Interestingly, concerns about the viability of democracy in an internationally integrated economy have intensified in the West precisely at the time that the Western argument for the inherent interdependence of market economics and liberal democracy has finally prevailed in many previously resistant parts of the globe, such as the former Soviet bloc and much of the "Third World." *See* Richard Falk, *Regionalism and World Order After the Cold War*, 1995 ST. LOUIS-WARSAW TRANSATLANTIC L.J. 71, 71–74 (1995) ("[W]orld order in the near future is no longer likely to be provided to nearly the extent, as during the Cold War, by hegemonic state actors . . . . [T]he weakening of the state, in general, is producing various adverse types of societal vulnerability to the integrative tendencies in the global economy."); Robert F. Housman, *Democratizing International Trade Decision-making*, 27 CORNELL INT'L L.J. 699, 702 (1994) ("[A]t a time when the democratic preachings of the developed world seem to be having their greatest effect on the actions of developing and transition nations, these same developed nations are rushing head first into international trade agreements that offend the essential principles of democracy.").

[97] Seidenfeld, *supra* note 49, at 1568 & nn.282–86. *But see* Donald Wittman, *Why Democracies Produce Efficient Results*, 97 J. POL. ECON. 1395, 1407–08 (1989).

[98] *See generally* VERDIER, *supra* note 59.

redistributive policies and protects non-trade values. Or it may do so if gains from trade exceed losses, or are reinvested in wealth-redistributing ways. But all these qualifications are a far cry from the notion that liberalizing trade is always the indisputably superior policy in a democratic society.

If the conventional wisdom on the absolute superiority of trade liberalization must be qualified, then the current legislative framework, which sets up a streamlined mechanism to achieve trade liberalization, may also need qualification to obtain democratic legitimacy.

III. Implications

The substantive and institutional complications discussed above render the conventional wisdom on trade policy problematic. The constraints on the legislative process for trade policy in the United States find their political justification in part on the grounds that the superior trade policy choice is unambiguously trade liberalization. As Robert Hudec has written: "The procedural components of legitimacy complaints—lack of transparency, lack of participation—are not absolutes. When good policy seems to require it, they have been overriden in the past, and they will be again in the future."[99]

If the factors determining the margin and distribution of gains from trade are more complex than often assumed, however, the superiority of trade liberalization as a policy choice becomes qualified. By extension, a legislative framework set up to produce trade liberalization becomes potentially problematic from the perspective of democratic legitimacy.[100]

What to do? At the very least it seems that the current trade legislative framework must be altered to provide for greater protection of concerns that the democratic majority might rationally have: concerns about wealth distribution and about "non-trade" goals such as environmental conservation. This is easier said than done, however. This Essay concludes by pointing to a few areas for further research.

---

[99] *See* Hudec, *supra* note 19, at 295, 299.

[100] To justify free trade, proponents must argue that the benefits from trade obtain notwithstanding the contrary majority preference; that is, that trade benefits the public *interest* in potential opposition to the majority of *preferences*. A more complex theory of democratic governance, in which constraints on majoritarian preferences are justified as necessary to achieve a larger public good, becomes necessary. For a thorough discussion of the "public interest" and "public preferences" as competing concepts of the public good, and of the role of constitutional theory in mediating these competing concepts, see Chantal Thomas, *Constitutional Change and International Government*, 52 HASTINGS L.J. 1, 32–39 (2000).

A. Institutional Quandaries at the Domestic Level

Returning trade policy to the legislature might not only create serious impediments for international and intergovernmental coordination; it might also fail to improve significantly the correlation between trade policy and majority will. As Part II.A demonstrates, interest-group influence on the legislature remains strong.

How to improve the democratic legitimacy of the trade legislation framework? One possibility is to insist on a broader representation of interest groups within the mechanisms that exist for holding trade policy accountable to domestic politics. The Presidential Advisory Committee on Trade Policy, for example, consists largely of industry groups, with labor, environmental, and other "public interest" groups underrepresented. An area for further research is the implications of "public interest" groups, such as environmental and consumer organizations, for public choice theory.[101]

Perhaps the process of making trade law and policy in the United States cannot become more democratic until it becomes more salient for voters, and until real alternatives begin to present themselves in the context of the electoral process. That is, until American voters care enough about trade to demand greater accountability, and until American political parties can present the electorate with real choices on trade policy, one might expect trade policy to be dominated by interest groups.

B. Institutional Quandaries at the International Level

Scholars have devoted increasing attention to the potential for creating channels of public access and participation at the international level and the relationship between such reform, and the legitimacy of international organizations.[102] In some ways, creating more space for international "civil society" and transnational "networks" may neutralize the institutional obstacles to democratic decisions at the domestic level. On the other hand, the "global governance" frameworks might be subject to public choice problems no less significant than those characterizing domestic governance frameworks. At the very least, however, greater trans-

---

[101] The appearance of the public interest group as legislative advocate appears to problematize capture theory and the Mancur Olson model set forth above—while losses or costs resulting from lower levels of consumer or environmental protection will be as diffuse as benefits resulting from greater consumption possibilities, there are many groups seeking to protect the former, but few looking to promote the latter.

[102] *See generally, e.g.*, THE ROLE OF LAW IN INTERNATIONAL POLITICS: ESSAYS IN INTERNATIONAL RELATIONS AND INTERNATIONAL LAW (Michael Byers ed., 2000); THE LEGITIMACY OF INTERNATIONAL ORGANIZATIONS (Jean-Marc Coicaud & Veijo Heiskanen eds., 2001); LEGALIZATION AND WORLD POLITICS (Judith L. Goldstein et al. eds., 2001); GOVERNANCE IN A GLOBALIZING WORLD (Joseph S. Nye & John D. Donahue eds., 2000); EFFICIENCY, EQUITY, LEGITIMACY, *supra* note 19.

parency seems essential to improving the democratic legitimacy of international organizations.[103]

C. Substantive Quandaries: Implications for the Developing World and Distributive Justice Among and Within Nations

If it is the case that trade liberalization will cause greater redistribution of wealth from rich countries to poor countries, then it is not clear that trade protection, though it might be preferable for the popular majority in the United States, accords with goals of global justice and equity. For those concerned with the "equity" side of the equity-efficiency equation, the question of reconciling domestic redistribution with international redistribution remains extremely difficult. For example, labor groups in the United States might argue for trade protection on the basis of equity and democracy, seeking to ensure greater job security and higher wages for those in the United States. From the perspective of developing countries, however, such an argument might seem deliberately designed to avoid sharing American wealth with workers from poor countries.

Ideally, of course, there is some way out of this apparently zero-sum equation: global labor protections and other social regulations might increase productivity across the board. Clearly, however, greater attention must be devoted to reconciling the interests of the economically vulnerable within developed and developing countries. This is an area for further research.

CONCLUSION

The problems for democratic legitimacy presented by the current trade legislative framework in the United States are those problems that challenge governance frameworks globally: how to ensure greater responsiveness to public participation, and how to integrate trade policy with the whole range of policy goals, including equitable redistribution and social protection. Devising solutions for those problems requires an appreciation of the importance of complexity in trade policy, both substantively and institutionally.

---

[103] *See* Robert Howse & Kalypso Nicolaidis, *Legitimacy and Global Governance: Why Constitutionalizing the WTO Is a Step Too Far*, *in* EFFICIENCY, EQUITY, LEGITIMACY, *supra* note 19, at 227; Thomas, *supra* note 100, at 42–43.

ESSAY

RESTORING INVESTOR TRUST IN AUDITING STANDARDS AND ACCOUNTING PRINCIPLES

MELVYN I. WEISS*
ELIZABETH A. BERNEY**

In this Essay, Melvyn Weiss and Elizabeth Berney address the current problems in the regulatory and legal framework surrounding corporate auditing and accounting in light of recent accounting fraud scandals at companies such as Enron and WorldCom. First, they present the historical development of the accounting profession and its responsibilities, and analyze the recent historical development of pressures on accountants and corporate managers that incentivize fraud. They then argue that regulators of corporate accounting, including Congress, have failed to effectively address the threat of fraud. Finally, they address these current shortcomings by offering specific proposals to increase the scope of audits; enhance regulatory oversight; increase the amount, meaningfulness, and usability of information companies and auditors provide to the public; improve auditor education; protect auditor independence; improve internal audits; and improve the legal accountability of auditors and companies.

Federal Reserve Board Chairman Alan Greenspan recently testified before the Senate Committee on Banking, Housing, and Urban Affairs that "our market system depends critically on trust—trust in the word of our colleagues and trust in the word of those with whom we do business. Falsification and fraud are highly destructive to free-market capitalism and, more broadly, to the underpinnings of our society."[1] The loss of trust has reached epidemic proportions. Of special concern is the sudden and dramatic increase in fraud allegations against Fortune 500 companies.[2] It is not only the everyday frauds and lack of transparency in the mid-size

---

* Co-founder and senior partner, Milberg Weiss Bershad Hynes & Lerach LLP. J.D., New York University Law School, 1959; B.A., Baruch College of the City College of New York, 1957.

** Attorney, Milberg Weiss Bershad Hynes & Lerach LLP. J.D., University of Chicago Law School, 1978; B.S., Cornell University, 1975.

[1] *Federal Reserve Board's Semiannual Monetary Policy Report to Congress: Hearing Before the Senate Comm. on Banking, Hous., & Urban Affairs*, 107th Cong., at 11 (2002) [hereinafter *Reserve Board Hearings*] (statement of Alan Greenspan, Chairman, Fed. Reserve Bd. of Governors). *See also* Mike Allen, *Bush Signs Corporate Reforms Into Law; President Says Era of 'False Profits' Is Over*, WASH. POST, July 31, 2002, at A4 (quoting President Bush stating, "[c]orporate corruption has struck at investor confidence, offending the conscience of our nation").

[2] A recent comprehensive GAO study found that the average size of companies that restated their financials increased from $500 million market capitalization in 1997 to $2 billion market capitalization in 2002. U.S. GEN. ACCOUNTING OFFICE, FINANCIAL STATEMENT RESTATEMENTS: TRENDS, MARKET IMPACTS, REGULATORY RESPONSES, AND REMAINING CHALLENGES 4 (2002) [hereinafter GAO FINANCIAL RESTATEMENTS REPORT].

industrial and consumer companies that have become a way of life, and the daily fudging that has made comparability between companies in the same industry impossible. Now, it is also the WorldCom-class financial frauds that have made investing so perilous.

Corporations have been disseminating numbers that are increasingly unreliable. The number of companies issuing financial restatements has increased dramatically in recent years, from 116 companies in 1997 to 270 companies in 2001.[3] As the *Washington Post* has noted, "[f]or every high-profile company that restates its financial results, hordes of lesser-known firms revise their numbers far from the public spotlight."[4] Restatements run across the entire gamut of industries, not merely "new economy" companies as is commonly thought.[5] In fact, "about 10 percent of all listed companies announced at least one restatement" between January 1997 and June 2002.[6]

Individual and institutional investors rely on financial reports every day. If corporate reports cannot be trusted, investors are victimized and suffer immensely. It becomes impossible for investment managers to do their job of sifting through companies' public statements, comparing companies within industries, and making the decisions that will earn or cost them and their clients millions or billions of dollars.[7] Suppliers of goods, services, and capital, who depend on the creditworthiness of their customers in deciding with which ones to do business, suffer as well. The lack of reliable financial reporting destroys the fundamental purpose of our securities laws, "to substitute a philosophy of full disclosure for the philosophy of *caveat emptor* and thus to achieve a high standard of business ethics in the securities industry."[8]

Auditors have not met their obligations to the investing public and bear significant responsibility for the current state of affairs.[9] When Xerox restated four years of its revenues by a stunning $6.4 billion in June 2002,

---

[3] Lynn Cowan, *More Large-Cap U.S. Companies Restate Results Than Small Fry*, WALL ST. J., Aug. 9, 2002, at A2. *See also Accounting and Investor Protection Issues Raised by Enron and Other Public Companies, Oversight Hearing Before the Senate Comm. on Banking, Hous. & Urban Affairs*, 107th Cong. (2002) [hereinafter *Accounting and Investor Protection Issues*] (prepared statement of Prof. Joel Seligman, Washington Univ. Sch. of Law) ("Last year, a study by Financial Executives International, a trade group for corporate executives, found that public companies had revised their financial results 464 times between 1998 and 2000, *nearly as many restatements as in the 20 previous years combined*, and the problem probably worsened last year.") (quoting Alex Berenson, *The Biggest Casualty of Enron's Collapse of Confidence*, N.Y. TIMES, Feb. 10, 2002, § 4, at 1).

[4] Cowan, *supra* note 3.

[5] *Id.*

[6] GAO FINANCIAL RESTATEMENTS REPORT, *supra* note 2, at 4.

[7] *See* JOSEPH E. STIGLITZ, THE ROARING NINETIES 126 (2003) ("When share prices reflect bad information, resources are likely to be badly deployed.").

[8] *In re* Initial Pub. Offering Sec. Litig., 241 F. Supp. 2d 281, 299 (S.D.N.Y. 2003) (quoting SEC v. Capital Gains Research Bureau, 375 U.S. 180, 186 (1963)).

[9] JORDAN E. GOODMAN, READING BETWEEN THE LIES 12 (2003).

the *Washington Post* quoted former SEC Chief Accountant Lynn E. Turner as saying, "It boggles the mind that a company this size and its executives can make an error of $6.4 billion and no one sees it until years later . . . . It should make investors wonder if the auditors would even notice Mount Everest if they were driving by it."[10] To make matters worse, auditors' failures are also being used as a shield by corporate wrongdoers to avoid their legal responsibilities. Share-holder securities fraud class actions have encountered corporate defendants arguing that they are not liable because auditors permitted their actions. Examples are many, but they include such high-profile cases as Enron,[11] Microstrategy,[12] and Xerox.[13] Meanwhile, auditors argue that finan-cial statements are those of management and that accountants are only required to render an opinion based on Generally Accepted Accounting Principles ("GAAP") and Generally Accepted Accounting Standards ("GAAS"). They take the position that they are not responsible if a client corporation commits fraud that goes undetected by the auditors. For example, Dennis Nally, chairman of PriceWaterhouseCoopers LLP, recently declared that auditors are merely "responsible for a fair presentation of the financials"[14] and that "the auditor is not responsible for detecting fraud."[15] Nally furthermore termed the public's reliance on accountants to keep companies honest an "expectation gap."[16]

Accountants' defenses are particularly difficult to overcome under the federal securities laws, which require plaintiffs to prove that a defendant committed wrongdoing with scienter.[17] Moreover, if plaintiffs do not have sufficient evidence in advance to name the accountants as defendants under the heightened standards of the Private Securities Litigation Reform Act of 1995 (the "PSLRA"),[18] or the accountants are dismissed from a case because plaintiffs cannot meet the high scienter standards under the Act,[19] the corporation and corporate insiders can then blame the

---

[10] Kathleen Day, *Xerox Restates 5 Years of Revenue; '97–'01 Figures Were Off by $6.4 Billion*, WASH. POST, June 29, 2002, at A1.

[11] *In re* Enron Corp. Sec., Derivative & Erisa Litig., Civ. No. H-01-3624, 2003 U.S. Dist. LEXIS 7632, at *45 (S.D. Tex. Apr. 24, 2003). In *Enron*, the defendant "said that he relied on his staff, [a law firm] and [auditor] Andersen to make sure the disclosures were correct and complied with the rules." *Id.*

[12] *In re* MicroStrategy, Inc. Sec. Litig., 148 F. Supp. 2d 654, 666 (E.D. Va. 2001).

[13] Defendant Xerox's Motion to Dismiss, at 6, Carlson v. Xerox, No. 00 Civ. 1621 (D. Conn. filed Dec. 2, 2002).

[14] Cassell Bryan-Low, *Accounting Firms Aim to Dispel Cloud of Corporate Fraud*, WALL ST. J., May 27, 2003, at C1.

[15] *Id.*

[16] *Id.*

[17] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, §§ 21D(b)(1) & (2) (1995) (codified at 15 U.S.C. §§ 78u-4(b)(1) & (2)).

[18] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, §§ 27(a)(2)(A), 27A(c), 201(g)(2), 109 Stat. 737, 738, 750, 758–59 (1995) (codified in various sections of 15 U.S.C.).

[19] *See In re Rite Aid Corp. Secs. Litig.*, 269 F. Supp. 2d 603, 608 (E.D. Pa. 2003) (not-

absentee accountants, using the classic "empty chair defense," a powerful jury argument for avoiding management liability.

This Essay reviews the historical development of auditing and how corporate executive compensation increased minefields in financial reporting, and suggests a number of proposals to improve the current situation. These proposals include requiring broader audits in order to improve the auditor's ability to detect fraud; increasing regulatory oversight of audit activities and removing responsibility for writing GAAP and GAAS from an institution dominated by the accounting industry; educating auditors with special training in fraud detection; requiring companies and auditors to make more meaningful information publicly available; increasing the independence of auditors; implementing stronger internal audit systems in companies; and increasing the legal accountability of auditors and company managers.

I. Historical Development

In order to solve the problems of inadequate auditing and financial reporting, it is useful to revisit the historical development of the audit function. Auditing began as a management tool to detect employee fraud.[20] This was corporate management's greatest need through the beginning of the twentieth century.[21] At this early juncture, there was no concept of auditor independence.[22] Auditors sat on corporate boards and did not have a unified professional society or status as a profession.[23] There was no disciplinary oversight authority governing how auditors conducted themselves.[24] They were not public accountants.[25] The financial statements

---

ing that, before the passage of the PSLRA, the scienter bar for accountants was already high).

[20] Rufus Wixon, The Accountant's Handbook 1-5 (4th ed. 1957).

[21] R. Gene Brown, *Changing Audit Objectives and Techniques*, *reprinted in* The Evolution of Audit Thought and Practice 696–99 (T.A. Lee ed., 1988); *see also* John L. Carey, The Rise of the Accounting Profession From Technician to Professional 1896–1936, at 25 (1969) (advertisements providing clues as to the nature of accounting practice in the late nineteenth century included among accountants' duties "searching out and reporting upon the possible shortages of cash and securities of trusted employees" and organizing offices to adopt proper safeguards against employee shortages); Peter F. Drucker, The Essential Drucker 6 (2001) ("During the 1920s and 1930s . . . [a]ccounting went from bookkeeping to analysis and control").

[22] Even after the advent of the securities laws, a popular auditing text advised that "open-mindedness" was an important quality for auditors: "While it is important that the [auditing] applicant shall be firm and not easily swayed, still he should not be too opinionated and should be amenable to reason. This is necessary because he must listen to instructions and follow them irrespective of his own opinions. A person who might obstinately refuse to do what he was told would not succeed as a public accountant." Paul E. Bacas et al., Auditing Procedure, 23 (rev. ed. 1948).

[23] *See* Dale L. Flesher et al., *The First Century of the CPA*, 182 J. of Acct. 51, 51–52 (1996); *see also* Carey, *supra* note 21, at 36–371.

[24] Flesher, *supra* note 23, at 52.

[25] *Id.*; *see also* American Inst. of Certified Pub. Accountants, *The Enron Crisis: The AICPA, the Profession & the Public Interest, a Brief History of Self-Regulation* (Feb. 20,

they prepared for their clients did not have to comply with any meaningful standards.[26] The 1946 edition of a popular accounting text, initially issued in 1924, explained,

> [T]here is no compilation or code of accounting principles, and accountants' reports provide ample evidence that accounting is not controlled by a body of principles which result in uniformity of procedure and treatment. There are several reasons for this condition.
>
> In the first place, a principle is a fundamental truth, a fundamental law, or a fundamental assumption which forms the basis of reasoning or conduct; but many of the rules and procedures of accounting are not of so fundamental a nature.[27]

It was in this atmosphere that the famous *Ultramares Corp. v. Touche & Co.*[28] decision arose, in which accountants were held not liable to a lender that had relied on an inaccurate balance sheet certified by the accountants. It is no wonder that, at a time when accounting was not held out to be a disciplined profession upon which the public might rely, Judge Cardozo's majority opinion held that accountants held no negligence responsibility to third-party users of financial reports.

After the stock market crash of 1929, financial institutions, primarily the commercial banks that were victimized by fraudulent corporate financial statements, realized they needed protection—much the way institutional investors are coming to realize the same thing today. As a result, banks put pressure on accountants to reform themselves. The American Institute of Accountants, predecessor to the American Institute for Certified Public Accountants, or AICPA, was formed, and the concept of independence was first formulated.[29]

2002), *at* http://www.aicpa.org/info/regulation02.htm.

[26] *See* JOHN L. CAREY, THE RISE OF THE ACCOUNTING PROFESSION TO RESPONSIBILITY AND AUTHORITY 1937-1969, at 5 (1970) ("Prior to the stock-market crash of 1929, there were no authoritative standards governing corporate financial reports"). Carey recounts that at the 1912 annual meeting of an early accountants' association, in response to criticism regarding the enormous variation in accountants' treatment of inventory valuations, various expenses, and other items, and the suggestion that a committee be appointed to examine the feasibility of adopting uniform accounting standards, the presenter of a paper stated, "All of us know perfectly well there is no book for accountants . . . you have to take every case on its own merits and deal with it as you see it. . . . You cannot lay down definite principles, but you can lay down broad principles that will cover most cases. When you come to interpret the principles, you will have to take the case in point and do the best you can with it upon a conservative basis." CAREY, *supra* note 21, at 78–79.

[27] H.A. FINNEY, PRINCIPLES OF ACCOUNTING, INTERMEDIATE 194 (3d ed. 1946).

[28] 174 N.E. 441, 448 (N.Y. 1931).

[29] *See* CAREY, *supra* note 26, at 20–21.

At about the same time, Congress decided that the investing public also needed a financial watchdog.[30] During the deliberations leading up to the securities acts of 1933 and 1934, legislators debated whether the government or the private sector should perform the watchdog function. The accountants begged to take on this role,[31] and ultimately, the profession was granted that franchise.[32] Congress delegated to the SEC the task of overseeing the accountants.[33]

Despite the fact that the earliest function of the auditor was to look for employee fraud, and that Congress, the SEC, and commercial lenders designated the accounting profession—at its own behest, even insistence—as the financial watchdog, the profession has generally declined to accept responsibility for ferreting out fraud.[34] Accountants maintained an approach of merely examining the books with a "degree of skepticism."[35] Moreover, the terminology of accounting continued to be "somewhat loose and vague,"[36] and "individual judgment" was favored over general standards.[37]

---

[30] The legislative history of the Securities Exchange Act of 1934 described the objects of accounting at that time: "(a) To give to management that accurate information and aid which is essential to the successful conduct of a business, and (b) to give to actual and prospective creditors that accurate information essential to the determination of the volume of credit which may safely be extended and the conditions under which it may be allowed .... The need of accurate information is still paramount; but, under the conditions of today, the next object in order of importance has become 'to give stockholders, in understandable form, such information in regard to the business as will avoid misleading them in any respect and as will put them in possession of all information needed, and which can be supplied in financial statements, to determine the true value of their investments.'" H.R. REP. NO. 1838, 73d Cong., at 12 (1934), *reprinted in* FEDERAL SECURITIES LAWS: LEGISLATIVE HISTORY, 1933–1982 (1983), at 305 [hereinafter SEC. LAWS LEGISLATIVE HISTORY]. Moreover, some of the motivation behind the enactment of the securities laws was to increase investor confidence by ensuring accurate and honest accounting. *See In re* Enron Corp. Sec., Derivative & ERISA Litig., Civ. No. H-01-3624, 2002 WL 31854963, at *5 (S.D. Tex. Dec. 19, 2002) (quoting United States v. O'Hagan, 521 U.S. 642, 658 (1997)).

[31] SEC. LAWS LEGISLATIVE HISTORY, *supra* note 30, at 305.

[32] *Id.*; *see* CAREY, *supra* note 26, at 5–6.

[33] SEC. LAWS LEGISLATIVE HISTORY, *supra* note 30, at 305. *See* CAREY, *supra* note 26, at 5–6.

[34] Bryan-Low, *supra* note 14 (quoting comments of PricewaterhouseCoopers Chairman Dennis Nally).

[35] CONSIDERATION OF FRAUD IN A FINANCIAL STATEMENT AUDIT, Statement on Auditing Standards No. 99, § 13 (American Inst. of Certified Pub. Accountants 2002). Effective for audits of financial statements for periods beginning on or after December 15, 2002, the statement warns that "the auditor's exercise of professional sceptcism is important when considering the risk of material misstatement due to fraud." *Id.*

[36] FINNEY, *supra* note 27, at 194.

[37] *Id.* at 195.

II. SKEWED INCENTIVES AND MORAL HAZARDS FACING EXECUTIVES AND ACCOUNTANTS

A. *Understanding the Stimuli for Committing Financial Fraud*

We live in an era when there are enormous incentives for top executives to cheat. For instance, federal investigators recently examined whether large executive incentive packages for three senior officers motivated accounting improprieties at Computer Associates.[38] The grant to these executives of shares valued at one billion dollars was conditioned on the company's share price closing above a certain target for sixty days within any twelve months during a five-year period.[39]

The lives of some CEOs are full of material excess, and some are practically deified by their companies and the business press. As the *Wall Street Journal* noted, "[b]y now, L. Dennis Kozlowski's $6,000 shower curtain is the stuff of corporate legend. So, too, is the $2 million birthday party in Sardinia thrown by the former Tyco International Ltd. chief executive for his wife, complete with an ice sculpture of Michelangelo's David spewing vodka from its nether regions."[40] Bill Black, former litigation director for the Federal Home Loan Bank Board's efforts to clean up the savings and loan debacle of the 1980s, recently discussed the enormous incentives, control and perquisites of upper management:

> If I am the CEO, all the internal controls are designed by me. All of them report to me. I choose the external controls such as the auditor. I can fire the outside auditor. People are not willing to look at me as a thief. These frauds create record profit and an enormous turnaround. What are the normal things you do for a CEO at that point? You give them a raise. You give them a bonus. You give them stock dividends, and the stock appreciates like all crazy because of what's happening to the income statement and the balance sheet. These guys get private air forces. The perks can run into the tens of millions.[41]

At WorldCom, compensation-committee members reportedly referred to former CEO Bernard Ebbers as "God," "Jesus Christ," and "Superman," and awarded him compensation to match, including a $400 million

---

[38] Michael Schroeder et al., *Accounting Crackdown Focuses Increasingly on Top Executives*, WALL ST. J., Apr. 12, 2002, at A1.

[39] *Id.*

[40] Mark Maremont, *Finally a CEO Faces a Jury: Kozlowski*, WALL ST. J., Sept. 25, 2003, at C1.

[41] Jake Bernstein, *Trash For Cash: From S&Ls to Enron: Bill Black Offers a Primer on CEO Fraud*, TEX. OBSERVER, May 24, 2002, *available at* http://www.texasobserver.org/showArticle.asp?ArticleID=738.

loan and a $1.5 million annual severance package (amounts they are now trying to recoup).[42] Former Enron CEO Kenneth Lay reportedly received more than $205 million from 1999 to 2001 in salary, cash bonuses, and cashed in stock and options in the company.[43] Lay received the bulk of these funds "by repeatedly borrowing from Enron, and then turning in stock to repay the loan, [thus] Mr. Lay was taking out cash directly from the company when the company's need for cash was growing."[44] Executives fear that failure to achieve their numbers can result in the loss of their privileged lifestyles. The more we provide executives with excess incentive compensation and other perks, the greater the risk of improper conduct.[45]

Many of the problems we see today have been caused by material changes in the way corporate managers are compensated.[46] Prior to the 1970s, company managers predominantly fell into two categories: the entrepreneurs—who owned stock because they created or originally took the risk and financed the company—and the professional managers—who served as the interim caretakers of other people's invested assets.[47] Significant wealth, when made, was generally earned by the creators and founders.[48] The managers were paid well, but did not receive the massive compensation of the successful entrepreneurs.

The landscape shifted when mergers and acquisitions started to proliferate. The use of "junk bond" assets, created to inflate companies' value artificially so that their stock could be used to acquire other com-

---

[42] Lori Calabro, *Above Board*, CFO MAG., Oct. 2003.

[43] Floyd Norris, *Enron's Collapse: Market Place; For Chief, $200 Million Wasn't Quite Enough Cash*, N.Y. TIMES, Jan. 22, 2002, at C1.

[44] *Id.*

[45] The accounting profession's own standards list as a risk factor an environment in which a "significant portion of management's compensation [is] represented by bonuses, stock options, or other incentives, the value of which is contingent upon the entity achieving unduly aggressive targets for operating results, financial position, or cash flow." CODIFICATION OF ACCOUNTING STANDARDS AND PROCEDURES, Statement on Auditing Standards No. 82, § 316.17(a) (American Inst. of Certified Pub. Accountants 1997).

[46] *See* Schroeder, *supra* note 38. *See also* Roger L. Martin & Mihnea C. Moldoveanu, *Capital Versus Talent: The Battle That's Reshaping Business*, HARV. BUS. REV., July 2003, at 36 (CEO compensation packages soared by 434% on average between 1991 and 2000); Stiglitz, *supra* note 7, at 118 ("Executive pay would become a topic of growing controversy, in the later part of the decade [1990s], as stock options enabled the likes of John Chambers at Cisco, Dennis Kozlowski of Tyco, Sanford I. Weill of Citicorp and David Kowarsky of Merrill Lynch to pocket millions of dollars.").

[47] JOHN C. COFFEE, JR., WHAT CAUSED ENRON? A CAPSULE SOCIAL AND ECONOMIC HISTORY OF THE 1990S, at 7–9 (Columbia Law Sch., Ctr. for Law and Econ. Studies, Working Paper No. 214, 2003), *available at* http://papers.ssrn.com/sol3/delivery.cfm/SSRN_ID373581_code030124630.pdf?abstractid=373581.

[48] 77 CONG. REC. 2910–55 (1933) (statement of Rep. Sam Rayburn), *reprinted in* SEC. LAWS LEGISLATIVE HISTORY, *supra* note 31, at 175 ("The managements of these big corporations, as a rule, own an insignificant percentage of the outstanding voting stock. For example, the management of 44 steam railroads owns only 1.2 percent of the common stock, the management of 36 mining corporations owns only 1.8 percent of the common stock.").

panies, fueled the change. The transactions fed by this financing taught corporate managers that they could achieve tremendous wealth by means of golden parachutes and other exit compensation packages.[49] These compensation packages were paid to corporate managers as incentives to, in effect, terminate the existence of their companies by permitting them to be taken over.[50] Many times, managers permitted these transactions without regard to the longstanding interests of the communities in which they did business, the companies' employees, or the many businesses that serviced these corporations.[51] Many of the junk bonds used in these transactions were then sold to savings and loan institutions to increase the S&L's capital, thereby enabling these institutions to obtain more FDIC-insured deposits that were lent improperly and recklessly, often to entities in which the bankers had interests.[52]

Without the complicity of accountants, such bonds could not have been viewed as having the value necessary to accomplish these transactions.[53] The accountants avoided marking the junk bonds to their realistic market values as required by prudent generally accepted accounting principles.[54] Auditors also closed their eyes to other accounting mischief in the savings and loan industry.[55] Ultimately, these practices caused savings and loan banks to melt down,[56] necessitating government takeovers to bail out the savings and loan industry.[57]

---

[49] *See* COFFEE, *supra* note 47, at 9, n.10.

[50] *Id.*

[51] *See generally* MARTIN MAYER, THE GREATEST-EVER BANK ROBBERY: THE COLLAPSE OF THE SAVINGS AND LOAN INDUSTRY (1990).

[52] *See generally id.*

[53] *Id.* at 293–94 (describing involvements of the (then) Big Eight accounting firms with failed and fraud-ridden savings and loan institutions). Mayer notes that the "GAO, a branch of Congress, studied the eleven largest thrift failures in 1987 and found seven of them audited with so little regard for honest practice that it referred the firms involved to the AICPA for disciplinary action. In spring 1990, no action had been taken . . . . A draft of a new AICPA guide for auditing S&Ls emerged in 1989, but many of the scams still pass the AICPA's audit test." *Id.*

[54] Gary Weiss, *Commentary: Congress Will Huff and Puff and . . . Do Little*, BUSINESSWEEK ONLINE, Feb. 25, 2002, *at* https://www-secure.businessweek.com/tms-subs-s/bin/auth.cgi?H66038d508b08b89df795da19a77533a1:url=http:Zb2fZb2fwww.businessweek.comZb2fcgi-binZb2fregisterZb2farchive.cgiZb3fcZb3dZb26yZb3d02Zb26wZb3d08Zb26hZb3db3771014.htm&ear=0&ipr=0&grp=8Zc1Zc6Zc51&fmt=si32&oip=140.247.202.15&rc=iv&kid=400001.148&ss=env.

[55] Saeed Roohani et al., *The S&L Crisis: A Learning Experience For Accountants*, 7 J. BANK COST & MGMT. ACCOUNTING 60, 72 (1994) ("[A] major factor cited by the GAO Report is that auditors failed to properly communicate insolvency information of the troubled S&L's to the public.").

[56] *Id.*

[57] *See* MAYER, *supra* note 52, at 2 (noting that the monetary cost of the S&L bailout to taxpayers was probably more than $140 billion out of pocket, and would exceed $500 billion when government borrowings to pay for the bailout are included in the computation), 19 ("Who could have stopped this atrocious theft from the government insurance funds and didn't? The accounting profession, which acquiesced in the most outrageous and dishonest 'practice' that misstated both the balance sheets and the profit-and-loss accounts, year after year.").

When this mergers and acquisitions activity ceased to allow the managers to enrich themselves, other methods had to be found to entice top managers to work at companies. As a result, the use of stock options as a form of management incentive compensation became rampant in the 1990s.[58] And with that development, stock options, rather than M&A dealmaking, became the new stimulus for cheating.[59] The use of stock options in historically large amounts has led to accounting fraud beyond anything anyone could have imagined.[60]

B. Sources of Pressure for Misreporting

It seems clear that misreporting is stimulated from the top. To quote Alan Greenspan:

> If a CEO countenances managing reported earnings, that attitude will drive the entire accounting regime of the firm. . . . It has been my experience on numerous corporate boards that CEOs who insist that their auditors render objective accounts get them. And CEOs who discourage corner-cutting by subordinates are rarely exposed to it.[61]

The Tyco situation demonstrates how pressure is exerted from the top down. After receiving its outside counsel's report, Tyco's current management concluded, in a report filed with the SEC in December 2002, that Tyco suffered from

> pressure on, and inducements to, . . . unit managers to increase current earnings, including by decisions as to what accounting treatment to employ; and a lack of a stated and demonstrable commitment by former senior corporate management to set ap-

---

[58] *See* Stiglitz, *supra* note 7, at 116 ("By 2001, options accounted for an estimated 80 percent of the compensation of American corporate managers.").

[59] *See* Kris Maher & Kemba J. Dunham, *Options Frenzy: What Went Wrong? Who Got the Most From Exercising Options?*, WALL ST. J., Dec. 17, 2002, at B1 (listing ten biggest recent gains by executives exercising options, varying from $147,252,540 to $706,076,907; many of the listed executives were subsequently involved in high-profile corporate scandals).

[60] *See* COFFEE, *supra* note 47, at 8–9. *See also* Bethany McLean & Peter Elkind, *Partners in Crime*, FORTUNE, Oct. 2003, at 78, 81 ("The problem was that [Enron's] new businesses, ranging from trading paper products to creating a global broadband network, consumed billions of dollars—and many of them never made money. Yet Enron wasn't about to admit that to the world, because the news would have crushed the company's highflying stock—and Enron executives had a fortune in options at stake.").

[61] RESERVE BOARD HEARINGS, *supra* note 1, at 11 (statement of Alan Greenspan, Chairman, Federal Reserve Bd. of Governors).

> propriate standards of ethics, integrity, accounting, and corporate governance.[62]

Even the accounting profession's own standards list "[m]anagement setting unduly aggressive financial targets and expectations for operating personnel" as a risk factor associated with misstatements arising from fraudulent financial reporting.[63]

The accountants themselves are subject to a variety of similar influences that impair their independence, in the form of pressure exerted by client companies—who, after all, pay the bills—to produce favorable audit results.[64] In February 2001, when the SEC issued an early round of proposed rules regarding auditor independence, it stated that "it is plain that there is ample basis to conclude that the more a person, including the auditor, has at stake in a judgment, the more likely his or her judgment is to be affected."[65] The SEC stressed that the influences that it was concerned with "can be 'extremely subtle.'"[66] Former Federal Reserve Chairman Paul Volcker, testifying about the impact of an accounting firm's outside relationships on audit practices, said that a real threat was posed by the "insidious, hard-to-pin down, not clearly articulated or even consciously realized, influences on audit practices."[67] During his tenure in the 1990s, former SEC Commissioner Arthur Levitt was "worried that

---

[62] Tyco Int'l Ltd. Current Report (Form 8-K), SEC File No. 001-13836, at 4 (Dec. 30, 2002).

[63] CODIFICATION OF ACCOUNTING STANDARDS AND PROCEDURES, Statement on Auditing Standards No. 82, § 316.17(a) (American Inst. of Certified Pub. Accountants 1997).

[64] The Supreme Court recognized the critical need for auditor independence twenty years ago:

> By certifying the public reports that collectively depict a corporation's financial status, the independent auditor assumes a *public* responsibility transcending any employment relationship with the client. The independent public accountant performing this special function owes ultimate allegiance to the corporation's creditors and stockholders, as well as to the investing public. This 'public watchdog' function demands that the accountant maintain total independence from the client at all times and requires complete fidelity to the public trust.

United States v. Arthur Young & Co., 465 U.S. 805, 817–18 (1984). More recently, other courts have noted the deleterious effects of incentives that impair auditor independence. *See, e.g.*, *In re* Complete Mgmt. Inc. Secs. Litig., 153 F. Supp. 2d 314, 334–35 (S.D.N.Y. 2001) (accountants' desire to maintain consulting fees "created incentives for the auditors to seek to please [the company's] management at the expense of accuracy and/or completeness"). Congressional testimony in the Enron matter similarly noted that "[a]t its core, Enron involved an audit failure. The outside auditor both appeared to operate with significant conflicts of interest and to have been too beholden to a highly aggressive corporate management." *Accounting and Investor Protection Issues*, *supra* note 3, at 6.

[65] Auditor Independence Requirements, Exchange Act Release No. 33-7919, 65 Fed. Reg. 76,008, 76,017 (Dec. 5, 2000).

[66] *Id.* (quoting written testimony of John D. Hawke, Jr. (July 26, 2000)).

[67] *Id.* (quoting written testimony of Paul A. Volcker (Sept. 13, 2000)).

auditors not only lacked the incentive to do the job that they were supposed to do; they had perverse incentives."[68]

But at the same time, as auditor independence was eroding, the organized accounting profession had, incredibly, been lobbying effectively to neutralize the laws that would hold auditors accountable. A major success attributable to these lobbying efforts was the passage of the PSLRA,[69] which dramatically increased pleading difficulties[70] and substituted proportionate liability for joint and several liability, making it less attractive to sue accountants.[71] Under proportionate liability, even if a lawsuit is successful, only a portion of the total liability is assessed against the accountants; this limited recovery may not justify the cost of the suit.

In addition, the Supreme Court's 1994 decision in *Central Bank of Denver, N.A. v. First Interstate Bank of Denver, N.A.*,[72] eliminating liability in private litigation for aiding and abetting a securities fraud violation, removed the principal tool used by victims of such conduct to sue

---

[68] Stiglitz, *supra* note 7, at 133.

[69] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, §§ 27(a)(2)(A), 27A(c), 201(g)(2), 109 Stat. 737, 738, 750, 758–59 (1995) (codified in various sections of 15 U.S.C.).

[70] 15 U.S.C. § 78u-4(b)(2) requires that a complaint asserting claims under 10(b) of the 1934 Securities and Exchange Act must "state with particularity facts giving rise to a strong inference that the defendant acted with the required state of mind." Even Professor John Coffee, Jr., who recently declared that he "sees no problem with the PSLRA's heightened pleading standards, as they have been interpreted by some courts," John C. Coffee, Jr., *Understanding Enron: "It's About the Gatekeepers, Stupid,"* 57 BUS. LAW. 1403, 1410, n.34 (2002), conceded that "[a]uditors were the special beneficiaries of many of [the PSLRA's] provisions" *Id.* at 1410. Coffee described the special difficulties shareholders now face in pursuing auditors: whereas shareholders may point to corporate insiders' stock sales as facts giving rise to a "strong inference" of fraud, "it is rarely possible to [make a pleading similar to an insider information charge] with respect to the auditors, who by law cannot own stock in their client. In short, the plaintiff faces a 'Catch 22' dilemma in suing the auditor: it cannot plead fraud with particularity until it obtains discovery, and it cannot obtain discovery under the PSLRA until it pleads fraud with particularity." *Id.* at 1410 n.35.

[71] Prior to passage of the PSLRA, senior author Weiss warned that "[a]doption of legislation abrogating joint and several liability . . . would lead to more fraud and less investor protection." Melvyn I. Weiss, *Report of Litigation Crisis Greatly Exaggerated By Big 6; Big 6 Accounting Firms and Securities Fraud*, ACCT. TODAY, Nov. 1, 1993, at 10 [hereinafter Weiss, *Litigation Crisis*]. *See also* Melvyn I. Weiss, *Auditors, Public Both Need Tough Liability Rules*, ACCT. TODAY, Jan. 24, 1994, at 9 [hereinafter Weiss, *Tough Liability Rules*] ("Eliminating joint-and-several liability would be particularly unwise in light of the unique role and primary duty that accountants have in assuring the reliability of audited financial statements.").

[72] 511 U.S. 164, 191 (1994). *Central Bank* was based on a strict interpretation of the Securities and Exchange Act: "Because the text of [Securities and Exchange Act] § 10(b) does not prohibit aiding and abetting, we hold that a private plaintiff may not maintain an aiding and abetting suit under § 10(b)." *Id.* at 191. The Supreme Court further explained that "[t]he absence of § 10(b) aiding and abetting liability does not mean that secondary actors in the securities markets are always free from liability under the securities Acts." *Id.* The Supreme Court offered that secondary actors "may be liable as [ ] primary violator[s] under 10b-5, assuming *all* of the requirements for primary liability under Rule 10b-5 are met," and acknowledged that "[i]n any complex securities fraud, moreover, there are likely to be multiple violators." *Id.*

accountants.[73] Before the *Central Bank* decision, aiding and abetting liability had played a crucial role in policing accountants; the dissenting justices in *Central Bank* noted, "[i]n *hundreds* of judicial and administrative proceedings in every Circuit in the federal system, the courts and the SEC [had] concluded that aiders and abettors [were] subject to liability under § 10(b) and Rule 10b-5."[74] In addition to eliminating aiding and abetting liability, "[i]t is generally agreed that *Central Bank* foreclosed a cause of action . . . for conspiracy to violate § 10(b) and Rule 10b-5, in addition to aiding and abetting."[75] Given the scope of the corporate frauds and losses of shareholder value during the past few years, it is surprising that Congress has not overturned this decision and restored aiding and abetting securities fraud by statute, despite sixty years of prior jurisprudence stating that it was the law.[76] Nor has it restored conspiracy to commit securities fraud.

The fate of our firm's lawsuit against WorldCom on behalf of its shareholders in June 2001, a year before the $3.8 billion WorldCom fraud became public knowledge,[77] is a prime example of what happens when auditors are not vigilant and the legal hurdles shareholders face have been made impossibly high. In July 2002, *Forbes* magazine published an article entitled *Asleep at the Switch*.[78] The article explained, "WorldCom book-cooking was laid out chapter, line and verse in a shareholder suit over a

---

[73] *Id. See also Shapiro v. Cantor*, 123 F.3d 717, 719–21 (2d Cir. 1997) (*Central Bank* marked "the end of any free-standing duty by collateral participants in securities transactions to blow the whistle . . . .") (quoting Robert A. Prentice, *Locating that "Indistinct" and "Virtually Nonexistent" Line Between Primary and Secondary Liability Under Section 10(b)*, 75 N.C. L. REV. 691, 760 (1997)). *Shapiro* held that following *Central Bank*, accountants "must themselves make a false or misleading statement (or omission) that they know will reach potential investors" in order to be liable under § 10(b) of the securities laws. *Id.* at 724 n.142.

[74] *Central Bank*, 511 U.S. at 192 (Stevens, J., dissenting).

[75] *In re* Enron Corp. Sec., Derivative & Erisa Litig., Civ. No. H-01-3624, 2002 WL 31854963, at *22 (S.D. Tex. Dec. 19, 2002) (citing cases).

[76] *Central Bank*, 511 U.S. at 197–98 n.8 (Stevens, J., dissenting) ("the private right against aiders and abettors . . . has become a part of the established system of private enforcement") (citing cases, a law review article, and legislative sources). *But see Shapiro*, 123 F.3d at 720-21 (citing cases with conflicting views as to existence of accountants' aiding and abetting liability prior to *Central Bank*).

[77] On June 25, 2002, WorldCom announced that it intended to restate its financial statements for 2001 and the first quarter of 2002 by a total of $3.8 billion. *See* Press Release, Securities & Exchange Commission, SEC v. WorldCom, Inc., Litig. Release No. 17588, Accounting and Auditing Release No. 1585 (June 27, 2002) (on file with authors). The SEC thereafter announced its filing of a civil action seeking money penalties and injunctive relief against WorldCom. *Id.* Such an SEC enforcement action involves no recovery for shareholders and minimal deterrence. The largest civil penalty ever assessed in an SEC enforcement action to date was $10 million (for the Xerox $6.4 billion accounting fraud). *See* Press Release, Securities & Exchange Commission, Xerox Settles SEC Enforcement Action Charging Company with Fraud, Release No. 2002-52, (Apr. 11, 2002) (on file with authors).

[78] Neil Weinberg, *Asleep at the Switch*, FORBES, July 22, 2002, at 38.

year ago. Sadly, a judge . . . tossed it out as directors, auditors, regulators—and the press—snoozed."[79]

The *Forbes* article also explained that the complaint "was backed by 100 interviews with former WorldCom employees and related parties," and lauded the fruits of that work as "startling in their breadth and detail."[80] The *Forbes* article gave samples of the complaint's detailed allegations:

> A former New York sales rep told of WorldCom's cutting bandwidth prices in half for [a] client . . . , then booking the order twice—once at the old rate and once at the new one. A Tulsa, Okla. quality-assurance analyst said WorldCom's balance sheet listed assets that included receivables as much as seven years past due.[81]

Yet the judge dismissed the case "with prejudice,"[82] so that it cannot be refiled, and said that the shareholders' complaint against WorldCom was a classic example of "puzzle pleading"[83] that did not meet the "heightened pleadings requirements for this type of case."[84] This decision demonstrates that these heightened pleading standards have been disastrous for investors.

With the courts reacting to Congress's call for tougher pleading standards, and while one accounting scandal after another was breaking out in recent years, what was the AICPA, the institution designed to ensure auditor independence, doing? In December 2002, former SEC Chairman Arthur Levitt said in *Forbes* magazine, "I think the AICPA under [Barry] Melancon's leadership has been the least effective, most backward, most obstructionist group that I encountered in my eight years running the SEC."[85] The *Forbes* article noted that the AICPA's chief executive "fought the SEC over auditor independence rules that Levitt and [former SEC chief accountant Lynn] Turner say would have stopped conflicts involving auditors selling consulting services to their audit clients."[86] According to Douglas Carmichael, then the director of the Center for Financial Integrity at Baruch College in New York, "[t]he AICPA also

---

[79] *Id.* at 38–39.
[80] *Id.*
[81] *Id.*
[82] *In re* MCI WorldCom, Inc. Sec. Litig., 191 F. Supp. 2d 778, 794 (S.D. Miss. 2002). There remains the possibilty of an appeal of the suit against MCI Worldcom once it emerges from bankruptcy.
[83] *Id.* at 782.
[84] *Id.* at 782.
[85] Elizabeth McDonald, *The Man With Nine Lives*, FORBES, Dec. 9, 2002, at 60.
[86] *Id.*

didn't tighten their rules . . . [because] auditors like weak rules, since they can use them as a defense when sued."[87]

In 1993 and 1994, senior author Weiss wrote a series of articles in *Accounting Today* arguing that until the accounting profession clearly accepts responsibility for ferreting out fraud, accountants will fail as watchdogs.[88] The need for accountants to assume that responsibility has become even more evident today. But the issues are also becoming more complex, as the profession explores the boundaries between financial fraud on the one hand, and so-called "aggressive accounting" on the other.

An excellent example of this complexity can be seen in the investigation of accounting at Tyco undertaken by David Boies, a prominent outside attorney, at the request of its board. Tyco's restatement in December 2002 erased $382 million in earnings for its fiscal year ending September 30, 2002.[89] The Tyco report is typified by statements such as, "Aggressive accounting is not necessarily improper accounting."[90] Reporting on this, a *New York Times* article asked,

> Can accounting that follows the stated rules still be unreliable?
>
> . . .
>
> After a year of corporate scandals in which some of the most outrageous financial reporting appears to have complied with generally accepted accounting principles, or GAAP, the answer appears to be yes . . . .
>
> Time and again, [Boies's] report says, Tyco aggressively used accounting gimmicks to make its reported financial statements look far better than they otherwise might.[91]

Accounting and auditing appear to be of little help if, when supposedly independent outside directors appoint a world-class lawyer to investigate, the lawyer is able to minimize the magnitude of the restatement by calling these inconsistencies "accounting gimmicks." It is virtually impossible not to conclude that there were systematic improprieties when a cor-

---

[87] *Id.* at 60–61. Carmichael has since been appointed chief auditor of the new Public Company Accounting Oversight Board. *See* Robert Schmidt, *Accountant Critic Named Top Auditor for Audit Board*, BLOOMBERG NEWS SERV., Apr. 17, 2003, 18:24.

[88] *See* Weiss, *Litigation Crisis*, *supra* note 71 at 10; Melvyn I. Weiss, *CPAs Sought Role of Assuring Financial Integrity; Accountants Obligated to Furnish Investors with Accurate Financial Information*, ACCT. TODAY, Dec. 13, 1993, at 36; Melvyn I. Weiss, *CPA Firms Wage Scorched-Earth Legal Tactics; Evaluation of Cost of Litigation for Accounting Firms*, ACCT. TODAY, Jan. 3, 1994, at 22; Weiss, *Tough Liability Rules*, *supra* note 71, at 9.

[89] Tyco Int'l Ltd., Current Report (Form 8-K), SEC File No. 001-13836, at 5 (Dec. 30, 2002).

[90] *Id.* at 7.

[91] Kurt Eichenwald, *Pushing Accounting Rules to the Edge of the Envelope: Report Focuses on Tyco's Aggressive Tactics*, N.Y. TIMES, Dec. 31, 2002, at C1.

poration overestimates earnings by $382 million. When companies repeatedly use aggressive tactics to increase reported earnings, these methods should be recognized as unacceptable. The *New York Times* article explains:

> While the accounting rules allow for interpretations ranging from the conservative to aggressive, companies are effectively graded pass-fail: either they receive a signature from their accountants attesting to their compliance with the rules, or they do not. There is no indication in a company's audited results whether or not it is fluffing its numbers through aggressive tactics.[92]

Incentives and conflicts can, of course, also result in auditors' failing to report fraud that is known to them, in addition to passively failing to uncover fraud. For instance, the GAO's review of various cases reported that:

> Arthur Andersen . . . was [Waste Management, Inc.]'s independent auditor and wrote unqualified opinions for the company during the periods that were eventually restated, although Arthur Andersen had identified the company's improper accounting practices *and quantified those practices in relation to the company's financial statements.* According to the complaint filed by [the] SEC, *these misstatements were presented annually to the company's management*, along with "Proposed Adjusting Journal Entries" to correct the errors.[93]

In addition, the SEC's recent study of five years of enforcement actions found that "violations by auditors . . . resulted largely from auditors failing to gain sufficient evidence to support the issuer's accounting, failing to exercise the appropriate level of skepticism in responding to red flags, and failing to maintain independence."[94]

III. Shortcomings of Regulators

As these problems grew, investors left it to the regulating authorities to determine what auditors needed to do. Unfortunately, investors cannot rely on the legislative or regulatory systems to protect them. The Sarbanes-Oxley Act,[95] passed by Congress in the summer of 2002, when

---

[92] *Id.*

[93] GAO Financial Restatements Report, *supra* note 2, at 216 (emphasis added).

[94] Sec. & Exch. Comm'n, Report Pursuant to Section 704 of the Sarbanes-Oxley Act of 2002, at 3, *available at* http://www.sec.gov/news/studies/sox704report.pdf.

[95] Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, § 104(b)(1), 116 Stat. 745,

Congress should have had the greatest motivation to regulate, falls short of what is needed in several respects. First, Sarbanes-Oxley left the formidable pleading and liability standards and limitations on discovery imposed by the PSLRA untouched, thereby perpetuating investors' difficulties in holding wrongdoers accountable.[96]

Second, it is questionable whether the legislation's new CEO and CFO certifications requirements[97] will be more efficacious than prior practices in preventing fraud. The new rules, effective immediately, require CEOs and CFOs to sign a written statement accompanying each periodic report filed under Sections 13(a) and 15(d) of the Securities Exchange Act, certifying that the report complies with the requirements of those sections[98] and certifying that the report "fairly presents, in all material respects, the financial condition and results of operations of the issuer."[99] The new certification requirements make us wonder: what were CEOs and CFOs thinking throughout all those years when they signed their companies' SEC filings? We predict that it will only be a matter of time before executives who are prone to creating illusory financial results revert to past practices.

Third, the Public Company Accounting Oversight Board ("PCAOB")[100] established by the Act is already self-destructing and using its public trust for self-aggrandizement. At its first formal meeting, the board members voted themselves annual salaries of $452,000 each—three times that of a federal judge—prompting a prominent Columbia University economist to suggest in the *New York Times* that the PCAOB was on its way to proving Noble Laureate George Stigler's theory that regulated industries end up co-opting their regulators.[101]

Moreover, the PCAOB is charged with inspecting larger registered accounting firms only once per year and smaller registered accounting firms only once every three years.[102] Although this sparse schedule may be adjusted by rule,[103] it seems evident that the board cannot rely on sur-

---

(2002).

[96] However, while pleading standards and discovery limitations remain the same, Sarbanes-Oxley Act § 804 (codified at 28 U.S.C. § 1658 (Supp. 2003)), took a small step in the right direction by extending the extraordinarily short statute of limitations in private securities fraud cases to the earlier of two years after discovery of the fraud or five years after the violation occurs. Thus, frauds that remain hidden for many years still may escape legal redress, but this situation is somewhat ameliorated for frauds that remain hidden for an intermediate term.

[97] Sarbanes-Oxley Act § 906 (codified at 18 U.S.C. § 1350(a)-(b) (Supp. 2003)).

[98] *Id.*

[99] *Id.* Section 302 of the Sarbanes-Oxley Act, effective upon SEC rule-making, imposes additional certification requirements on CEOs and CFOs for each annual or quarterly report.

[100] Sarbanes-Oxley Act § 104(b)(1) (2002) (codified at 15 U.S.C.S. 7214 (Supp. 2003)).

[101] Moshe Adler, *Employees as Regulators*, N.Y. TIMES, Jan. 27, 2003, at A25.

[102] Sarbanes-Oxley Act § 104(b)(1) (codified at 15 U.S.C.S. 7214 (Supp. 2003)).

[103] Sarbanes-Oxley Act § 104(b)(2) (codified at 15 U.S.C.S. 7214 (Supp. 2003)).

prise inspections comparable to those conducted by health inspectors, who catch sight of cockroaches on the counters by walking into kitchens, unannounced, at any time.

It appears that the PCAOB will not make frequent inspections[104] nor will it make use of the element of surprise. It is thus highly unlikely that it will uncover fraud—at least not the complex fraud that exists today. In addition, it will be useless in ferreting out or deterring the everyday failures of auditors to unveil the misrepresentations that result from so-called judgment areas in accounting. These areas include maintaining adequate reserves for uncollectible receivables, anticipating future cost exposure from executory contracts, and assuring write-downs of unproductive assets; these are all areas in which it is easy for companies to inflate their balance sheets because management holds a high degree of discretion in valuations.

Fourth, Sarbanes-Oxley's requirement that the lead (or coordinating) audit partner be rotated[105] does not go far enough—there is no requirement that other key members of an audit team be rotated from time to time.[106] In addition, the auditor independence provisions of the new law only prohibit performing certain specified non-audit services contemporaneously with audit services;[107] other non-audit services still may be performed if approved by the company's audit committee.[108]

Finally, concerns were raised within hours after the passage of Sarbanes-Oxley that the Act's whistle-blower protection provisions[109] were being watered down when the White House issued an interpretation stating that the administration would provide job protection for employees cooperating with a House or Senate investigation.[110] According to Senator Patrick J. Leahy, the provisions were intended to apply to whistle-blowers

---

[104] *See id.* § 104(d)(1) (codified at 15 U.S.C.S. 7214 (Supp. 2003)). *But see id.* § 104(d)(3) (codified at 15 U.S.C.S. 7214 (Supp. 2003)).

[105] Sarbanes-Oxley Act § 203 (codified at 15 U.S.C. § 78j-1) (Supp. 2003)).

[106] The SEC rules issued January 28, 2003, require the lead and concurring partners on an audit to rotate out after five years, followed by a five-year time-out before returning to work on the company's audit, and thus "went beyond the Sarbanes-Oxley Act." Arthur Levitt, Jr., *The SEC's Repair Job*, WALL ST. J., Feb. 10, 2003, at A14. Former SEC Chairman Levitt noted that this "will go far in chilling the cozy relationships that can develop between auditors and their clients," but still asked whether auditing firms—not just partners—should be rotated, as "[a] fresh set of eyes may make sense." *Id.*

[107] Sarbanes-Oxley Act § 201 (codified at 15 U.S.C. § j-1 (Supp. 2003)).

[108] Sarbanes-Oxley Act §§ 201–202 (codified at 15 U.S.C. § 78j-1 (Supp. 2003)). Former SEC Chairman Levitt recently noted that "[o]ne area where the SEC's line-drawing has been questioned was its failure to ban firms from providing tax services for their audit clients." Levitt, *supra* note 106. Levitt wrote that he supports an outright ban on such work, but noted that new disclosure and audit committee oversight requirements increase the pressure on audit committees not to allow auditors to perform tax-shelter work that is still permitted under Sarbanes-Oxley. *Id.*

[109] Sarbanes-Oxley Act § 806 (codified at 18 U.S.C. § 1514A (Supp. 2003)). The whistle-blower protection applies only when an employee reasonably believes that the conduct reported constitutes a securities laws violation. *Id.*

[110] Allen, *supra* note 2.

who provide evidence of fraud to individual lawmakers even before an investigation is launched.[111]

Prophetically, according to a PricewaterhouseCoopers Management Barometer survey, only nine percent of executives surveyed in October 2002 characterized Sarbanes-Oxley as a good and adequate response to problems in accounting and reporting.[112]

More than a year after the Act's passage, the news remains full of new revelations of fraudulent practices. These recent issues include "whether [the New York Stock Exchange's former chief] Dick Grasso or the [E]xchange's board looted the NYSE in approving [Grasso's] $187.5 million pay and retirement package"[113]; revelations of "blatant governance conflicts at the [New York Stock] [E]xchange"[114] such as that "[f]or years the [exchange] has had a board and compensation committee composed of representatives from . . . member firms [regulated by the exchange]"[115]; revelations that Freddie Mac had understated income "by a least $4.5 billion"[116]; and the commencement of SEC charges against mutual fund managers and firms in the mutual funds "market timing" scandal.[117]

IV. Proposals for Reform

The accounting industry can no longer be a franchise without responsibility and accountability. The only reason to give accountants their franchise is so that they will fulfill a responsibility to help protect the investing public. The investment community has an important opportunity to ensure—through lobbying and other pressure—that intelligent, sophisticated changes are implemented to assure the proper functioning of our accounting watchdog system.[118] The remainder of this Essay sug-

---

[111] *Id.*

[112] Scott C. Newquist & Max R. Russell, Putting Investors First 141 (2003) (citing statistics contained in Press Release, PricewaterhouseCoopers, *Senior Executives Less Favorable on Sarbanes-Oxley* (July 23, 2003), *available at* http://www.barometersurveys.com/production/BarSurv.nsf/vwResources/PR_PDF_Files_2003/$file/mg030723.pdf.

[113] Susanne Craig & Kate Kelly, *Reed Plans NYSE Board Overhaul*, Wall St. J., Oct. 29, 2003, at A1.

[114] *Id.*

[115] *Id.*

[116] Patrick Barta & John D. McKinnon, *Fannie Mae Made $1.1 Billion Error in Its Accounting*, Wall St. J., Oct. 30, 2003, at A1.

[117] Deborah Solomon et al., *Milestone for "Timing" Scandal: Putnam is First Mutual-Fund Firm to Face a Regulatory Complaint; "It's about Cheating," Says Galvin*, Wall St. J., Oct. 29, 2003, at C1; John Hechinger et al., *Civil Charges are Expected in Fund Scandal*, Wall St. J., Nov. 3, 2003, at C1.

[118] Participants in the recent GAO Forum on Governance and Accountability "generally agreed that the demand side (investors and other users of financial information), has not been as involved as it needs to be to make financial reporting more meaningful and understandable . . . . Voluntary disclosures are rare and only in industries that demand this type of information . . . . Change is going to have to come from the demand side." U.S.

gests measures to increase the scope of audits; enhance regulatory oversight; increase the amount, meaningfulness, and usability of information companies and auditors provide to the public; improve auditor education; protect auditor independence; improve internal audits; and improve the legal accountability of auditors and companies.

A. The Scope of Audits

Fraud detection must be recognized as a responsibility of auditors. The scope of audits must be expanded, and looking for fraud must be made an affirmative audit obligation.

Recently, PricewaterhouseCoopers publicly embraced the need for fraud detection as part of the audit function.[119] The *New York Times* characterized such a change as risky for auditors, noting that such a course might expose them to increased liability.[120] Yet it is hard to imagine greater liability than the profession has already taken on by failing to increase its fraud detection responsibility.[121] If clients knew that audits required fraud detection, they would also know that they could not shop around for compliant auditors.

Auditors must go beyond the recently proposed Statement of Auditing Standards ("SAS") 99,[122] which creates an affirmative duty to perform only "minimum" procedures to detect material fraud, including making inquiries of management.[123] Beyond this limited duty, auditors need to ask probing questions, expand their inquiries beyond the most senior personnel, and develop and implement a variety of approaches to detecting fraud.

Fraud detection will require the improvement of many different aspects of the financial reporting process within individual companies. For example, the internal audit staff must be encouraged to act independently and trained to look into the areas of the company's operations that are most fraught with potential for wrongdoing. Audit committees must ensure that the reporting system requires the chief internal auditor to report directly to the audit committee of the Board of Directors. The PCAOB should develop a mechanism to oversee these developments. Meanwhile, auditors need to focus training on how to conduct examinations and fol-

---

GEN. ACCOUNTING OFFICE, COMPTROLLER GEN. OF THE UNITED STATES, GAO FORUM ON GOVERNANCE AND ACCOUNTABILITY, "CHALLENGES TO RESTORE PUBLIC CONFIDENCE IN U.S. CORPORATE GOVERNANCE AND ACCOUNTABILITY SYSTEMS," REP. NO. GAO-03-419SP, at 15 (2003) [hereinafter GAO FORUM].

[119] Jonathan D. Glater, *Pricewaterhouse Taking a Stand, And a Big Risk*, N.Y. TIMES, Jan. 1, 2003, at C1.

[120] *Id.*

[121] *Id.*

[122] CONSIDERATION OF FRAUD IN A FINANCIAL STATEMENT AUDIT, *supra* note 35.

[123] *Id.*

low specific inspection protocols,[124] especially in areas where there have been significant and frequent accounting abuses, such as improper revenue recognition, improper determinations of immateriality, and off-balance sheet debt for which the company is ultimately responsible (as in Enron).

One major area of abuse, premature income recognition,[125] has led to numerous instances of wrongdoing involving "side-letter agreements"—agreements with purchasing customers that if they are unable to resell a product, the original seller company will take it back with a refund.[126] Despite that contingency, the seller company immediately recognizes the sale as entirely complete. When a company sells a product but gives the buyer a right of return, the company may not record the revenue if the obligation is contingent on resale of the product.[127]

For instance, on an occasion where the audit system worked to an extent—albeit belatedly—database management company Informix was forced to restate overstated revenues of $311 million between 1994 and 1996 after auditors discovered side-letter agreements with customers.[128] Another infamous side-letter agreement was Kmart's secret arrangement with American Greetings Corporation in 2001, for which two Kmart executives were later charged with fraudulent reporting of revenue.[129]

---

[124] Forensic audits, favored by some participants in the recent GAO Forum, "would require that auditors undertake an attitudinal shift in their degree of skepticism and presume the possibility of dishonesty at various levels of management, including collusion, overriding of controls, and falsification of documents." GAO FORUM, *supra* note 118 at 19 & n.6.

[125] From January 1997 to June 2002, premature revenue recognition accounted for almost thirty-eight percent of 919 announced restatements, and was associated with $56 billion of the $100 billion in market capitalization lost by restating companies during the same period. GAO FINANCIAL RESTATEMENTS REPORT, *supra* note 2 at 5, 28. *See also* COFFEE, *supra* note 47, at 21.

[126] SEC. & EXCH. COMM'N, *supra* note 94, at 7, 9 n.23. The report found that eighty-one SEC enforcement matters during the five-year period from July 31, 1997 through July 30, 2002 involved improperly timed revenue recognition, including (1) holding books open after the close of a reporting period; (2) improper recognition of "bill and hold" sales, consignment sales, side-letter agreements, and other contingency sales; and (3) improper recognition of revenue from multiple element or bundled contracts. The same study identified twenty-five enforcement matters involving side-letter agreements.

[127] FINANCIAL ACCOUNTING STANDARDS BOARD, Statement of Financial Accounting Standards No. 48, Revenue Recognition When Right of Return Exists, par. 6 (Financial Accounting Standards Board, June 1981).

[128] GOODMAN, *supra* note 9, at 11.

[129] *See* MARCIA LAYTON TURNER, KMART'S TEN DEADLY SINS: HOW INCOMPETENCE TAINTED AN AMERICAN ICON 3 (2003) ("The [Kmart executives] agreed to make American Greetings its sole source of greeting cards and received a $42.3 million payment from the company at a time when it was struggling to meet Wall Street earnings expectations. Although American Greetings indicated to auditors that there were 'no strings attached' to the payment, there was actually a side agreement with Kmart that if their exclusive contract was rescinded, the payment had to be returned. That fact made it improper for Kmart to include the payment as revenue when it did. Kmart's financial statements from the first three quarters of 2001 were impacted by these maneuvers, which made the company's financial health seem stable when it clearly was not.").

Under current standards, an auditor must obtain sufficient competent evidence through inspection, observation, inquiries, and confirmations to afford a reasonable basis for an opinion regarding the financial statements under audit.[130] Yet, particularly in light of the SEC's finding that auditor violations resulted largely from auditors' failing to gain sufficient evidence to support the issuer's accounting,[131] we must re-examine the auditing requirements for confirmation of completed sales to ensure that auditors can ferret out these and other abuses.

In their examinations of areas that have been prone to abuse, auditors can no longer rely upon the integrity of third parties from whom confirmations are solicited. Enron and other cases involving financial fraud have indicated that the economic power of the audit client over customers and suppliers can be so strong that it induces counterparties to supply false information to auditors. Legislation can help the auditors avoid the consequences of such wrongful conduct by imposing severe civil and criminal penalties upon any third party who supplies an auditor with false information.

Also, when there are specific practices that are known to be subject to abuse in a certain industry, such as swap transactions, derivatives, and other hard-to-understand transactions, the audit file should be required to reflect the assumptions and beliefs underpinning the auditors' assessments of the items they are auditing. The auditors should then report to the public on their understanding of the substance of the transactions and the accounting methodologies employed.

When auditors evaluate a client's valuations for corporate transactions, they should carefully examine companies' valuation practice in previous similar transactions. Sometimes the reason that accountants accepted a particular practice in the past was that the amount involved was immaterial to the company's overall financial picture. Yet continuing the practice when the amounts increase in size can cause material misstatements, so evaluations of these practices must be revisisted. Auditors also need to focus on qualitative factors as well as quantitative factors in order for the financial statements not to be misleading, and in order to make fair determinations of materiality.[132]

Outside auditors must also assess the quality of work of the Board of Directors' audit committee. The more active and thorough the audit committee, the greater the assurance of management oversight and its effectiveness. To determine the quality of the audit committee's work, the outside auditor should prepare a checklist tailored to the client's operations assessing the independence, financial knowledge, and integrity of

---

[130] CODIFICATION OF ACCOUNTING STANDARDS AND PROCEDURES, Statement on Auditing Standards No. 31, § 326.01 (American Inst. of Certified Pub. Accountants 1980).

[131] Sec. & Exch. Comm'n, *supra* note 94, at 7.

[132] *See* CONSIDERATION OF FRAUD IN A FINANCIAL STATEMENT AUDIT, *supra* note 35 § 13 (advising auditors to regard clients' stated information skeptically).

the Board of Directors. For example, if a company has an important research and development project where the public has been led to believe that new products will be ready for market within a certain time period, a checklist item should be whether the audit committee interviewed the project's senior technicians to obtain their assurances that the timetables are reasonable. Furthermore, the auditor should perform background checks of each member of the audit committee and the rest of the Board of Directors.

B. Regulatory Oversight

Enhanced regulatory oversight by the PCAOB would be an important mechanism for preventing abuses. The PCAOB should create a notice to be included in all confirmation requests sent by auditors advising those whose account verifications are being sought that supplying false information—as proposed above—is a crime and subject to severe penalty. In addition, the notice should advise that the PCAOB staff will perform its own periodic investigations into the truth and adequacy of confirmations.

The PCAOB should compare how companies in the same industry account for similar items. For instance, the PCAOB could check the amounts and percentages of reserves established by competitors in an industry; anomalies would be red flags indicating possible abuse.

GAAP and GAAS should be maintained by an organization not dominated by the accounting profession and industry. Instead, the PCAOB should use the powers granted to it by Sarbanes-Oxley[133] to make critical GAAP and GAAS determinations.

Pursuant to the powers granted to the Board to establish record retention rules for auditors,[134] the PCAOB should mandate that auditors cannot destroy *any* document created or used by auditors in the scope of an audit, including items such as "to do" notes.

A system should be implemented to monitor inquiries from the audit staffs to technical personnel of audit firms. All such requests should be sent to the PCAOB and both the inquiry and the responses should be reviewed. The PCAOB should have a technical staff dedicated to investigating the underlying situations giving rise to such audit-field-staff requests and the responses thereto.

The PCAOB should have—and exercise—the teeth necessary to sanction individual auditors and audit firms for improper audits.

---

[133] Sarbanes-Oxley Act § 103 (codified at 15 U.S.C.S. 7214 (Supp. 2003)). *See also* NEWQUIST, *supra* note 112, at 142 ("[U]nless the new Board uses its [standard setting] powers well, . . . its ability to impose and enforce penalties will quickly diminish.").

[134] Sarbanes-Oxley Act § 104(e) (codified at 15 U.S.C.S. 7214 (Supp. 2003)).

C. Information Provided to the Public

Far more information—in understandable form—needs to be provided to the users of financial statements. Information about the aging of accounts receivables should be provided to help investors, and analysts advising investors, make judgments about the collectability and quality of companies' accounts receivables. Similarly, information on stale inventory, that is, inventories remaining after the normal turnover cycles, would provide a useful measure of the health of corporate sales.

In addition, more information on reserves, including the share of reserves that have been established against general portfolios and the share of reserves established against specific accounts of concern, and the basis for establishing the reserves, will help investors and analysts make judgments regarding the adequacy of companies' reserves. The reserve percentage being used for each industry sector in which a company operates would also be helpful.

Investment and interest rate assumptions should be provided wherever a company's financials employ such assumptions. Such information is critical for determining pension plan adequacy and a variety of other measures of corporate financial health. This is especially true of companies such as insurance companies that use interest rate assumptions for actuarial calculations to determine or estimate the need for further reserves or to calculate estimated future values.

Assumptions concerning the technological viability of equipment should be provided to enable financial statement users to examine whether depreciation schedules are appropriate.

When revenue is recognized prior to receipt of income for a recognized item, companies should be required to state the basis for the earlier recognition of revenue. Disclosures should describe which items are being recognized early. Among other things, such information can help analysts assess the impact on future revenues of the early income recognition.

Assumptions used to determine whether items are material should also be disclosed. Materiality determinations are often critical, and can mean the difference between a company's missing or "making" its numbers.

Financial reports should illuminate "gray areas"—those areas where management's discretionary judgment was used. Accountants also should be obliged to seek out and disclose publicly much more information about these accounting "gray areas," because they are highly subject to manipulation. Often a principle is clear, such as the need for setting up reserves for doubtful accounts, but a judgment call needs to be made as to the extent of the reserve that is set up. Other examples of areas involving material judgment are percentage of completion accounting methods, and accounting for derivative instruments, loss reserves, and

reserves for inventory and receivables. Judgment calls all too often make the difference between a corporation showing a profit or a loss. Better disclosures by management are needed to understand the true financial health of a company.

Currently, auditors must assess the adequacy of such disclosures in light of the circumstances and facts known to the auditors at the time of the audit.[135] However, auditors should assess the adequacy of such disclosures in light of auditors' own investigations as well. They could then disclose the aggressiveness of the company's practices compared to industry practices, ranges of acceptable practices, and risks. For instance, auditors should examine how many receivables are past due beyond the normal cycle and for how long, which loss reserve percentages are being used for general portfolio, and so forth.

It also should be clearly understood that an auditors' duty is to the investing public. According to the *New York Times*, Auditor-Trak, a publisher of accounting industry data, found that 348 accounting firms resigned from clients in 2002, up from 286 in 2001.[136] The question that must be asked is, once an auditor distrusts the audit process, why should the auditor have the right to resign?

Auditors are currently required, if management has not provided reasonable justification for a change in accounting principles, to note in an audit opinion that the change was made without reasonable justification.[137] This requirement does not go far enough. If auditors resign, the reasons for the resignation should be disclosed to the public. Auditors should be required to forfeit fees if they fail to provide the results of the audit to the company's investors. The shareholders hired the accountants to do the audit. By resigning, auditors deprive the investing public of the results of this work, though the shareholders are ultimately paying.

If the client company interferes with an audit, the auditors should disclaim an opinion and disclose the basis for the disclaimer.

D. Auditor Education

To enhance auditors' ability to detect fraud, special training must be given and tools developed to improve the detection of markers of possible wrongful behavior.[138]

---

[135] *See* CODIFICATION OF ACCOUNTING STANDARDS AND PROCEDURES, Statement on Auditing Standards No. 32, § 431.02 (American Inst. of Certified Pub. Accountants 1980).

[136] Glater, *supra* note 119.

[137] CODIFICATION OF ACCOUNTING STANDARDS AND PROCEDURES, Statement on Auditing Standards No. 58, § 508.57 (American Inst. of Certified Pub. Accountants 1989).

[138] The inadequacy of auditor training in fraud detection has long been noted. *See* Roohani et al., *supra* note 55, at 72 ("There has been little effort in teaching or training accountants to detect financial fraud. Detecting fraudulent financial transactions that took place during the late 1980's by some high flying financial institutions requires special education and training beyond the scope of traditional accounting study.").

The first order of business must be education that permits the auditor to understand the industry in which the company being audited is materially active. Only by understanding the industry and the audit client's competitors will detection of aberrant results (such as unusual inventory turnover) become possible. Requiring a better understanding of the client's business and industry is imperative.[139]

As transactions become increasingly complex, they become more difficult for observers to understand. If observers do not have a proper understanding of the underlying transactions, the implications of exotic transactions such as activities in the futures markets and the off-balance-sheet activities that we recently observed in Enron will not be brought to light. Therefore, there must be people within the auditing firm or consultants hired by them who can completely understand the nature of those financing arrangements and the true economic meaning of the transactions in order to be able to render an opinion with respect to them. This has been a material shortcoming in the audit world in the past.

E. Auditor Independence

Further steps need to be taken to improve auditor independence. If an auditing partner is receiving a major portion of his or her income from a particular client, the auditing firm should ensure that a senior regional staff person reviews the audit.

Other steps to enhance and assure auditor independence include stopping the revolving door: clients should not be permitted freedom to hire personnel from their auditors without restraint. Corporate personnel hired from a company's outside auditing firm are often involved in manipulations because they know how to pressure their former colleagues into compliance. This problem was recently acknowledged by the court in the suit brought by Enron's shareholders against Arthur Andersen.[140] In her decision sustaining the claims against Arthur Andersen, Judge Harmon noted that

> [the] complaint points to more than three hundred accounting and finance positions at Enron, many in mid-level and senior management, that were filled with former Arthur Andersen auditors and professionals. The complaint comments that Enron De-

---

[139] Joseph T. Wells, *The Fraud Examiners*, 10-03 J. ACCT. 76, ¶ 3 (2003) *available at* http://www.aicpa.org/pubs/jofa/oct2003/wells.htm ("[B]ecoming an antifraud expert doesn't take months—it takes years. Formal education in the fraud examination field is new and limited—but growing . . . . Until four years ago, only 19 colleges out of 900 offered a class in fraud examination. That number now exceeds 150, with 200 more schools planning to add a course shortly. But for the time being, it is necessary for most budding fraud examiners to learn this craft on the job.").

[140] *In re* Enron Corp. Sec., Derivative & ERISA Litig., 235 F. Supp. 2d 549, 674 (S.D. Tex. 2002).

> fendants were comfortable with this fact because they knew the Arthur Andersen auditors were less likely to question improper accounting if done by their former co-workers and bosses, who were now officers and managers at Enron.[141]

Similarly, a leading financial commentator concluded that improper practices at Waste Management, including shifting current expenses to later periods and depreciating assets over excessively long periods to reduce current-year depreciation expenses, which "should have popped up as red flags[,] . . . were not disclosed because of the close relationship between management and the auditing firm itself . . . . Until 1997, every CFO and chief accountant at Waste Management had been recruited from the ranks of Arthur Andersen auditors."[142]

Rotation of auditing personnel would help alleviate this problem. To maintain continuity, not all personnel need be rotated at one time. For instance, the manager on an audit can be rotated every three or four years, and the partner can be rotated on a staggered or alternate three- or four-year schedule.

Recent auditor independence rules proposed by the SEC for comment involve yet another possibility—fixed auditor terms.[143] The Sarbanes-Oxley Act authorized a study of this proposal, now being undertaken by the General Accounting Office.[144] This approach to accounting independence would require a public company to retain its auditor for a fixed term of five or seven years with no right to terminate and change auditors at the end of each fixed term.[145] Former SEC Chairman Harold Williams testified that an advantage of this approach was that "the auditor would be assured of the assignment and, therefore, would not be threatened with the loss of the client and could exercise truly independent judgment."[146]

Auditor independence can also be enhanced by forbidding auditors from playing another dual role that recently received attention after Sprint Corporation's auditor, Ernst & Young, sold several top Sprint executives questionable tax shelters.[147] Auditors should be prohibited from advising both a company and any of its executives on any matter. This issue was left untouched by Sarbanes-Oxley.

---

[141] *Id.*

[142] GOODMAN, *supra* note 9, at 10.

[143] Sarbanes-Oxley Act § 203.

[144] *Id.* § 207.

[145] *Id.* § 203.

[146] *Accounting and Investor Protection Issues, supra* note 3, at 141 (prepared statement of Harold M. Williams, Former Chairman, Securities & Exchange Commission).

[147] Jonathan D. Glater & Stephen Labaton, *Auditor Role in Working for Executives is Questioned*, N.Y. TIMES, Feb. 8, 2003, at C1; Rebecca Blumenstein & Carol Hymowitz, *Inside the Tough Call at Sprint: Fire Auditor or Top Executives?*, WALL ST. J., Feb. 10, 2003, at A1.

F. Internal Audits

There is also a need to press for stronger internal audits by auditors who report directly to members of a company's audit committee. Audit committees must ensure that the internal auditors who are trained in fraud detection have the freedom to roam the company and choose their targets as they see fit. It may also be useful to explore a recent suggestion that employees be represented on audit committees, inasmuch as employees have a real stake in their companies' long-term future.[148] As SEC Chief Accountant Robert K. Herdman said last year, "the role of the audit committee is central to insuring the integrity of published financial statements on which investors rely."[149]

G. Legal Accountability

The certified statements that CEOs and CFOs are now required to sign under Sarbanes-Oxley[150] are unlikely to have much effect.[151] Improvements in legal accountability are far more likely to have an impact.

There is an urgent need to reinstate aiding and abetting as a ground for accountants' legal liability.[152] If only those who are "primarily" involved in securities fraud are subject to liability, the real purpose of the securities law—the protection of investors—will be thwarted. As the massive frauds in the years since *Central Bank* eliminated aiding and abetting liability demonstrate, removing the legal disincentives for auditors to enable fraud, or to fail to take reasonable steps to ferret out fraud, has had serious consequences for the investing public.

Finally, it is time to rectify the mistakes made when Congress passed the Private Securities Litigation Reform Act, and to return to shareholders the same procedural rights that are available to litigants in other cases, such as the right to discovery at an early stage of the case, so that accounting fraud can be uncovered. The automatic stay of discovery imposed by the PSLRA whenever any defendant moves to dismiss[153] is highly unfair to fraud victims, given the already significant (and unfair) advantages provided to defendants under the PSLRA. These defendant advantages include heightened pleading standards,[154] safe-harbor protec-

---

[148] *See* Adler, *supra* note 101.

[149] SEC Chief Accountant Robert K. Herdman, Making Audit Committees More Effective, Speech at the Tulane Corporate Law Institute (Mar. 7, 2002), *available at* http://www.sec.gov/news/speech/spch543.htm.

[150] Sarbanes-Oxley Act § 302.

[151] *See, e.g.*, NEWQUIST, *supra* note 112, at 141 ("[T]he benefit of the provision [of Sarbanes-Oxley] requiring that CEOs personally certify financial statements is primarily psychological. In truth, they've always had that responsibility.").

[152] *See supra* notes 72–76 and accompanying text.

[153] Pub. L. No. 104-67, § 27D(b)(3) (codified at 15 U.S.C. § 78u-4(b)(3) (Supp. 1995)).

[154] *Id.* § 21D(b)(1) and (2) (codified at 15 U.S.C. § 78u-4(b)(1) and (2) (Supp. 1995)).

tions,[155] elimination of joint and several liability,[156] proof of loss causation requirements,[157] and more difficult damages standards and limitations.[158]

V. Conclusion

The best deterrence is the adoption of truly meaningful and achievable methods for holding malefactors accountable. The investing public has learned at too great a cost that self-regulation, government regulation, and our criminal justice system cannot or will not do the job. The government regulatory system will always be underfunded. Criminal cases are rare because they are too hard to prove, and law enforcement resources are needed elsewhere. Investors need to be able to hold wrongdoers accountable. The measures proposed here would give investors the tools to do so.

---

[155] *Id.* § 27A and § 21E (codified at 15 U.S.C. § 78u-5 (Supp. 1995)).
[156] *Id.* § 21G (codified at 15 U.S.C. § 78u-4(f) (Supp. 1995)).
[157] *Id.* § 21D(b)(4) (codified at 15 U.S.C. § 78u-4(b)(4) (Supp. 1995)).
[158] *Id.* § 21D(e) (codified at 15 U.S.C. § 78u-4(e) (Supp. 1995)).

ARTICLE

ACCOUNTING FOR SOCIAL SECURITY AND ITS REFORM

HOWELL E. JACKSON*

How well did the Social Security system do last year? According to the most recent annual report prepared by the system's Board of Trustees, the Social Security trust funds showed a $165.4 billion net increase in assets in 2002 and reported accumulated reserves of nearly $1.4 trillion by year end. Unfortunately, these glowing reports are a cash-flow illusion, revealing only the differences between the system's annual cash receipts and its yearly payments for benefits and administrative expenses. In this Article, Professor Howell Jackson argues that the United States should measure Social Security through an accrual accounting method, which recognizes commitments to make future benefit payments when those obligations are actually incurred. Were the finances of the Social Security system restated under principles of accrual accounting, the Social Security trust funds would have had to report a loss of several hundred billion dollars in 2002. Moreover, as of December 31, 2002, an accrual-based balance sheet of the Social Security system would have revealed more than $14.0 trillion of accrued liabilities to Social Security participants and beneficiaries. Even allowing for the system's $1.4 trillion of accumulated reserves as well as the value of excess future taxes to be paid by current participants over the rest of their working lives, the Social Security trust funds had unfunded obligations on the order of $10.5 trillion as of year-end 2002.

In addition to misrepresenting the magnitude of the Social Security system's looming financial crisis, the current accounting system for Social Security distorts public debate over Social Security reform proposals and confuses the relationship between Social Security and the rest of the federal budget. Accrual accounting, in contrast, would provide a clearer picture of the true state of the Social Security system's current financial shortfall and the extent to which the system's burden on future generations is increasing each year. Accrual accounting would also create political incentives for political leaders to address Social Security's difficulties in a timely manner, and enhance the quality of public debate over the relative merits of competing reform proposals.

---

* Finn M. W. Caspersen & Household International Professor of Law, Harvard Law School. M.B.A. & J.D., Harvard University, 1982; A.B., Brown University, 1976. E-mail: hjackson@law.harvard.edu. I would like to thank Robert Clark, Peter Diamond, Elizabeth Garrett, Michael Graetz, Daniel Halperin, John Langbein, Jeffrey Liebman, Elena Kagan, Louis Kaplow, Maya MacGuineas, Olivia Mitchell, Peter Orszag, Robert Pozen, Mark Roe, Sylvester Schieber, Dan Shaviro, Kent Smetters, A. Haeworth Robinson, Sheila Weinberg, and David Wilcox for helpful comments and suggestions, as well as participants in the Law & Economics Seminar at Harvard Law School, the Shirley A. Webster Lecture at the University of Iowa School of Law, the Regulatory Policy Seminar at the Harvard Kennedy School of Government, and the Stanford Institute for Economic Policy Conference on U.S. Budget Policy and Practice. Steven Goss and his staff at the Social Security Administration's Office of the Chief Actuary were extremely generous in supplying data for this Article. Tony Feuerstein, Charles Gray, and Wang Xin provided valuable research assistance. The Olin Center for Law, Economics and Business and the Harvard Law School Summer Research Program provided funding for research on this Article. An abbreviated and simplified version of some of the ideas presented here appeared in Howell Jackson, *It's Even Worse Than You Think*, N.Y. TIMES, Oct. 9, 2003, at A37. A .pdf version of this Article is available at http://www.law.harvard.edu/students/orgs/jol.

Following this Article, economists, political scientists and legal scholars comment on Professor Jackson's proposal. They consider its merits, costs, and implications, focusing on its short- and long-term accuracy, effectiveness in conveying meaningful financial information, applicabilitity in the context of other government programs, and viability in comparison to alternative metrics for evaluating Social Security.

TABLE OF CONTENTS

I. A Critique of the Current Accounting System for Social Security 68
- *A. Analysis of Accounting for the Trust Funds* 69
  - *1. Standard Presentation* 69
    - *a. Short-Range Projections* 70
    - *b. Long-Range Projections* 72
  - *2. Criticisms of the Standard Presentation* 78
    - *a. Short-Range Projections* 78
    - *b. Long-Range Projections* 83
    - *c. Biases with Respect to Reform Proposals* 84
- *B. Consolidation Within the Federal Budgtet Process* 86
  - *1. The Current Approach to Consolidating Social Security Trust Funds* 87
  - *2. Criticisms of Standard Presentation* 91
    - *a. Misleading Features of the Standard Presentation* 92
    - *b. Familiar Criticisms of Relying on Total-Budget Aggregates* 93
      - *i. Proper Characterization of Trust Fund Reserves* 93
      - *ii. The Economic Realities of Social Security* 95
    - *c. A Deeper Critique* 96
    - *d. Biases for Reform Proposals* 98
      - *i. Perverse Budgtetary Incentives for Social Security Reforms* 98
      - *ii. Distorting Effects on Reforms That Change Trust Fund Operations* 99
      - *iii. Generating Reform Proposals with Ulterior Motives* 100

II. An Accrual Accounting System for Social Security 101
- *A. Possible Components of an Accrual-Based System of Accounting* 101
- *B. A GAAP-Style Accrual Accounting System for Social Security* 104
  - *1. Balance Sheet* 105
  - *2. Income Statement* 109
- *C. Comparing Accrual-Based Financial Statement with Other Measures of Long-Term Solvency* 112

1. *Closed-Group Unfunded Obligation* 112
2. *Open-Group Unfunded Obligation Seventy-Five-Year Horizon* 114
3. *Open-Group Unfunded Obligation: Infinite Horizon* 116
4. *GAAP-Style Accrual Accounting v. Alternative Measures of Long-Term Solvency* 117
   a. *Distinguishing Benefits Earned-to-Date from Total Benefits* 118
   b. *The Inclusion of Benefits-To-Be-Earned and Future Taxes* 119
5. *Modified Accrual Accounting* 120

D. *Accrual Accounting and the Unified Federal Budget* 123
1. *Full Consolidation Within the Federal Budget* 123
2. *An Unconsolidated Approach to Social Security Accounting* 126
   a. *Budgetary Aggregates* 127
   b. *Interest Payments and Other General Revenue Contributions* 127
   c. *Measures of Public Debt Outstanding* 127
   d. *Separate Financial Statements for Trust Funds* 129

III. *Accrual Accounting and Social Security Reform* 130

A. *Clarifying the True Nature of Social Security Finances* 132
1. *Debunking the Debilitating Myth of Current Surpluses* 132
2. *Focusing Public Attention on the Trust Funds' Mounting Implicit Debt* 132
3. *Developing Appropriate Targets for the Implicit Debt of Social Security* 134

B. *Enhancing the Quality of Public Debate over Reform Proposals* 135
1. *Exposing Short-Range Cash-Flow Effects* 136
2. *Recognizing the Value of Reforms with Future Cash-Flow Effects* 136
3. *Flagging Reform Proposals That Expand Accrued Liabilities* 137
4 *Eliminating the Cliff Effect* 137
5. *Identifying Back-loaded Reform Proposals* 137
6 *Flagging the Amount of Excess Future Taxes* 138

C. *Altering the Optics of Reform Proposals* 138
1. *Distinguishing Accrued Rights of the Elderly and Other Participants* 138
2. *Differentiating Prospective and Retroactive Effects* 139
   a. *The Possibility of Prospective Benefit Cuts* 139
   b. *More Modest Methods of Enhancing Benefits* 140
3. *Changing Expectations of the General Public* 140

D. *Accrual Accounting and Individual Accounts* 140

1. Posing the Problem of Transition Costs *140*
2. Eliminating Budgetary Anomalies *141*
3. Comparing Traditional Benefits and Individual Accounts *141*
E. Social Security Reform, the Federal Fisc, and National Savings *142*
1. Budgetary Effects of Fully Unconsolidated Treatment *142*
2. Budgetary Effects of General Revenue Contributions *142*
3. Tracking Gross Public Debt and Unfunded Trust Fund Liabilities *143*
F. Substantive Implications of Accrual Accounting *143*
IV. Postscript on Normative Baselines *144*
A. Accrual Accounting and the Federal Government *146*
1. Accrual Accounting in General *146*
2. Accrual Accounting for Social Insurance *148*
B. The Estimate of Public Pension Plan Obligations in Other Contexts *151*
1. Measures of National Public-Pension-Plan Obligations *152*
2. Economic Estimates of the Value of Social Security Benefits *155*
a. Life Cycle Savings Literature *156*
b. Household Wealth Literature *157*

This Article proposes a change in the way we account for Social Security.

The financial problems of Social Security are frequently debated on editorial pages and in policy circles.[1] For almost a decade, politicians and analysts have been warning of a looming crisis, and numerous commissions and pamphleteers have advanced a host of solutions to the system's problems, from reductions in benefits to increased taxes to a range of privatization proposals.[2] Though much ink has been spilled and many speeches given, Social

---

[1] *Compare* Felix Rohatyn, *Free, Wealthy & Fair*, WALL ST. J., Nov. 11, 2003, at A18 ("Our dependence on foreign capital and our budget deficits must be reduced by greater fiscal discipline at home including the politically painful reform of entitlements such as Social Security."), *with* Robert M. Ball, *Social Security Needs Tweaking, Not 'Reform,'* BOSTON GLOBE, Oct. 11, 2003, at A19 ("Social Security doesn't need to be 'reformed' because it hasn't failed. The system just needs some timely maintenance work.").

[2] Two prominent reform proposals, discussed at various points in this Article, appeared in 1994–1996 ADVISORY COUNCIL ON SOC. SECURITY, REPORT OF THE 1994–1996 ADVISORY COUNCIL ON SOCIAL SECURITY: FINDINGS AND RECOMMENDATIONS (1997) [hereinafter 1994–1996 ADVISORY COUNCIL REPORT] and THE PRESIDENT'S COMM'N TO STRENGTHEN SOC. SECURITY, REPORT OF THE PRESIDENT'S COMMISSION: STRENGTHENING SOCIAL SECURITY AND CREATING PERSONAL WEALTH FOR ALL AMERICANS (2001) [hereinafter BUSH COMMISSION REPORT]. For an overview of various approaches to Social Security reform, see generally GEN. ACCOUNTING OFFICE, SOCIAL SECURITY: DIFFERENT APPROACHES FOR ADDRESSING PROGRAM SOLVENCY (July 1998); *see also* Kathryn L. Moore, *Privatization of Social Security: Misguided Reform*, 71 TEMP. L. REV. 131, 139 (1998).

Security has undergone no fundamental changes in more than twenty years.[3] As of this writing, there is no prospect for serious reform until at least the completion of the next presidential election and, most likely, not even then.

For a variety of reasons, politicians have difficulty addressing problems like Social Security reform. The stakes are high, both financially and politically. The primary beneficiaries of the program are the elderly, who are both politically active and well-represented.[4] The most obvious solutions—raising taxes or cutting benefits—are politically unpalatable, perhaps even toxic.[5] While stock market investments appeared to offer a painless fix for at least part of the system's difficulties a few years ago, subsequent volatility in stock prices has deflated public support for reforms of this sort.[6] For politicians, Social Security presents a problem with substantial political risk and no easy answers.

But the current accounting system for Social Security also contributes to the current political gridlock. Almost all public discussion of Social Security finances are based on annual reports that the Social Security Board of Trustees publishes each spring.[7] The trustees' reports are, in essence, statements of annual cash flows, comparing the system's annual cash receipts to its yearly payments for benefits and administrative expenses. This format highlights the system's current annual "surpluses" and its mounting reserves. These favorable financial reports undermine efforts to address Social Security reform proposals by locating the Social Security crisis far in the future. If, however, the financial statements of the Social Security system were prepared in accordance with the principle of accrual accounting—under which commitments to make future benefit payments are recognized when those obligations are incurred—the public would understand that promises being made today to future retirees are the cause of Social Security's financial problems.

Consider, for example, the reported financial performance of the Social Security trust funds in 2002. According to the trustees' report, the combined Social Security trust funds ran a cash-flow surplus of $165.4 billion in 2002

---

[3] For an overview of the political process that led up to 1983 reforms, which included a combination of both benefit reductions and tax increases, see generally PAUL LIGHT, ARTFUL WORK: THE POLITICS OF SOCIAL SECURITY REFORM (1985).

[4] *See* John Lloyd, *'Sans Nothing,'* FIN. TIMES, Apr. 19, 2003, at P1 (noting the political power of elderly voters and the organizations that represent them).

[5] *See, e.g.*, Susan Page, *Social Security Debate May Be Ready to Ignite*, USA TODAY, Dec. 3, 2002, at A11.

[6] *See, e.g.*, Claudia Deane & Dan Balz, *GOP Puts Stock in 'Investor Class,'* WASH. POST, Oct., 27, 2003, at A1 ("Bush's plan for partial Social Security privatization, for example, was blunted by the drop in the market, and many congressional Republicans remain wary of promoting the idea . . . ."); Janet Kidd Stewart, *Jilted Shareholders Rethink Faith in Stocks*, CHI. TRIB., Sept. 30, 2001, at 1.

[7] *See, e.g.*, Kathy M. Kristof, *Averting the Crisis in Social Security*, L.A. TIMES, Mar. 30, 2003, at C3 (citing most recent Trustees Report for information on financial condition of Social Security); John Attarian, *Time is Running Out to Fix Medicare, Social Security*, DETROIT NEWS, June 26, 2003, at 17 (same).

and accumulated record levels of reserves, equal to $1.4 trillion, by the end of the year.[8] These financial facts, which are widely reported in the popular press and well known to political leaders,[9] imply that the financial position of the Social Security trust funds at the end of 2002 was quite good.[10] While the trustees' longer-term projections suggest that Social Security will face serious problems several decades in the future, the short-term outlook is so favorable that there is little incentive for political leaders to expose themselves to the potentially painful choices necessary to address the system's problems by altering benefits or raising revenues.

If, however, the trust fund accounts were presented under a system of accrual accounting, public perceptions of the problems facing Social Security would change dramatically. Restated under accrual accounting, the Social Security trust funds would have had to report a loss on the order of several hundred billion dollars for 2002 and comparable or even larger annual losses for much of the last decade.[11] A Social Security trust funds balance sheet prepared in accordance with principles of accrual accounting would show accrued liabilities of approximately $14 trillion at the end of 2002—that is, more than ten times the system's current reserves of $1.4 trillion.[12] In other words, as of December 31, 2002, the Social Security trust funds had unfunded accrued liabilities of $12.6 trillion, or 122% of the country's gross domestic product (GDP).[13] Even if one factors in the present value of excess taxes that current participants will pay into the Social Security trust funds over the balance of their working lives, the system's current unfunded obligations are on the order of $10.5 trillion.[14]

Moving Social Security to an accrual-based accounting system would greatly enhance the quality of debate over Social Security reform proposals. In a variety of ways, cash-flow accounting biases public debate over Social Security reform proposals, encouraging politicians to obscure the extent of Social Security's mounting liabilities and favoring certain kinds of reform, particularly those that increase short-term cash-flow surpluses while simultaneously increasing the system's long-term fiscal obligations.[15] Cash-flow accounting also makes certain kinds of reform proposals—notably some individual-account proposals and any reforms that include investments in private capital markets—look prohibitively expensive.[16] An accrual-accounting sys-

---

[8] *See* 2003 TRS. OF THE FED. OLD-AGE & SURVIVORS INS. & DISABILITY INS. TRUST FUNDS ANN. REP. 2 [hereinafter 2003 TRUSTEES REPORT]. *See infra* text accompanying notes 26–31.

[9] *See, e.g.*, Allan Sloan, *Washington Talks a Blue Streak As Social Security's Red Ink Runs*, WASH. POST, Mar. 25, 2003, at E3 (citing 2003 Trustees Report for level of current reserves and annual cash flow surpluses).

[10] *See infra* text accompanying notes 26–38, note 80.

[11] *See infra* text accompanying notes 159–165.

[12] *See infra* text accompanying notes 152–158.

[13] *See infra* text accompanying notes 152–158.

[14] *See infra* Part II.C.5.

[15] *See infra* text accompanying notes 128–129.

[16] *See infra* text accompanying note 86. *Cf.* Robert C. Pozen, *The Virtues of Moving From Cash to Accrual Accounting for Social Security*, 41 HARV. J. ON LEGIS. 199 (2004).

tem would clarify both the current status of Social Security's finances and the relative merits of competing reforms.

Another benefit of adopting accrual accounting for Social Security is the effect that such a change would have on the country's overall fiscal policy. Currently, the federal government regularly includes annual trust-fund surpluses in the budgetary aggregates reported to the general public. Even though the trust funds are supposed to be "off-budget," current practice allows the federal government to use annual trust fund surpluses to support other spending programs and disguise the growth in financial obligations that future generations will have to shoulder.[17] Were Social Security to restate its accounts in accordance with the principles of accrual accounting, politicians would be less likely to use trust funds' reserves in this way and the public would gain a clearer picture of the country's mounting financial commitments.

This Article is divided into four Parts. Part I presents a critique of the current system of accounting for Social Security and advances the claim that accrual accounting offers a valuable alternative framework for evaluating the annual performance and overall solvency of the Social Security system. The argument focuses first on the traditional accounting treatment of the Social Security trust funds and then considers the relationship between those trust funds and the overall federal budget. In both contexts, current practices are inherently misleading and distort the debate over Social Security reform proposals in important ways.

Part II presents an alternative framework for Social Security accounting based on principles of accrual accounting. After reviewing the different kinds of assets and liabilities that might be recognized on a set of financial statements for Social Security, this Part sketches out a system of balance sheets and income statements for the trust funds in 2002 using an accrual-based system of accounting. It then compares this system of accrual accounting with alternative measures of long-term solvency that experts sometimes use to assess Social Security solvency, and explores the relative merits of various approaches. Based on this analysis, it proposes a modified system of accrual accounting of Social Security that incorporates the strengths of both Generally Accepted Accounting Principles (GAAP)-style accrual accounting and alternative measures of long-term solvency. Part II concludes with a brief discussion of how such a modified system of accrual accounting

[17] *See infra* text accompanying notes 87–101.

BOX ONE

THE RISING CONCERN OVER THE IMPLICIT DEBT OF THE FEDERAL GOVERNMENT

Over the past year, commentators and government officials have begun to take a greater interest in the implicit obligations of the federal government. For example, in a recent study, Gokhale and Smetters have estimated the government's fiscal imbalance to be $44.5 trillion at year-end 2002, of which they attribute a $7.0 trillion long-term fiscal imbalance to Social Security. *See* JAGADEESH GOKHALE & KENT SMETTERS, FISCAL AND GENERATIONAL IMBALANCES: NEW BUDGET MEASURES FOR NEW BUDGET PRIORITIES (AEI Pamphlet No. 4, Apr. 2003); *see also* LIQUN LIU ET AL., MEANINGFUL MEASURES OF FISCAL DEFICIT AND DEBT: THE CASE FOR INCORPORATING ENTITLEMENT DEBTS (Texas A&M Private Enterprise Research Paper No. 2010, May 2002). The federal government itself has begun to compile and report similar measures of long-run shortfalls, *see* 2002 FIN. REP. OF THE UNITED STATES 62–65 (reporting $4.6 trillion unfunded obligations of Social Security over the next seventy-five years); ANALYTICAL PERSPECTIVES: BUDGET OF THE UNITED STATES GOVERNMENT: FISCAL YEAR 2004, at 45–49 (similar), and even the most recent Social Security trustees' report has supplemented its reports to include additional measures of the system's long-run unfunded obligations. *See infra* text accompanying notes 166–185. These recent developments reflect concerns for fiscal imbalance similar to the ones that motivated this Article. These broader analyses also indicate that the problems of Social Security are not unique. Other federal programs, most notably Medicare and other retiree health programs, are also seriously underfunded over the long run. Accordingly, aspects of the accrual accounting proposals for Social Security advanced in this Article might profitably be extended to other contexts.

This Article differs from these other writings in its focus on the significance of accrual accounting for a single, extremely important federal program. Unlike other work on fiscal imbalance, the emphasis here is arguing why a system of accrual accounting would provide a more accurate overview of the financial obligations of Social Security than do its current statements based largely on the presentation of annual cash flows. Also in contrast to other work on the subject, this Article analyzes the manner in which the current Social Security trustees' reports and most other official writing on the subject distort public understanding of the Social Security system's financial condition and needlessly complicate the debate over Social Security reform. While other studies offer a birds-eye view of the federal government's overall fiscal imbalance, this Article provides a ground-level assessment of how a new system of financial reporting could improve our understanding of one major social insurance program and the prospects for its sensible reform.

for Social Security might be integrated with other aspects of the federal budget to produce a more accurate and useful presentation of the government's overall financial position.

Part III explains how a modified system of accrual accounting could improve the ongoing debate over Social Security reform. Most obviously, a system of accrual accounting for Social Security—with substantial annual losses and growing unfunded obligations—would increase pressure on the political leadership to address the problems of Social Security in a timely manner. In addition, a system of accrual accounting would clarify the nature of the choices being presented to the American public and improve the incentives for political leaders to adopt responsible reform proposals that promote the long-run stability of Social Security.

Part IV concludes with a postscript on normative baselines for this proposal and argues that its approach is more consistent with these baselines than is the current system of cash-flow accounting. As explained below, the essential problem of accounting for Social Security is that the government first makes commitments to Social Security participants while they are working and contributing payroll taxes and then, many years later, must honor these commitments by making payments after the workers retire. In other governmental contexts where there is a substantial temporal disconnect between commitment and payment, government accounts are increasingly expressed on an accrual basis. What this Article proposes is simply an extension of this trend to our most important social-insurance program.[18]

Before turning to the substance of the argument, a few additional introductory points are in order. While issues of governmental accounting will undoubtedly strike many as mundane and technocratic, this Article presents a radical, in some circles even heretical, proposition for a social-insurance program such as Social Security. If adopted, this proposal could have a profound effect on the way the country thinks about Social Security and perhaps other social-insurance programs for the elderly, such as Medicare.

This Article does not, however, advance two more far-reaching propositions. First, it does not argue that the accrual-based statements should be the only financial statements prepared for the Social Security trust funds. Cash-flow statements and long-term projections of liquidity also provide valuable insight into the financial position of the Social Security system and should continue to be prepared. Accrual accounting should, however, be the dominant lens through which the Social Security system presents its financial posture to the world. The burden is on critics of this proposal to explain why accrual accounting should be the dominant format of financial presentation

---

[18] This Article is not, moreover, the first to explore this territory. Several years ago, the accounting oversight body responsible for developing generally accepted accounting practices for governmental entities adopted new rules that require social-insurance programs, such as Social Security, to disclose certain accrual-accounting elements similar to, though not nearly as extensive as, the ones recommended in this Article. *See infra* text accompanying notes 242–257. The logic underlying these reforms in government accounting supports the proposals offered here.

in so many other areas of financial and economic analysis, but not for the Social Security trust funds.

The second proposition from which this Article disassociates itself is the notion that the movement toward an accrual accounting system necessarily implies adoption of full, advance funding for Social Security: that each generation should pre-fund the full value of that generation's retirement benefits. It is perfectly possible to imagine a Social Security system with partial pre-funding and financial statements prepared on an accrual basis. In a growing economy with an expanding population, it may make perfect sense for future generations to pay for a portion of the cost of the retirement benefits of current generations, particularly if current generations have borne the cost of retirement for generations of past workers. It is, however, critical to keep control over the magnitude of liabilities that are being passed on to future generations. One of the virtues of accrual accounting is that it offers a perfect tool for monitoring the size of these liabilities and therefore provides a superior framework for ensuring the degree of pre-funding necessary to keep the growth of unfunded Social Security obligations in balance with other economic and social goals.

I. A Critique of the Current Accounting System for Social Security

Social Security historically has employed a cash-flow method of accounting. In part, this presentation reflects the traditional operating philosophy of the program. For many years, Congress and the Social Security Administration ran the program on a pay-as-you-go basis. The lion's share of annual revenues—raised primarily in the form of payroll taxes paid by current employees—were immediately transferred to current benefi-ciaries.[19] As long as inflow was adequate to meet outflow, the system was considered to be in balance. Indeed, for many decades, if the inflow exceeded the promised benefits, Congress interpreted the cash-flow surpluses as a sign that benefits should be raised so as to make use of the system's excess cash reserves.[20]

While generally consistent with traditional federal government accounting practices,[21] cash-flow accounting leaves much to be desired in the context of public retirement programs such as Social Security. This Part briefly summarizes current accounting practices for the Social Security program and then explains how these practices distort public understanding of

---

[19] *See* Eric M. Patashnik, Putting Trust in the US Budget 72 (2000) (describing benefits increases and payroll tax reduction designed to reduce size of projected reserves).

[20] *See* Light, *supra* note 3, at 33–57; *see also* Merton C. Bernstein & Joan Brodshaug Bernstein, Social Security: The System that Works 34–35 (1988).

[21] As described below, the federal budget is increasingly making use of accrual accounting for certain programs. *See infra* Part IV.A. Indeed, the traditional presentation of Social Security finances in the trustees' annual reports no longer complies with generally accepted accounting standards for government entities. *See infra* Part IV.A.2.

the program. It begins with the Social Security trust funds themselves and then turns to the relationship between the trust funds and the unified federal budget.

A. Analysis of Accounting for the Trust Funds

Since the Social Security program's early days in the 1930s, Social Security revenues have been placed in "trust funds."[22] These trust funds have little in common with traditional trusts. Assets are not legally segregated for particular classes of beneficiaries, nor do Social Security trustees have any legal obligation to protect the interests of program participants.[23] Rather, the trust funds are accounting entries in the federal budget to which Social Security contributions and certain other accruals and payments are attributed and from which Social Security benefits and certain other expenses are paid.[24] The trust funds are, however, the principal vehicles through which politicians and the public analyze the program's financial status and therefore provide a useful starting point for analysis of the program's current accounting treatment.

1. Standard Presentation

Each year, the trustees of the Social Security system produce an annual report of the trust funds' financial condition. A cursory review of the trustees' most recent report—for the calendar year ending December 31, 2002—reveals two basic financial messages. In the short term, Social Security is doing quite well, but over the longer run, this program is on a course for financial ruin.[25]

---

[22] The trust funds were not included in the original Social Security legislation of 1935 but were added in 1939 amendments. *See* PATASHNIK, *supra* note 19, at 67–69 (explaining that concerns over constitutional challenges prevented initial legislation from including trusts).

[23] *See* 42 U.S.C. § 401(c) (2000) ("A person serving on the Board of Trustees shall not be considered to be a fiduciary and shall not be personally liable for actions taken in such capacity with respect to the Trust Funds.").

[24] 42 U.S.C. § 401 (2000).

[25] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 2. *See also* 2002 TRS. OF THE FED. OLD-AGE & SURVIVORS INS. & DISABILITY INS. TRUST FUNDS ANN. REP. 3 (2002) [hereinafter 2002 TRUSTEES REPORT].

TABLE ONE

Summary of Trust Fund Operations in 2002 (billions)
(Source: Table II.B1 of 2003 Trustees Report)

Trust Fund Assets (year end 2001)		**$ 1,212.5**
Income During 2002:		
Payroll Taxes:	$532.5	
Taxation on Benefits:	13.8	
Transfers	0.4	
Interest on Assets:	80.4	
Total Income	*$ 627.1*	
Expenditures During 2002:		
Benefit Payments:	$ 453.8	
Railroad Retirement	3.6	
Administrative Expenses	4.2	
Total Expenditures	*$ 461.7*	
Net Increase in Assets:	**$ 165.4**	
Trust Fund Assets (year end 2002)		**$ 1,378.0**

a. Short-Range Projections

According to the 2003 Trustees Report, the short-range prospects of Social Security are good because the combined Old Age, Survivors, and Disability Insurance (OASDI)[26] trust funds have more than a trillion dollars of reserves deposited with the U.S. Treasury and are projected to take in more in income than they put out in expenditures for at least the next decade. Table One provides the kind of typically reassuring presentation that one finds in the opening pages of the most recent Trustees Report.[27] It shows a rosy picture of current operations. The combined OASDI trust funds started the year with some $1,212.5 billion in assets deposited with the U.S. Treasury. Over the course of the year, the funds took in $532.5 billion in payroll taxes,[28] another $13.8 billion in taxes on Social Security benefits,[29] plus $80.4 billion

[26] Throughout, this Article refers to the combined Old Age, Survivors, and Disability Trust Funds. In fact, there are two separate trust funds—one for Old Age and Survivors Insurance and the other for Disability Insurance—and their individual financial status differs somewhat from that of the combined fund. *See* 42 U.S.C. § 401(a)–(b) (2000). For the purposes of this Article, however, these differences are not significant.

[27] 2003 TRUSTEES REPORT, *supra* note 8, at 4 tbl.IIB1.

[28] These revenues resulted from a levy of 12.40% imposed on taxable wages up to $87,000 in 2003 and divided evenly into employee and employer shares of 6.20% each. The level of covered payroll is adjusted annually to reflect cost-of-living increases. *See* Soc. Security Admin., *2002 Social Security Changes*, *at* http://www.ssa.gov/cola/colafacts2003.htm (last visited Nov. 17, 2003).

[29] This source of revenue comes from funds raised through the federal income tax system. A portion of Social Security benefits is treated as taxable income under the federal income tax,

from interest on trust fund assets deposited with the Treasury,[30] for a combined income of $627.1 billion for the year. Charged against this inflow, as seen in Table One, were some $461.7 billion in expenditures, consisting almost entirely of benefit payments. Overall, there was a "net increase in assets" of $165.4 billion during the year, which increased the funds' total assets to $1,378 billion at year end.[31]

Another way that the trustees commonly describe the favorable short-term outlook for the Social Security system is by reference to a "trust fund ratio."[32] This statistic is the ratio of the trust funds' total assets at the beginning of the year to the projected total expenditures over the course of the year.[33] In assessing the trust fund ratio, the trustees have set 100% as a minimum acceptable ratio on the theory that, as long as the trust funds had at least a year's worth of expenditures on hand, Congress would have time to respond to any unexpected short-term crisis in the system's financing; the trust funds could receive no additional revenues for a year and still be able to pay benefits at current levels until the end of the year.[34]

Measured in terms of trust fund ratios, the 2003 Trustees Report also shows a carefree short-term picture, as presented in Figure One. Based on the trustees' intermediate forecasts,[35] this chart shows the trust fund ratio increasing from 263% at the end of 2002 to a high of 452% a decade later—that is, cash reserves equal to nearly four and a half years of projected expenditures in 2012:[36] In the trustees' words, "[B]ecause the trust fund ratio for the combined funds is estimated to remain above 100% under the intermediate assumptions, the combined funds meet the short-range test of financial adequacy."[37] Even under the pessimistic, high-cost estimates, the trust funds are projected to remain above the 100% threshold throughout the coming decade.[38]

---

and a portion of the taxes levied in this manner is redirected to the trust funds. *See* I.R.C. § 86 (2000); 2003 TRUSTEES REPORT, *supra* note 8, at 18–19.

[30] *See* 42 U.S.C. § 401 (2000). The effective interest rate on OASI trust fund assets in 2002 was 6.4% while the effective interest rate on DI trust fund assets was 6.3%. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 21, 25.

[31] The trust fund expenditures include administrative costs ($4.2 billion) and transfers to the Railroad Retirement program ($3.6 billion).

[32] *See* 2000 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 15 (2000) [hereinafter 2000 TRUSTEES REPORT].

[33] *Id.* In 2002, for example, the trust fund ratio for the combined funds was 263%: total assets at the beginning of the year of $1,212.5 billion divided by total expenditures during the year of $461.7 billion. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 41.

[34] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 32.

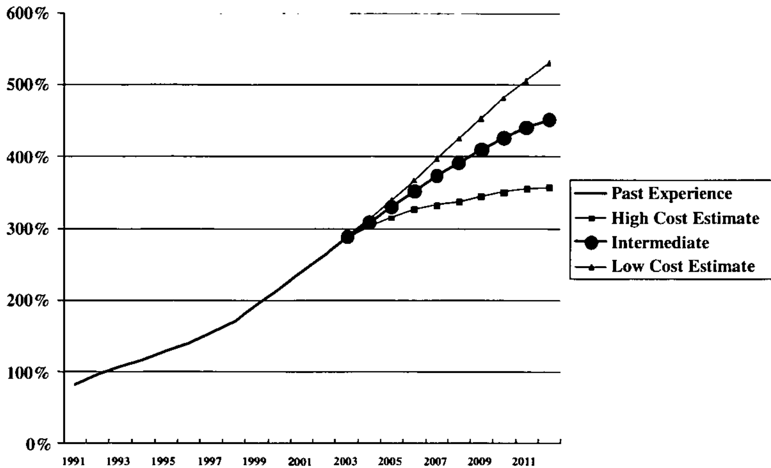
[35] The trustees' reports typically present three alternative cost estimates: high, low, and intermediate. Unless otherwise indicated, this Article will use the intermediate estimates, as these are the ones most commonly used in public discussions of Social Security. To the extent feasible, figures and tables will report all three estimates. For a discussion of the assumptions and methods underlying the actuarial estimates, see *id.* at 74–128.

[36] *See id.* at 41.

[37] 2000 TRUSTEES REPORT, *supra* note 32, at 15.

[38] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 7.

FIGURE ONE



b. Long-Range Projections

When discussing the trust funds' long-range prospects, the trustees' reports become decidedly more pessimistic. Notwithstanding a much-publicized bipartisan effort in the early 1980s to produce a permanent solution to the problem of Social Security,[39] the trustees have for some time now been warning of serious dangers confronting the system in the next decade.[40] The 2003 Trustees Report sounded the same alert:

> Under current law the cost of Social Security will increase faster than the program's income, because of the gain of the baby-boom generation, expected continuing low fertility, and increasing life expectancy. Based on the Trustees' best estimates, expenditures will exceed tax revenues starting in 2018 and throughout the remainder of the 75-year projection period. Social Security's combined trust

---

[39] *See* LIGHT, *supra* note 3, at 115–228.

[40] *See* 2000 TRUSTEES REPORT, *supra* note 32, at 3 ("On a combined basis, the OASDI program is not in 'close actuarial balance' over the next seventy-five years. In addition, the individual OASI and DI Trust Funds are not in close actuarial balance. These conclusions are the same as those shown in the 1999 Annual Report."); *see also* 2003 TRUSTEES REPORT, *supra* note 8, at 7–10.

> funds are projected to allow full payment of benefits until they become exhausted in 2042.[41]

Based on these projected long-range shortfalls, the trustees regularly urge Congress and the President to revisit the issue of Social Security and make appropriate adjustments to forestall the looming crisis.[42]

The 2003 Trustees Report includes several different accounting presentations to support the report's claims of long-range insolvency. One presentation, summarized in Figure Two, compares long-range estimated income rates to cost rates over the next seventy-five years, a period chosen because it includes the life expectancy of almost all current workers and retirees.[43] The rates represent the system's projected annual cash inflows from taxes and outflows for expenses and are expressed as a percentage of taxable wages. The income rates rise fairly modestly over the period, as payroll taxes are not scheduled to increase beyond their current 12.4% of taxable payroll, and income generated by the taxation of Social Security benefits has only a modest effect on overall inflow.[44] The system's cost rates, in contrast, rise dramatically in the second and third decades of the century, surpassing income rates (under intermediate assumptions) by 2018 and generating an annual combined trust fund deficit of 6.71% of payroll by 2077.

---

[41] 2003 TRUSTEES REPORT, *supra* note 8, at 16. *See also* 2002 TRUSTEES REPORT, *supra* note 25, at 18.

[42] *See* 2000 TRUSTEES REPORT, *supra* note 32, at 30–31 ("In view of the size of the financial shortfall in the OASDI program over the next 75 years, we again urge that the long-range deficits of both the OASI and DI Trust Funds be addressed in a timely way."); *see also* 2003 TRUSTEES REPORT, *supra* note 8, at 17; 2002 TRUSTEES REPORT, *supra* note 25, at 18; 2001 TRS. OF THE FED. OLD-AGE & SURVIVORS INS. & DISABILITY INS. TRUST FUNDS ANN. REP. (2001) [hereinafter 2001 TRUSTEES REPORT].

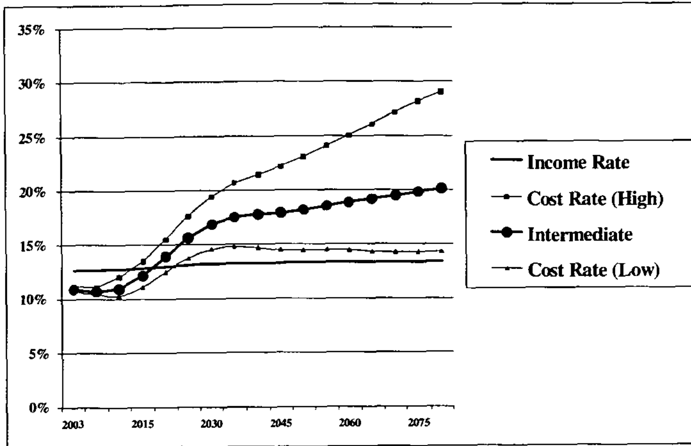
[43] *See* Stephen C. Goss, *Measuring Solvency in the Social Security System*, *in* PROSPECTS FOR SOCIAL SECURITY REFORM 16, 19 (Olivia S. Mitchell et al. eds., 1999).

[44] For example, under the trustees' intermediate projections, the combined income rate for OASDI funds increases by only 0.73 percentage points during the seventy-five-year period, from 12.70% of taxable wages in 2003 to 13.43% in 2080. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 47.

FIGURE TWO

Long-Range Estimated Income and Cost Rates

(from Table IV.B1 of 2003 Trustees Report)



The substantial imbalance between projected income and benefit rates lies at the heart of the long-range crisis facing the Social Security system. It is, however, a somewhat complicated phenomenon to comprehend because it combines the short-range surplus that will accumulate between now and some time in the next decade with longer-term deficits. Admittedly, a quick review of charts such as Figure Two suggests that the long-term deficit is, crudely speaking, bigger than the short-range surplus. Such eyeball assessments are unreliable, however, particularly when dealing with a period extending seventy-five years into the future. Moreover, a simple comparison of income and benefit rates does not take into account the current reserves of the combined trust funds (that is, the more than $1,378 billion in assets deposited with the Treasury as of December 31, 2002) or the interest that can be expected to accrue on these reserves.[45] Recognizing that simply comparing cost and income rates yields an incomplete picture, the trustees' reports offer two different ways of conceptualizing the long-range imbalance of the combined OASDI trust funds.

The first, which is illustrated in Tables Two and Three, presents what are known as summarized income and cost rates. In essence, these summarized figures attempt to normalize projected costs and benefits over three time

[45] In other words, the income rates reported in presentations such as Figure Two do not include income in the form of interest payments from the trust fund reserves. This interest does, however, figure into the annual report of trust fund activities summarized Table One.

periods (the following twenty-five, fifty, and seventy-five years), generating what are, in effect, average income and cost rates for the three periods.[46] The summarized presentations offer a more complete picture of income and expenses than do the simple comparisons of annual income and benefit rates discussed above. As can be seen from Table Two, these presentations include information not just about the payroll tax and the taxation of benefits but also about the annualized value of the current trust fund reserves, spread over the relevant period, plus the cost of building up a final trust fund balance equal to 100% of benefits projected for the year following the end of the period.

TABLE TWO

Components of Summarized Income and Cost Rates
(from Table IV.B11 of 2003 Trustees Report)

Valuation Period	Income Rate				Cost Rate		
	Payroll Tax	Taxation of Benefits	Beginning Fund Balance	Total	Disburse-ments	Ending Fund Balance	Total
Intermediate:							
25-years:							
2003–2027	12.39	0.47	1.48	14.34	12.53	0.56	13.09
50-years:							
2003–2052	12.39	0.62	0.88	13.89	14.59	0.25	14.84
75-years:							
2003–2077	12.39	0.70	0.69	13.78	15.56	0.14	15.70
Low Cost:							
25-years:							
2003–2027	12.39	0.43	1.47	14.28	11.50	0.48	11.98
50-years:							
2003–2052	12.39	0.54	0.87	13.80	12.77	0.20	12.97
75-years:							
2003–2077	12.39	0.58	0.67	13.65	13.12	0.11	13.23
High Cost:							
25-years:							
2003–2027	12.39	0.53	1.48	14.40	13.80	0.65	14.45
50-years:							
2003–2052	12.39	0.74	0.87	14.00	16.87	0.33	17.20
75-years:							
2003–2077	12.39	0.86	0.68	13.93	18.80	0.20	19.00

The value of summarized revenue and cost rates is that they allow analysts to compare the relative magnitudes of various components of the trust funds' long-term balance. These tables show, among other things, the relatively minor long-run significance of the trust funds' current reserves of $1,378 billion.[47] An analysis of the summarized components also confirms

[46] The summarized data reflect not an arithmetic mean, but an average of the discounted value of various components over the relevant period. Thus, the summarized rates take into account the time-value of money as well as the projected accrued interest on trust fund reserves. *See* Goss, *supra* note 43, at 20.

[47] For example, if one were to assume (under the intermediate-cost estimates) that the entire current reserves ($1,378 billion) would be amortized over the next twenty-five years, the reserves would be equivalent to an increase of 1.48% in payroll taxes, and, if adjustment were

the accuracy of an eyeball assessment of Figure Two—the long-run deficits of the out years substantially outweigh the funds' near-term surpluses in all but the trustees' low-cost estimates.

TABLE THREE

Summarized and Combined Income and Cost Rates

(from Table IV.B5 of 2003 Trustees Report)

Valuation Period	Income Rate	Cost Rate	Combined Rate
Intermediate:			
25-years:			
2003–2027	14.34	13.09	1.26
50-years:			
2003–2053	13.89	14.84	–0.95
75-years:			
2003–2077	13.78	15.70	–1.92 ⇦
Low Cost:			
25-years:			
2003–2027	14.28	11.98	2.31
50-years:			
2003–2052	13.80	12.97	0.83
75-years:			
2003–2077	13.65	13.23	0.42
High Cost:			
25-years:			
2003–2027	14.40	14.45	–0.05
50-years:			
2003–2052	14.40	17.20	–3.20
75-years:			
2003–2077	13.93	19.00	–5.07

The statistic that encapsulates all of this information can be found in Table Three, showing summarized and combined income and cost rates. The table restates the total income and cost rates shown in Table Two and then indicates the difference between those two rates for each period of analysis. The figure that the arrow points to—minus 1.92—is the one that policy analysts often employ as a shorthand measure of Social Security's long-term problems. This measure, sometimes referred to as the "actuarial deficit," reflects the difference between seventy-five-year revenue and cost rates under the trustees' intermediate assumptions. The actuarial deficit is often described as the amount that the payroll tax rate would have to be increased (starting today and continuing for the full seventy-five-year period of analysis) to bring the Social Security trust funds into long-range balance.[48] Typically, reform measures—ranging from reductions in benefits to increases in

---

made for the 0.56% cost rate associated with building an adequately funded reserve at the end of the period, the increase would be only 0.92%. Over longer periods, the current trust reserves have even less significance.

[48] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 16 ("This deficit indicates that financial adequacy of the program for the next 75 years could be restored if the Social Security payroll tax were immediately and permanently increased from its current level of 12.4 percent (for employees and employers combined) to 14.32 percent.").

the taxation of Social Security benefits to gains from improving the yield on trust fund assets—are measured in terms of their effect in reducing or eliminating the current actuarial deficit.[49]

According to standards of long-range "close" actuarial balance that the trustees have established, a 1.92% actuarial deficit in the combined rate for the seventy-five-year horizon is unacceptable. According to the trustees' standards, the combined ratio for this period should be no greater than 5% of the summarized cost rate for the same period, that is, 5% of 15.70%, or 0.79%.[50] The intuition underlying this standard is that over the long run the trust funds should be considered in actuarial balance only if projected revenues meet 95% of projected costs.

Another way of presenting the trust funds' long-range solvency is to use charts showing the projected trust fund balances in future years.[51] In essence, these charts reflect a continuation of those in Figure One, which was limited to ten years of trust fund ratios. Figure Three is an example of this sort of long-range presentation. This figure reveals that (under intermediate estimates) the trust fund ratios will begin to decline in approximately 2016, will pass the trustees' minimum prudent reserve level (100% of projected annual benefits) in 2039, and will be entirely depleted of resources in 2042.[52] At that point, if this scenario were to come to pass and no intervening legislation had been enacted, the combined trust funds would have insufficient resources to honor the projected cost of all promised benefits. In fact, the funds' projected income for that year (roughly 13.26% of taxable payroll) could cover less than three-quarters of projected benefits (roughly 17.80% of taxable payroll).[53] By the end of the seventy-five years, the trust funds' projected coverage would be about two-thirds of projected benefits (that is, 13.42% of taxable payroll in income versus 19.92% of taxable payroll in benefits).[54]

---

[49] For example, President Bush's Commission on Social Security used this measure as a benchmark for evaluating various reform proposals. *See* BUSH COMMISSION REPORT, *supra* note 2, at 69–70. *See also* 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 11 (using the same measure of insolvency to assess proposed reforms).

[50] See 2003 TRUSTEES REPORT, *supra* note 8, at 63–66. In their assessment of combined rates, the trustees actually use a sliding scale, under which combined rates must precisely match cost rates over a ten-year period (that is, over the short range) and are allowed to deviate gradually by up to 5.0% over the full seventy-five-year horizon. *Id.*

[51] *See id.* at 15.

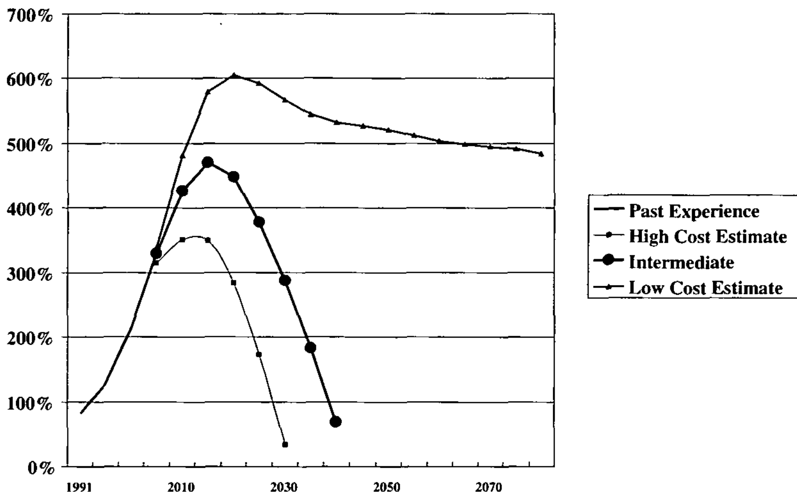
[52] *See id.* at 16. Annual estimates of projected trust fund ratios are available from the SSA Office of the Chief Actuary. SSA Office of the Chief Actuary, *Estimated Trust Fund Ratios: Calendar Years 2003–80, at* http://www.ssa.gov/OACT/TR/TR03/lr4B3.html (last visited Oct. 20, 2003).

[53] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 47 tbl.IV.B1 (projected cost and income rates for 2040 and 2075). These ratios reflect the differences in projected income and benefit rates shown in Figure Two (intermediate estimate). Annual estimated projections of cost and income rates are available from the Office of the Chief Actuary. SSA Office of the Chief Actuary, *Estimated Annual Income Rates, Cost Rates, and Balances: Calendar Years 2003–80, Intermediate Assumptions, at* http://www.ssa.gov/OACT/TR/TR03/lr4B1-2.html (last visited Oct. 20, 2003).

[54] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 47 tbl.IV.B1.

FIGURE THREE

Long-Range Trust Fund Ratios
(from Table IV.B3 of 2003 Trustees Report)



2. *Criticisms of the Standard Presentation*

Although the information embodied in the trustees' reports is well presented and of considerable value, there are serious limitations in their standard presentations. Put simply, the optimistic short-range projections, and the relatively pessimistic long-range projections, are both too optimistic. Moreover, in many respects, these presentations distort public debate over competing reform proposals.

a. Short-Range Projections

As explained above, the take-home message of the trustees' short-range projections is that the next ten years appear favorable for the Social Security trust funds. To be sure, there is some truth to this message, at least in the sense that the trust fund reserves on deposit in the Treasury will grow during the coming decade. But this assessment is based exclusively on an analysis of cash flow. It takes no account of changes in the system's liabilities over the course of each year—either liabilities that were liquidated (benefit promises fulfilled) or the new liabilities generated (new benefits promised).

Indeed, the trustees' reports have a somewhat curious attitude with respect to the system's liabilities. Reading through reports, one is constantly warned that the primary problem facing the Social Security system is the retirement of the baby-boomer generation starting a decade or two in the future.[55] At this point, the reports advise, the system will owe this generation a substantial level of benefits, which it will be unable to pay even after drawing down substantial reserves that will have accumulated over the intervening years.[56] Thus, the short-range surplus will be overwhelmed by liabilities arising fully formed several decades in the future.

The dichotomy between short-range feast and long-range famine is inaccurate and misleading. Social Security faces a long-term crisis because it has already made substantial pension promises to current workers and retirees, and each year promises to make additional pension payments to most current workers. When the 2003 Trustees Report suggests that 2002 was a good year for the Social Security system and emphasizes a $165.4 billion increase in the system's assets in that year,[57] the data presented give no sense of how the growing cash surpluses compare to other aspects of the system's operations during the year. To be sure, an astute reader might infer that all was not well with the system from the report's ample evidence that current surpluses will be insufficient to meet long-run benefit commitments. But it is all but impossible to determine from the 2003 Trustees Report whether these long-run problems are the consequence of promises made before 2002 (that is, promises made to current retirees and older workers) or promises to be made to baby boomers and others at some point in the future. In short, the 2003 Trustees Report gives little sense of whether, all things considered, 2002 was a good year for the Social Security trust funds.

In referring to liabilities of the Social Security system, it is important to recognize that the character of Social Security's pension promises is complex and contestable. As a purely legal matter, Social Security benefits do not constitute binding obligations on the part of the federal government.[58] Because Social Security beneficiaries have no constitutionally protected right to receive any particular level of Social Security benefit, it remains problematic to define how individual benefits accrue over time. While some might argue that the absence of enforceable legal rights on the part of Social Security beneficiaries wholly justifies the formats adopted in the trustees' reports,[59] there are several strong grounds for rejecting this approach.

---

[55] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 2 ("The OASI and DI Trust Funds, individually and combined, are adequately financed over the next 10 years under the intermediate assumptions.").

[56] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 3 ("The combined OASDI Trust Funds are projected to become insolvent in 2042 under the long-range intermediate assumptions.").

[57] *See* Table One.

[58] *See* Box Two.

[59] For example, in a recent exposure draft regarding the accounting treatment of Fiduciary Activities, the Federal Accounting Standards Advisory Board (FASAB) proposed applying the new standards only to fiduciary activities in which non-federal parties have an "ownership interest

BOX TWO

THE LEGAL STATUS OF SOCIAL SECURITY BENEFITS

The Supreme Court has addressed the question of whether Congress has the authority to reduce or eliminate Social Security benefits for individuals who have paid Social Security taxes under a statutory regimen that provides for a certain level of benefits. In *Fleming v. Nestor*, 363 U.S. 603, (1960), the Court considered whether Social Security participants have a property interest in Social Security benefits that, under the Fifth Amendment of the U.S. Constitution, cannot be diminished without just compensation. The Court ruled that "a person covered by the [Social Security] Act has no[] . . . right in benefit payments as would make every [statutory] defeasance of 'accrued' interests violative of the Due Process Clause." *Id.* at 611.

The *Fleming* case involved a Bulgarian immigrant named Ephram Nestor, who had arrived in the United States in 1913 and paid Social Security taxes from 1936 until his retirement in 1955. *Id.* at 605. In 1954, Congress had amended the Social Security Act to deny retirement benefits to any beneficiary deported from the United States on the grounds of being or having been a member of the Communist party, and Nestor was deported on that basis in 1956. *Id.* Nestor, who would otherwise have been eligible for benefits given his years of payment of Social Security taxes, had his benefits revoked, and he challenged the revocation as violating the Fifth Amendment. *Id.* at 606. Writing for a majority, Justice Harlan rejected Nestor's claims, emphasizing that giving Congress flexibility to recalibrate Social Security benefits from time to time is a practical necessity and that the Social Security Act expressly reserves for Congress the right to "alter, amend or repeal any provision." *Id.* at 611 (quoting 42 U.S.C. § 1304). Though occasionally criticized and distinguished in subsequent decisions, the core holding of *Fleming* retains vitality: Congress can reduce or eliminate accrued Social Security benefits without having to provide just compensation to participants or other beneficiaries whose entitlements are thereby diminished. Section 1304 of the Social Security Act still explicitly preserves this authority by providing that "the right to alter, amend or repeal any provision of this Act is hereby reserved to the Congress." 42 U.S.C. § 1304 (West. 2000).

The only significant constitutional protection afforded Social Security participants is that individualized decisions to reduce or eliminate benefits must be conducted in accordance with statutory standards and effected through procedures meeting the requirements of the Due Process Clause of the Fifth Amendment. For a review of the doctrine in this area, see JULIE A. NICE & LOUISE G. TRUBEK, CASES AND MATERIALS ON POVERTY LAW: THEORY AND PRACTICE 284–374 (1997).

---

that the Federal Government must uphold," thus exempting social insurance programs such as Social Security. *See* ACCOUNTING FOR FIDUCIARY ACTIVITIES, Statement of Federal Financial Accounting Standards Exposure Draft, (Fed. Accounting Standards Advisory Bd. 2003) [hereinafter ACCOUNTING FOR FIDUCIARY ACTIVITIES], *available at* http://www.fasab.gov/pdf/trustf1.pdf (last visited Oct.

For one, an accounting treatment that reflects the accrual of benefits is more consistent with the statutory framework for the payment of Social Security benefits. As defined under the Social Security Act, an individual's Social Security benefits are based on the number of years the individual participates in the system. Individuals become eligible to receive retirement benefits once they have made contributions for ten years, and their statutory entitlements grow larger as they make contributions to the system.[60] The amount of benefits ultimately paid out of the trust funds is a function of a participant's lifetime earnings, based on the highest-wage years over the course of his or her working career.[61] When economists want to calculate the value of Social Security benefits, they routinely look to these statutory formulas to estimate how large a benefit a particular individual has accrued at a particular time or at the individual's retirement.[62] The trust funds' financial statements, however, make no effort to quantify this gradual accretion of statutory entitlements for individual beneficiaries.

The trustees' reports also fail to reflect the political reality of the trust funds' benefit structure. According to Social Security lore, President Roosevelt wanted Americans to think of themselves as having contributed to individual Social Security accounts—that is, accounts held in trust and funded through individual contributions—so as to create a sense of personal connection to the Social Security program and thereby increase public support for the system.[63] If this was, in fact, President Roosevelt's intention, history has seen his vision realized. Public opinion surveys confirm that older workers and retirees do feel that the federal government has an obligation (that is, has incurred a liability) to pay Social Security benefits as promised.[64] Although it is difficult to say exactly when that obligation arises (that is, accrues), there is a national consensus that somewhere after entry into the workforce and before the age of retirement, workers and their families do accrue—politically, if not constitutionally—an entitlement to receive Social Security benefits at approximately the levels indicated by the statutory structure.[65]

17, 2003). For further discussion of FASAB pronouncements in this area, see *infra* Part IV.A.2.

[60] The rules governing Social Security benefits are extraordinarily complex. The key statutory entitlement for participants is the primary insurance amount, defined in 42 U.S.C. § 415 (2002). For a good non-technical explanation of Social Security benefits, see C. EUGENE STEUERLE & JON M. BAKIJA, RETOOLING SOCIAL SECURITY FOR THE 21ST CENTURY 75–79 (1994).

[61] *See* STEUERLE & BAKIJA, *supra* note 60, at 75–79.

[62] *See infra* text accompanying notes 275–283.

[63] *See* STEUERLE & BAKIJA, *supra* note 60, at 25–27.

[64] *See* Lawrence R. Jacobs & Robert Y. Shapiro, *Myths and Misunderstandings about Public Opinion Towards Social Security in Framing the Social Security Debate*, *in* FRAMING THE SOCIAL SECURITY DEBATE: VALUES, POLITICS, AND ECONOMICS 355, 365–74 (R. Douglas Arnold et al. eds., 1998).

[65] *See* Jacobs & Shapiro, *supra* note 64, at 359 n.14 ("70 percent to 80 percent of Americans agreed that 'everyone who pays into Social Security should receive it, no matter what other income they have.'").

The trustees' reports are also inconsistent with methods used to account for analogous private pension plans. Under Generally Accepted Accounting Principles (GAAP), private firms are not permitted to ignore retirement benefits promised to current workers.[66] Whether offering pension or retiree health benefits, companies must accrue the costs of retiree benefits while employees are still in the workforce. The required rates of accrual are not based simply on the legal entitlements of workers, but rather on the projected level of benefits that workers are actually expected to receive—levels of projected benefits that are often much higher than workers' legal entitlements.[67] The premise of this GAAP requirement is that retiree benefits should be recognized on corporate financial statements as soon as it is reasonably clear that the benefits will have to be paid.[68] The financial statements of Social Security disregard this central GAAP principle.[69]

Finally, the inclusion of presently incurred promises is consistent with accounting's own definition of liabilities. Liabilities are defined as "probable future sacrifices of economic benefits arising from present obligations of a particular entity to transfer assets or provide services to other entities in the future as a result of past transactions."[70] This definition is intentionally expansive. The commitments need not be definite, so long as obligations are probable—"that which can reasonably be expected or believed on the basis of available evidence or logic but is neither certain nor proved."[71] Moreover, such commitments need not be legally enforceable obligations; rather, the term "obligations" as used in the definition of liabilities is broader, and includes duties that "one is bound to do by contract, promise, moral responsibility, and so forth."[72] The trustees' reports fail to recognize in a timely manner the substantial financial commitments that the trust funds will in all likelihood need to honor as a result of past contributions to the system.

This notion of timeliness is of critical importance to the construction of financial accounts. Liabilities are recognized when they become probable

---

[66] *See* EMPLOYERS' ACCOUNTING FOR PENSIONS, Statement of Financial Accounting Standards No. 87, 15–19 (Fed. Accounting Standards Advisory Bd. Dec. 1995) [hereinafter EMPLOYERS' ACCOUNTING FOR PENSIONS], *available at* http://www.fasb.org/pdf/fas87.pdf (last visited Nov. 17, 2003).

[67] JOHN H. LANGBEIN & BRUCE A. WOLK, PENSION AND EMPLOYEE BENEFIT LAW 372 (3d ed. 2000).

[68] *See* EMPLOYERS' ACCOUNTING FOR PENSIONS, supra note 66 at 15–16.

[69] As discussed *infra* text accompanying notes 231–240, federal government budgeting procedures increasingly are making use of accrual techniques to account for government activities that entail long-term commitments. Recognizing the value of accrual accounting, the Bush Administration recently proposed that all retirement benefits for government employees also be accounted for in this way, at least for internal management purposes. *See* CONGRESSIONAL BUDGET OFFICE, THE PRESIDENT'S PROPOSAL TO ACCRUE RETIREMENT COSTS FOR FEDERAL EMPLOYEES (June 2002), *available at* http://www.cbo.gov/showdoc.cfm?index=3580&sequence=0 (last visited Oct. 17, 2003) [hereinafter PRESIDENT'S PROPOSAL].

[70] *See* ACCOUNTING STANDARDS, Statements of Financial Accounting Concepts 191 (Financial Accounting Standards Bd. 1999).

[71] *Id.* at n.21.

[72] *Id.* at n.22.

economic commitments so that managers and other readers are made aware of the creation of financial commitments at the time those obligations are incurred and can be controlled.[73] As a matter of political reality and individual equity, the Social Security system's obligations to pay retirement benefits are incurred during the working life of individual participants, which is when accrual accounting would recognize those obligations. The trustees' reports, however, recognize those commitments much further in the future, after the participants have retired and begun to receive payments. At that point—regardless of the constitutional status of Social Security benefits—the federal government has long since assumed the practical responsibility to honor these commitments.

b. Long-Range Projections

Although the trustees' long-range projections quite properly warn of a looming crisis, these projections are also misleading and incomplete. Recall for a moment the basic message of this aspect of recent trustees' reports: over the next seventy-five years, the projected revenues of the trust funds will be insufficient to cover projected costs. The summary statistic provided is the 1.92% actuarial deficit (marked with the arrow in Table Three) for the trust funds' combined rate. This statistic is usually interpreted as the amount by which payroll taxes would have to be raised throughout the seventy-five-year period to bring the system into perfect balance. Converting these taxes to present value implies that if the Social Security trust funds had received a lump-sum contribution of roughly $3.8 trillion on December 31, 2002, the program's long-range problems would be solved.[74]

Assume for a moment that the federal government were to raise this staggering sum through a special assessment of some sort. Would the long-range problems of Social Security be solved once and for all? The trustees' long-range projections suggest that they would, but this is false. The trust funds would begin to fall out of actuarial balance again the very next year. Consider the long-range estimated income and cost rates shown in Figure 2. In 2078, at the end of the long-range projection period, the income and liabilities of the Social Security system will be substantially out of balance. Even

---

[73] *See* GEN. ACCOUNTING OFFICE, BUDGETING FOR FEDERAL INSURANCE PROGRAMS 8 (Apr. 23, 1998) [hereinafter BUDGETING FOR FEDERAL INSURANCE PROGRAMS] (citing among the benefits of accrual-based accounting, the provision of "an opportunity to control costs before the government is committed to making payments").

[74] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 61 (explaining components of this estimate). This figure is slightly higher than the estimate of the open-group unfunded obligation discussed below ($3.5 trillion), *see infra* text accompanying notes 172–179, because this estimate factors in the additional present value cost of building up a reserve at the end of the seventy-five-year projection period ($280 billion). To put this number in context, the total revenues of the United States government in fiscal year 2002 were just $1.85 trillion. *See* CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: FISCAL YEARS 2004-2013, at 4 tbl.1-2 (Jan. 2003) [hereinafter CBO JANUARY 2003 REPORT].

if the long-range Social Security problem defined in the 2003 Trustees Report were "solved," a new long-range problem requiring a similar solution would emerge in a decade or two.[75]

The trustees' long-range projections suffer from the same defect as their short-range projections in failing to account for the accrual of future obligations. As described earlier, the long-range projections are designed to take into account all benefits paid and revenues received from current and future participants over the projection period.[76] Thus, the long-range projections encompass the trust funds' obligations to these individuals inasmuch as all of the obligations are liquidated over the seventy-five-year period. What the long-run projections fail to consider are the trust funds' accrued liabilities to generations of workers who will receive benefits after this period. After seventy-five years, at the end of 2077, millions of workers and retirees will have substantial claims on the system, and the trust funds will have neither reserves to honor these claims nor revenue streams to support them. The actuarial deficit measure, large though it is, underestimates the true long-range shortfall of the system, because it fails to account for the future accrual of Social Security promises to individuals who have not yet entered the workforce but will have substantial claims on the system at the end of seventy-five years.[77]

c. *Biases with Respect to Reform Proposals*

The way in which the trustees' reports present the financial problems of Social Security also significantly distorts the public debate over reform proposals. To begin with, the take-home message of these reports—short-range surplus and long-range deficits—invites politicians to postpone the difficult and politically charged problems of Social Security reform. The trustees' reports' repeated dichotomy between short- and long-term prospects suggests that the system is currently making some sort of profit—which it is not—and can safely be left alone until the real problems appear somewhere down the road. Experts in the field know this is not the case, and regularly warn that the system's problems will be much harder to solve if reform ef-

---

[75] *See* BUSH COMMISSION REPORT, *supra* note 2, at 69–70. This defect in traditional social security accounting is sometimes referred to as the "cliff" problem. *Id.* at 70. A good example of this phenomenon is the growth of the Social Security trust fund deficits in the years immediately following the 1983 reforms. At the beginning of this period, when the reforms had just been enacted, the system was in long-range balance, although it quickly fell into long-range imbalance again. *See infra* Figure 5.

[76] *See supra* text accompanying notes 40–54.

[77] Admittedly, the long-range measures of actuarial balance do contemplate the accumulation of trust fund reserves at 100% of projected annual expenditures at the end of the period. This level of reserves, however, does not ensure long-run solvency. After all, the trust funds currently have reserves in excess of 200% of annual expenditures, *see supra* text accompanying notes 26–31, and the system is widely perceived to be on the edge of crisis.

forts are delayed.[78] But the trustees' reports undercut this warning by emphasizing the magnitude of the system's short-term cash surpluses.

Consider, for example, the press coverage that followed the release of the 2003 Trustees Report. On the day after the report's release, for example, the *Wall Street Journal* reported:

> Social Security's long-term prognosis brightened a bit. In annual reports released yesterday . . . Social Security gained a year on its expected life span, to 2042 . . . . [T]he findings slightly diminished the political pressure to shore up Social Security . . . . Yesterday, the leading House Democrat on Social Security, Rep. Robert Matsui of California, said the report showed that "those who claim that Social Security is bankrupt are misleading the public.[79]

In many respects this account is typical of the press coverage that the trustees' reports generate, portraying the problem, to the extent that it exists, as located far in the future.[80] The *Journal*'s own conclusion that the funds' long-term financial health improved in 2003 helps explain why release of the 2003 report "slightly diminished the political pressure to shore up Social Security."[81]

The trustees' principal measure of long-range actuarial deficit—the negative 1.92% of total payroll highlighted in Table Three—introduces another more subtle but equally pernicious bias in the reform debates, best illustrated by a report of the Social Security Advisory Council released in early 1997.[82] The report included three different proposals for solving the financial problems of Social Security, each supported by a different coalition of council members. The requirement for each solution was that it eliminate the (then-estimated) 2.17% actuarial deficit over the seventy-five-year long-

---

[78] *See, e.g.*, BUSH COMMISSION REPORT, *supra* note 2, at 66.

[79] *See* John D. McKinnon, *Medicare Outlook Appears to Worsen*, WALL ST. J., Mar. 18, 2003, at A3 (reporting on both Social Security and Medicare Reports).

[80] While the discussion in the text focuses on the *Wall Street Journal* account, other press coverage of the 2003 Trustees Report was generally comparable. *See, e.g.*, Vicki Kemper, *Benefits Outlook Is Mixed Bag*, L.A. TIMES, Mar. 18, 2003, at A20 ("The long-term financial outlook for Social Security is somewhat brighter than a year ago . . . ."); Janelle Carter, *Medicare Funds to Run Out Sooner But Social Security on Better Footing, Fund Trustees Say*, BOSTON GLOBE, Mar. 18, 2003, at A2 ("The Social Security trust fund is slightly stronger than it was a year ago . . . ."); Larry Lipman, *Outlook Better for Social Security Fund*, ATLANTA J. CONST., Mar. 18, 2003, at A1 ("Updating their projections, federal officials forecast Monday that Social Security will stay solvent a year longer than expected . . . ."). To be sure, some journalists do a better job dissecting the Trustees Report, *see, e.g.*, Sloan, *supra* note 9 (offering a more critical analysis of the 2003 Trustees Report); Marie Cocco, *We're Sailing Right into a Fiscal Hurricane*, NEWSDAY, May 27, 2003, at A23 (similar), but the 2003 Trustees Report was generally portrayed in contemporaneous press accounts as reporting good news, at least for the short term.

[81] McKinnon, *supra* note 79, at A3.

[82] 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2.

range projection period.[83] Not surprisingly, all of the Council's proposals included elements that increased short-term costs for the trust funds, and one group proposed a series of reforms that would substantially increase the trust funds' deficits for the next few decades, only to recoup these losses with new (and unspecified) tax increases starting far in the future.[84] One of the unfortunate consequences of focusing on long-range actuarial deficits is that it encourages irresponsible recommendations of this sort, which depend on future sacrifices of an unrealistic magnitude.[85]

As President Bush's Social Security Commission discovered, cash-flow accounting can also understate the effectiveness of certain solutions. A key component of the Bush Commission's proposals was a reduction in Social Security benefits to offset voluntary contributions to individual retirement accounts.[86] Under cash-flow accounting, payments into the privatized accounts constitute an immediate diminution in trust fund resources and an immediate expense. The reduction in benefits—that is, future outlays—associated with these contributions, however, is not reflected until the benefits are actually due to be paid. Typically these benefits will not be payable until many years in the future, and oftentimes outside of the seventy-five-year estimation period. Accordingly, under cash-flow accounting, the fiscal benefits of the Bush Commission's proposal were substantially understated.

B. Consolidation Within the Federal Budget Process

This Part reviews the manner in which the finances of the trust funds are currently integrated into other federal budgetary accounts. After reviewing current practices, it offers several criticisms of the present approach and then discusses its distorting effect on public debate of Social Security reform. Finally, it reveals how accrual-based accounting for Social Security would reduce the incentive to obscure budget realities by integrating annual trust fund revenues with other budgetary aggregates.

---

[83] *Id.* at 11. To be fair, the trustees' standard for long-range actuarial balance also requires that certain standards be met for interim periods beginning ten years from the beginning of the period of analysis. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63. Under these more complex requirements, a lump-sum future payment would not suffice. Reform proposals, however, routinely are evaluated solely in terms of their effect on the seventy-five-year actuarial deficit, confirming the basic point that the seventy-five-year projection tends to become the dominant measure of long-range solvency.

[84] 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 30–33 (Option III).

[85] Admittedly, there are legitimate grounds for structuring reform proposals that gradually come into effect. Indeed, one of the advantages of accrual accounting is that it facilitates transitional reforms that are more difficult to impose under current practices. *See infra* text accompanying notes 217–270. At some point, however, reform proposals are so severely backloaded as to exceed any legitimate claim of transitional relief, and the advisory report recommendation discussed in the text arguably crosses this boundary line. In any case, regardless of whether one agrees with this assessment, it is unquestionably true that the actuarial deficit measure for reform proposals fails to distinguish between changes instituted in the near term from those scheduled to be implemented at the end of the seventy-five-year period.

[86] *See* BUSH COMMISSION REPORT, *supra* note 2, at 25–41.

1. The Current Approach to Consolidating Social Security Trust Funds

TABLE FOUR

The CBO Budget Outlook in August 1998
(By Fiscal Year, in Billions of Dollars)

	Actual 1997	1998	1999	2000	2001	2002	2003 . . .	2008
Total Deficit (-) or Surplus	-22	63	80	79	86	139	136 . . .	251
Off-Budget Surplus	81	104	117	125	131	138	146 . . .	186
On-Budget Deficit (-) or Surplus (Excluding Social Security and Postal Service)	-103	-41	-39	-46	-45	1	-10 . . .	64
Memorandum:								
Off-Budget Surplus								
Social Security	81	105	117	126	130	138	146 . . .	186
Postal Service	a	a	a	a	1	0	0 . . .	0
Total	81	104	117	125	131	138	146 . . .	186

[a] Less than $500 million

Source: Congressional Budget Office, The Economic and Budget Outlook: An Update at ix, 26 (Aug. (Summary Table 1 & Table 2-1. The Budget Outlook Under Current Policies).

The federal government routinely misleads the public about the government's fiscal posture by including trust fund surpluses in budgetary aggregates. Consider, for example, excerpts from the opening paragraphs and accompanying summary table (reproduced in slightly altered form in Table Four) from an August 1998 Congressional Budget Office (CBO)[87] report, which was issued when the federal government was just beginning to report the surpluses of the late 1990s:[88]

[87] Although several federal agencies routinely prepare information on the federal budget, this Article focuses on the presentations of the CBO. A similar analysis would follow if the work of the executive branch's Office of Management and Budget (OMB) or the independent General Accounting Office (GAO) were used instead. *See generally Social Security: Long-Term Financing Shortfall Drives Need for Reform: Hearing Before the House Comm. on the Budget*, 108th Cong. (2002) (statement of David M. Walker, Comptroller General) [hereinafter *GAO Testimony*]; OFFICE OF MGMT. & BUDGET, MID-SESSION REVIEW: FISCAL YEAR 2002 (July 15, 2002) [hereinafter 2002 MID-SESSION REVIEW] (OMB budgetary analysis).

[88] Several times a year, the CBO prepares reports of this sort for the Senate and House Committees on the Budget. In recent years, the CBO has produced a substantial report on the budget and economic outlook in January with an update in August. For a list of recent CBO budget documents, see Congressional Budget Office, *Congressional Budget Office—Publications*, *at* http://www.cbo.gov/byclasscat.cfm?class=0&cat=0 (last visited Nov, 18, 2003).

> The Congressional Budget Office (CBO) projects that the federal budget for fiscal year 1998 will record a total surplus of $63 billion, or 0.8 percent of gross domestic product (GDP). If current policies remain unchanged, the surplus is expected to rise to $80 billion in 1999 and reach $251 billion (nearly 2 percent of GDP) by 2008 . . . . Excluding the surplus in Social Security and the net outlays of the Postal Service (both of which are legally classified as off-budget), the CBO's new projections show an on-budget deficit of $41 billion in 1998, which gives way to surpluses in 2002 and in 2005 through 2008.[89]

This artfully drafted prose encapsulates a common approach to discussing budgetary aggregates in Washington, D.C.

The first sentence provides the report's bottom line: the federal government was projected to run up a $63 billion surplus in fiscal year 1998.[90] This figure is repeated on numerous occasions throughout the rest of the report and was cited in numerous press accounts heralding what was widely reported as the first federal budgetary surplus since 1969.[91] Most readers could, therefore, be excused for not lingering over the third sentence of the opening paragraph, in which the report notes a distinction between off-budget surpluses and on-budget deficits for the year. What that convoluted construction explains is that certain sources of revenue—that is, those associated with Social Security and the U.S. Postal Service—are "legally classified as off-budget," meaning that Congress has voted to exclude these items from budget aggregates.[92] The sentence further informs that if those revenues were so excluded, something called the "on-budget" would show a deficit for 1998 and four of the five following years.[93] This on-budget figure was projected to reach a surplus on the order of $60 billion, but not until ten years into the future, in 2008.

Since 1998, the overall budgetary picture of the United States has followed a tortuous path, first improving through 1999 and 2000 and then, starting in 2001, deteriorating significantly.[94] For a time, particularly during the final years of the Clinton Administration, budget officials became some-

---

[89] CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: AN UPDATE ix (Aug. 1998) [hereinafter CBO AUGUST 1998 UPDATE].

[90] Ultimately, the total surplus for Fiscal 1998 was $69 billion. *See* OFFICE OF MGMT. & BUDGET, BUDGET OF THE UNITED STATES, FISCAL YEAR 2003: HISTORICAL TABLES 22 tbl.1.1 (2002), *available at* http://www.whitehouse.gov/omb/budget/fy2003/pdf/hist.pdf) (last visited Oct.17, 2003) [hereinafter FY 2003 HISTORICAL TABLES].

[91] *See* CBO AUGUST 1998 UPDATE, *supra* note 89, at xiii. *See also* Christopher Georges, *Surplus Could Hit $60 billion this Year*, WALL ST. J., May 5, 1998, at A2.

[92] *See* STANLEY COLLENDER, THE GUIDE TO THE FEDERAL BUDGET 22–23, n.11 (1995) (describing section 13301 of the Budget Enforcement Act of 1990, as amended).

[93] *See infra* text accompanying notes 110–127 (exploring problems with this formulation).

[94] As it turned out, the total federal surplus was $126 billion in fiscal year 1999, $236 billion in fiscal year 2000, and $127 billion in fiscal year 2001. *See* FY 2003 HISTORICAL TABLES, *supra* note 90, at 22 tbl.1.1.

what more circumspect in their casual combination of Social Security trust fund surpluses and those of the on-budget accounts.[95] However, confusion and obfuscation still prevailed. The repeated and often incoherent references to Social Security lockboxes in the 2000 presidential campaigns are striking illustrations of this confusion,[96] and even publications of budgetary agencies included many of the problems noted above.[97]

TABLE FIVE

The CBO Budget Outlook in January 2002

(By Fiscal Year, in Billions of Dollars)

	Actual 2001	2002	2003	2004	...	2009	2010	2011	2012	...	Total 2003-2012
On-Budget Surplus or Deficit (-)	-33	-181	-193	-141	...	-23	4	131	319	...	-242
Off Budget Surplus *	161	160	178	195	...	274	290	307	322	...	2,505
Total Surplus or Deficit (-)	127	-21	-14	54	...	250	294	439	641	...	2,263
Memorandum:											
Social Security Surplus	163	163	179	195	...	274	290	307	322	...	2,505
Postal Service Outlays	2	3	1	a	...	0	0	0	0	...	a
Total Surplus or Deficit (-) as a percentage of GDP	1.3	-0.2	-0.1	0.5	...	1.7	1.9	2.7	3.7	...	1.6

a. Less than $500 million

* Off Budget surpluses comprise surpluses in the Social Security trust funds as well as the net cash flow of the Postal Service

Source: Congressional Budget Office, The Budget and Economic Outlook: Fiscal Years 2003-2012, at 2 (Jan. 2002) (Table 1.1. The Budget Outlook Under Current Policies).

With the reemergence of budgetary shortfalls, budgetary documents became even more contorted in their treatment of Social Security surpluses. Reproduced below is the opening paragraph of the CBO's January 2002 report, and an accompanying summary table (Table Five):

> The economic recession and recent laws have combined to sharply reduce the budget surplus projected a year ago. In January 2001, the Congressional Budget Office (CBO) projected that under

[95] For an interesting discussion of efforts to clarify the distinction between on-budget and off-budget surpluses during the final years of the Clinton Administration, see DOUGLAS W. ELMENDORF ET AL., FISCAL POLICY AND SOCIAL SECURITY POLICY DURING THE 1990S 44 (Nat'l Bureau of Econ. Research, Working Paper No. 8488, 2001).

[96] For a discussion of the lockbox debate, see *infra* note 113 and text accompanying note 222.

[97] *See* CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: FISCAL YEARS 2002–2011, at xiii (Jan. 2001) (reporting a "total surplus [that] will reach $281 billion in 2001" as well as a smaller "on-budget surplus" of $125 billion that is "[p]erhaps more important to some policymakers").

> the laws and policies then in force, the federal government would run surpluses in fiscal years 2002 through 2011 totaling $5.6 trillion. In CBO's new projections, that cumulative surplus has fallen to $1.6 trillion—a drop of $4 trillion . . . .[98]

In many respects, the January 2002 report is even more misleading than prior reports. Whereas prior reports were precise, even in summary statements, to speak in terms of "total" surpluses or deficits (which in turn were decomposed into on-budget and off-budget components), the January 2002 report is more casual in its terminology, referring simply to "the budget surplus" in its opening sentence. When the introductory paragraphs eventually turn to the off-budget component of the total-budget picture, the presentation is even more convoluted. The text explains that off-budget surpluses run throughout the period but that on-budget surpluses do not "reemerge" until 2010, three years before the end of the ten-year projection.[99] The astute reader might infer that the on-budget accounts must therefore be in deficit before 2010. But what the summary doesn't say—and what many readers would likely find salient—is that without Social Security cash-flow surpluses, the on-budget deficits for 2003 and 2004 were projected to be in the range of $180 to $190 billion.[100] This information is contained in tables appearing elsewhere in the 2002 report (summarized in Table Five) but does not factor into the introductory paragraphs.[101]

---

[98] CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: FISCAL YEARS 2003–2012, at xiii (Jan. 2002) [hereinafter JANUARY 2002 CBO OUTLOOK] (citation omitted).

[99] *Id.*

[100] These projections substantially exceeded actual results. *See* CBO JANUARY 2003 REPORT, *supra* note 74, at xvi (reporting on-budget deficits of $317 billion in 2002 and total budget deficits of $158 billion); CONGRESSIONAL BUDGET OFFICE, MONTHLY BUDGET REVIEW: AUGUST 2003 (2003) [hereinafter CBO AUGUST 2003 UPDATE] (projecting total budget deficits of $455 billion in 2003), *available at* ftp://ftp.cbo.gov/44xx/doc4491/08-2003-MBR.pdf.

[101] *See also* CBO JANUARY 2003 REPORT, *supra* note 74, at xvii–xix (failing to distinguish carefully on-budget and total budget deficits in summary paragraphs on budget outlook). Press coverage immediately following the CBO's August 2003 report estimating total budgetary deficits of $401 billion in FY 2003 and $480 billion in FY 2004, *see* Congressional Budget Office, *The Budget and Economic Outlook: An Update* (Aug. 2003), *available at* http://www.cbo.gov/showdoc.cfm?index=4493&sequence=0, uncritically accepted the CBO's projected total deficits, which included Social Security cash-flow surpluses. *See, e.g.*, *$1,400,000,000,000*, ST. LOUIS POST-DISPATCH, Aug. 31, 2003, at B2; *A Dismaying Report on Federal Deficits*, MINNEAPOLIS STAR TRIB., Aug. 31, 2003, at 12AA; Associated Press, *The Markets; Stocks & Bonds*, N.Y. TIMES, Aug. 28, 2003, at c7; *Deficit Delusions*, WASH. POST, Aug. 29, 2003, at A22; Linda Feldmann, *How 'Pushback' Plays for Bush on Iraq*, CHRISTIAN SCI. MONITOR, Aug. 28, 2003, at 2; David Lazarus, *Burning Money in Iraq*, S.F. CHRON., Aug. 29, 2003, at B1; R.C. Longworth, *The Long-Term Cost of the Deficit*, CHI. TRIB., Aug. 31, 2003, at C1; *Making the Kids Pay*, L.A. TIMES, Aug. 28, 2003, at 16. In only one newspaper article that I found—a Saturday editorial—was it noted that the projected deficit in FY2003 would have been $562 billion if limited to the government's operating budget. *See The Deeper Deficit*, BOSTON GLOBE, Aug. 30, 2003, at A10. During the late 1990s, when the federal government began to run total budgetary surpluses, the press seemed to do a slightly better job of noting the extent to which those surpluses depended on cash-flow surpluses in the Social Security trust funds. *See, e.g.*, *Bleeding Social Security*, L.A. TIMES, Sept. 29, 1998, at B6 ("[T]he surplus consists

2. *Criticisms of Standard Presentation*

Apart from the political appeal of trumpeting higher budgetary surpluses or lower budgetary deficits, total-budget figures are admittedly useful statistics to compile. Several dozen pages into its analysis, the CBO August 1998 Update explains the importance of this budgetary measure, namely its relationship to the amount of government debt held by the public:

> [T]he excess of total federal revenues over total outlays is estimated to grow over the next years, rising from $80 billion in 1999 to $251 billion in 2008. If those projected surpluses are actually realized, past borrowings from the public will be partially repaid, and the debt held by the public will fall to $2.3 trillion by the end of 2008. Such a reduction in borrowing by the Department of the Treasury will release resources for private investment, thereby enhancing productivity and economic growth . . . . [Social Security] trust fund surpluses have, by law, been invested in interest bearing government securities, and that interest is part of the funds' income. Those investments have, in turn, reduced the need to borrow from the public to finance other programs.[102]

The compilation of and emphasis on total-budget aggregates are thus justified on the grounds that they represent the net amount that the federal government must borrow from (in years of deficits) and can repay to (in years of surplus) the general public. By this logic, Social Security trust fund surpluses are properly included in this aggregate because these funds reduce the government's need to obtain funds from the private sector, both decreasing the government's net borrowing and "freeing up" financial resources for other investments.[103]

To the extent that the purpose of budgetary aggregates is to communicate to Treasury Department officials the net amount of federal bonds to be sold to (or redeemed from) the general public in the course of a year, the current emphasis on total budget deficits (and occasionally surpluses) is fully justified. The management of public debt auction is not, however, the most important function of federal budgetary aggregates. The principal reason for compiling these budgetary figures is, in my view, to communicate to the general public and their representatives in Washington the amount by which

---

almost entirely of Social Security payroll taxes. Take away those revenues and federal budgets over the next five years are expected to show deficits totaling $137 billion."); Steven Thomma, *Federal Budget in the Black*, FORT WORTH STAR-TELEGRAM, Oct. 1, 1998, at 1 ("Without the Social Security money, the government would have a deficit of $32 billion.").

[102] CBO AUGUST 1998 UPDATE, *supra* note 89, at 33.

[103] It was for reasons of this sort that a presidential commission under President Johnson recommended that Social Security be brought on-budget. *See* PRESIDENT'S COMM'N ON BUDGET CONCEPTS, REPORT OF THE PRESIDENT'S COMMISSION ON BUDGET CONCEPTS 26–27 (Oct. 1967) [hereinafter BUDGET CONCEPTS REPORT].

overall financial claims on the federal government increased or decreased over the course of the reporting period.[104] Measured against this standard, the current practices of the federal government fall far short of the ideal. In particular, the government's total-budget surplus or deficit is not the most salient number to present to the general public as the best measure of the change in the government's overall financial obligations over the course of the year. At a minimum it would be preferable for reporting agencies to emphasize what the CBO refers to as the "on-budget" aggregates. However, it also would be preferable for federal budgetary reports to highlight the annual growth in both the total explicit government debt and the implicit debt associated with government programs such as Social Security.

a. Misleading Features of the Standard Presentation

A threshold complaint about the CBO presentation of budgetary aggregates is that it is arguably inconsistent with federal law. Several times in the 1980s and again in 1990, Congress, with some fanfare, voted to move Social Security off-budget, and from time to time, politicians point to this fact with pride.[105] The CBO alludes to this fact in its August 1988 report when it mentions parenthetically that the trust funds are legally required to be off-budget.[106] Throughout the report, however, it speaks almost exclusively in terms of total-budget figures that consolidate off-budget accounts. At a minimum, this approach seems unfaithful to congressional intent.[107]

---

[104] *Cf.* David M. Walker, Truth and Transparency: The Federal Government's Financial Condition and Fiscal Outlook, Speech at the National Press Club (Sept. 17, 2003) ("When it comes to the U.S. Government's financial condition and fiscal outlook, the federal government's current measurement and scorekeeping approaches leave much to be desired. The result is an incomplete and misleading picture of the federal government's current financial condition and future fiscal outlook, as well as a delay in the timeliness in which we address important issues."), *available at* http://www.gao.gov/cghome/npc917.pdf). To the extent that the composition of the federal government's financial obligations changed over the course of the year, federal budgetary statements should also communicate these changes. For this reason, I favor more comprehensive measures of the government's fiscal imbalance along the lines proposed in JAGADEESH GOKHALE & KENT SMETTERS, FISCAL AND GENERATIONAL IMBALANCES: NEW BUDGET MEASURES FOR NEW BUDGET PRIORITIES 7–15 (2003). *Infra* Part II.C.5 explains how an accrual-based system of accounting for Social Security could help compare the system's mounting levels of implicit debt to the growth of the federal government's explicit debt.

[105] For a good summary of this history, see generally DAVID S. KOITZ, SOCIAL SECURITY AND THE FEDERAL BUDGET: WHAT DOES SOCIAL SECURITY'S BEING "OFF BUDGET" MEAN? (Congressional Research Serv. Report No. 98-422, 2001).

[106] *See supra* text accompanying notes 87–89.

[107] Congress's own record in this regard is hardly exemplary. During the 1960s, Social Security was added to the unified federal budget in response to recommendations from the President's Commission on Budget Concepts. *See* BUDGET CONCEPTS REPORT, *supra* note 103, at 26–27. Through the 1980s, when public concern over mounting federal deficits led to the passage of the Gramm-Rudman-Hollings Act, Pub. L. No. 99-177 (1985), Social Security surpluses were included in budgetary aggregates, although Social Security expenditures were largely exempt from the Act's automatic sequestration procedures. *See* KOITZ, *supra* note 105, at 12. In 1990, with the passage of the Budget Enforcement Act, Pub. L. No. 101-508 (1990), Social Security was officially moved off-budget. 2 U.S.C. § 900 (2000). Deficit targets were, however, statu-

To make matters worse, the CBO further obscures the significance of trust fund surpluses through a misleading aggregation of another off-budget entity—the Postal Service—with the Social Security trust funds.[108] A reader of these references to the Postal Service might reasonably infer that, because the CBO gives both entities the same degree of prominence, both entities must make roughly comparable contributions to the total-budget aggregates. The memorandum items set forth in Tables Four and Five—information typically not included in the CBO's own summary tables—reveal that the Postal Service has no meaningful effect on budgetary surpluses.[109]

b. Familiar Criticisms of Relying on Total-Budget Aggregates

This Part focuses on more substantive complaints regarding the CBO's and other government accounting agencies' reliance on total-budget aggregates that consolidate Social Security's cash-flow surpluses as the principal summary statistic measuring the government's overall operations for any given year.

i. Proper Characterization of Trust Fund Reserves

For those who follow federal budgetary policy closely, complaints about the consolidation of Social Security surpluses with on-budget accounts are familiar.[110] Although the trust funds are statutorily required to invest their surplus reserves in government securities, the funds are not available to finance general federal expenditures in the same sense that income-tax revenues are. Social Security reserves are earmarked for the payment of Social Security benefits, and the amount of benefits is expressly tied to trust fund balances.[111] Further, the manner in which the Social Security trustees account for these reserves in their annual reports suggests that these funds have

---

torily adjusted downward (that is, higher total-budget deficits were permitted) in recognition of the fact that the movement of Social Security off-budget would otherwise impose increased pressure for on-budget accounts.

[108] *See supra* text accompanying notes 87–89.

[109] *Cf.* CBO AUGUST 1998 UPDATE, *supra* note 89, at 33–34 (acknowledging minor role of Postal Service surpluses).

[110] *See* Alan J. Auerbach et al., *The Budget Output and Options for Fiscal Policy*, 91 TAX NOTES 1639, 1639 (2002); *see also* Bob Davis, *Clinton's Simple Goal, Saving Social Security, Involves Complex Plan to Cut Debt, Boost Savings*, WALL ST. J., Feb. 2, 1999, at A9 ("The [FY 2000 proposed] budget reiterates Mr. Clinton's pledge to devote 62 percent of the budget surplus over the next 15 years to bolstering Social Security. But that is less definitive than it sounds. Most of the budget surplus already comes from the payroll taxes that fund Social Security payments. The administration is counting the same money twice, many critics charge."); David Wessel, *The Outlook: Social Security With Subtitles*, WALL ST. J., Dec. 7, 1998, at A1 ("For the next few years, the surplus exists only because Social Security is collecting more taxes than it's paying in benefits; the rest of the federal budget isn't projected to go into the black until 2002.").

[111] *See* 42 U.S.C. § 401(h) (2000).

already been factored into the long-range estimates for trust fund revenues.[112] The Social Security surpluses have, in essence, already been allocated for a particular use—one with a high degree of political support. Accordingly, this critique runs, it invites misunderstanding for the CBO to trumpet a budgetary aggregate that absorbs the trust funds' current surpluses—and it encourages politicians to behave as if they had several hundred billion extra dollars to spend. The Clinton Administration, for example, appeared to fall into this trap in the late 1990s, when it called for total-budget surpluses to be used to solve the problems of the Social Security system.[113]

In essence, this critique rests on a challenge to the way in which the federal government currently classifies the ownership of trust fund assets. The premise of a unified federal budget—that is, a budget that turns on total-budget aggregates—is that intragovernmental transactions should be consolidated to get a complete picture of the government's cash flows and hence of its financing needs.[114] Critics of current practices dispute this premise, arguing that Social Security trust fund reserves are fundamentally different from the accounts of other departments maintained with the Treasury and even from other kinds of federal trust-fund accounts, such as the government's highway trust reserves, which are deposited with the Treasury pending passage of appropriation bills and distributed periodically to the states

---

[112] *See supra* Table Two (showing that under intermediate estimates, current trust fund balances contribute a summarized revenue rate of 0.69% of taxable payroll over the next seventy-five years).

[113] *See* ELMENDORF ET AL., *supra* note 95, at 43 (discussing internal debate within Clinton Administration on this point). The 2000 presidential campaign devolved into dueling lockbox proposals, which became fodder for late-night-television comedy routines. The lockbox concept was never clearly defined. Often, the term seemed to refer to the allocation of Social Security surpluses to the trust funds. *See id.* at 73. At other times, the term seemed to be used to refer to the allocation of a portion of on-budget surpluses to the Social Security trust funds. *See id.* Exactly how this second formulation was to have been implemented is unclear. Conceivably, retired public debt could have been allocated to the trust funds. Or, to similar effect, general revenues could have been directed to the trust funds and then used to purchase government bonds in the secondary market. Ultimately, the unified surplus disappeared before any lockbox concept could be implemented and references to lockboxes have largely fallen out of the public debate. *See* Douglas W. Elmendorf & Jeffrey B. Liebman, *Social Security Reform and National Savings in an Era of Budget Surpluses*, *in* 2 BROOKINGS PAPERS ON ECONOMIC ACTIVITY 11–18 (2000); *see also* RUDOLPH G. PENNER ET AL., SAVING THE SURPLUS TO SAVE SOCIAL SECURITY: WHAT DOES IT MEAN? (Urban Inst. Brief Series No. 7, 1999).

[114] *See* BUDGET CONCEPTS REPORT, supra note 103, at 26–27. When this approach is applied, the critical question is whether a particular entity should be characterized as governmental or nongovernmental. The notion that the Social Security trust funds should be classified as governmental is not altogether implausible. As mentioned above, the trusts are an accounting fiction managed by a group of six trustees, four of whom are ex officio government officials and two of whom are public representatives appointed by the President and confirmed by the Senate. *See* 42 U.S.C. § 401(c) (2002). The putative beneficiaries of the trust funds are current and future participants and their family members. But the legal right of these beneficiaries to trust assets are not robust, particularly with respect to prospective changes in benefit formulas and rates of taxation. *See supra* text accompanying note 59. It is for this reason that supporters of consolidation of Social Security surpluses can credibly assert that the trust funds should be considered governmental agencies.

and other recipients.[115] Critics of the consolidated approach regard the government's obligation to pay Social Security benefits as among the most substantial of our government's commitments[116] and, for this reason, regard obligations to the Social Security trust funds as equivalent, for accounting purposes, to debt held by the general public. As such, Social Security trust funds should not be consolidated with other intergovernmental transfers and on-budget aggregates should replace total-budget aggregates as the principal measure of fiscal balance.[117]

ii. The Economic Realities of Social Security

Total-budget aggregates are also inconsistent with the economic realities of Social Security. A fair reading of the CBO reports described above and similar publications suggests that the decision to include Social Security surpluses in budgetary aggregates is not simply based on a formal classification of the trust funds as governmental entities but also reflects a desire to present a complete picture of the government's financial obligations and the net effect of governmental borrowing from capital markets. Even if this is true, it is not necessarily accurate to view the transactions between the trust funds and the rest of the federal government in a given year as reducing the government's borrowing or enhancing the availability of capital to other borrowers.

Consider the consequences of the $165.4 billion net increase in trust fund assets deposited with the Treasury in 2002.[118] The Treasury is required to repay these funds, with interest, just as it would have to pay back interest and principal on securities issued directly into the capital markets. As discussed above, the trustees of the Social Security system are counting on accumulated surpluses and the interest payments thereon to provide for more than five percent of the trust funds' summarized income rates over the next seventy-five years.[119] If the trustees were to distribute the entire accumulated reserves of $1.4 trillion to currently covered workers and retirees in the form of a single government note issued to each participant beneficiary and make an offsetting deduction in the benefits of these individuals, there would be no change in the financial status of the trust funds (although the funds' liquidity would be entirely eliminated). This distribution would, however, unambiguously transform Treasury fund obligations to the trust funds into debt held by the public, suggesting that, as a practical matter, Social Security trust

---

[115] *See* PATASHNIK, *supra* note 19, at 113–34.

[116] *See supra* text accompanying notes 64–65.

[117] Note that "debt held by the general public" is the standard against which government accounting officials measure aggregate borrowings of the federal government. As currently used, this formulation does not include federal debt held in the Social Security trust funds.

[118] *See supra* Table One.

[119] *See supra* Figure Two. Under intermediate assumptions over the seventy-five-year period, the beginning fund balance accounts for summarized income rate equal to 0.69% of payroll as compared with a total summarized income rate of 13.78% of payroll.

fund assets deposited in the Treasury are functionally equivalent to debt held by the public.[120]

It is also debatable whether the current operations of the Social Security system should be understood to increase the availability of capital for other forms of investment. For some time, economists have theorized that pay-as-you-go public pension systems such as Social Security might actually reduce the supply of capital, as individuals reduce their private savings for retirement in anticipation of receiving benefits promised from their public pensions.[121] There is a growing body of theoretical and empirical literature estimating the actual value of the retirement income of Social Security benefits and measuring the effect of these benefits on other types of individual savings.[122] While the results of this research are not without ambiguities, the weight of current evidence does come down on the side of Social Security promises' somewhat depressing private savings rates.[123] If this is true, the CBO's belief that Social Security's current operations "release resources for private investment"[124] is inaccurate.

c. A Deeper Critique

Although the familiar criticisms of total-budget aggregates outlined above are valid, proponents of these criticisms have not pursued them to their logical conclusions. Generally speaking, the issue has been framed as whether or not annual Social Security operating surpluses should be consolidated with the rest of the government's operating budget. The sole remedy sought by critics of current practices is the removal of these off-budget sur-

---

[120] To be sure, one could resist this point by arguing, once again, that obligations to the trust funds are different from other government securities because the government has the legal right to reduce the amount of Social Security benefits or increase the amount of Social Security payroll taxes, thereby eliminating the need for the Treasury to honor its obligations to the fund. In my view, this is a dubious point, given the magnitude of changes necessary to obviate the trust funds' need for current reserves and projected operating surpluses over the coming decade.

[121] For a summary of this literature, see generally CONGRESSIONAL BUDGET OFFICE, SOCIAL SECURITY AND PRIVATE SAVINGS: A REVIEW OF THE EMPIRICAL EVIDENCE (July 1998).

[122] *Id.*

[123] *Id.* at 3.

[124] *See* CBO AUGUST 1998 UPDATE, *supra* note 89, at 33. A few more words about the relationship between pre-funding and Social Security's negative effect on savings may be in order here. The basic point in the text is that the promises that Social Security makes to U.S. workers each year diminish other forms of savings to some degree. Having a portion of the Social Security system's new commitments funded each year (through the purchase of government securities) to some degree offsets the system's negative effect on savings. So determining the net effect of Social Security on national savings would require a comparison of the extent of pre-funding and the size of the system's negative impact on savings each year. Because the annual accrual of obligations of the system are so much larger than the annual increase in its funding, the negative effects would likely outweigh the positive effects of pre-funding, but this is simply a conjecture, which is subject to empirical validation or rebuttal. Even without such further inquiry, however, the basic point that annual gross increases in funding likely overstate the net savings effect remains valid.

pluses from budgetary aggregates. For a variety of reasons, this response is incomplete and unsatisfactory.

Even in a year in which both the on-budget aggregates and the Social Security trust funds are in perfect balance,[125] it would be incorrect to conceptualize the federal government's overall financial posture as being in equipoise with respect to the Social Security system. Admittedly, the trust funds would have made no net deposits with the Treasury in such a year. However, it is equally clear that the amount of pension benefits promised by the Social Security system might have increased. As described above, critics' objections to total-budget aggregates turn on the special nature of Social Security benefits, not on the contractual arrangements governing trust fund deposits with the Treasury. Even though the level of the trust fund reserves may not have increased, covered workers and their families could have accrued additional benefits and the value of the Social Security system's "sacred obligation" to the American people might well grow.[126] Simply moving the trust fund surpluses off-budget fails to reflect this basic truth. Those who want federal budget aggregates to reflect the special nature of Social Security's obligations faithfully must do more than move the trust funds off-budget. They must support an accounting system that more accurately estimates the size and growth of the system's actual obligations to pay future benefits.

Similarly, to the extent that one is looking to federal budgetary aggregates to get a better sense of the net effect of government operations on capital markets, simple off-budget accounting treatment for Social Security is inadequate. In a year in which the trust funds and the federal budget are in perfect balance, the net sale of government securities to the capital markets would, of course, be zero. But if covered workers and retirees continue to accrue pension benefits from the Social Security system, economic theory predicts that there will continue to be some sort of offsetting reduction in the retirement savings that these individuals would otherwise make. In the extreme, if there were a one-for-one reduction in other savings, for each dollar of newly accrued Social Security benefits, there would be a dollar reduction in private savings.[127]

---

[125] As revealed in Figure Two, the revenue and cost rates of the trust funds are expected to be in balance at some point during the next decade.

[126] Press Release, Senator John Ashcroft (May 2, 1999) (on file with author) ("Honoring the commitment to retirement security for the American people must be a sacred obligation of the government. Promises made must be promises kept, for this generation and for future generations. I welcome the House's decision to end once and for all Washington's indefensible practice of dipping into Social Security to pay for irresponsible new spending, higher deficits, or tax cuts.").

[127] Of course, the actual effect is unlikely to be this strong. In a 1974 study, Martin Feldstein originally estimated the reduction in private savings to be on the order of 30% to 50% of Social Security benefits. *See* Martin Feldstein, *Social Security, Induced Retirement and Aggregate Capital Accumulation*, 82 J. POL. ECON. 905 (1974); *see also* Martin Feldstein, *Social Security and Private Savings: Reply*, 90 J. POL. ECON. 630 (1982). *See also infra* Part IV.B.2.a (review of the life-cycle literature). However, the direction of the effect, not its magnitude, is

In short, simple off-budget treatment of Social Security obscures two important and related effects. Properly constructed budgetary aggregates cannot ignore changes in the overall financial status of the Social Security system—particularly increases in unfunded liabilities. Further, increases in the system's liabilities (that is, its promised benefits) are what make private individuals change their savings patterns and thereby reduce funds available for other forms of investment. The only way that budgetary aggregates can reflect these factors is to rely on financial accounts that reflect the accrual of Social Security liabilities.

d. Biases for Reform Proposals

Beyond the failure of current budget presentations to provide a complete picture of the financial health of the Social Security system, the current primacy of total-budget aggregates distorts the public debate over Social Security reforms in numerous ways.

i. Perverse Budgetary Incentives for Social Security Reforms

To begin with, the time horizon for federal budgetary politics is much shorter than the seventy-five-year perspective that the trustees of Social Security consider in their annual reports. Because budgetary calculations typically project only five or ten years, the effect of reform proposals over this horizon is most salient for budgetary politics. Accordingly, reform initiatives that create short-range cash surpluses and long-range growth in expenditures are favored in budget-driven compromises. Illustrations of this phenomenon are manifold in the budget wars of the past fifteen years. For example, reform proposals often entail the extension of Social Security participation to all state and local workers, many of whom are currently exempt.[128] Such reforms bring an immediate infusion of payroll taxes into the Social Security system—and additional total-budget surpluses for the federal government—while postponing increases in benefit payments until later years. From the perspective of the federal budget, this is a highly attractive reform, irrespective of whether the reform actually improves the long-term viability of Social Security.[129]

---

what is important here. For a recent review and critique of related literature on the overall impact of private pensions on overall savings, see generally William G. Gale, *The Effects of Pension on Household Wealth: A Reevaluation of Theory and Evidence*, 106 J. POL. ECON. 706 (1998).

[128] *See, e.g.*, 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 19–20.

[129] To assess such a proposal's long-term effect, one would have to analyze the wage structure and other demographic facts about the new state and local workers brought into the system. Conceivably, if a sufficient percentage of these workers were low-income, the progressive nature of Social Security benefits could make their participation a long-run net loss for the system. Measured in terms of near-term (five- to ten-year) total-budgetary effects, however, the inclusion of these workers would be a clear benefit. For an otherwise careful study of the inclu-

To make the same point in slightly more abstract terms, consider the desirability of doubling the size of the Social Security system immediately. Having worked their way through the discussion thus far of the long-range deficit problems of the Social Security system, most readers would readily agree that doubling benefits and payroll taxes would be a very dubious choice for the federal government. But such a move would, if viewed through the lens of total-budget aggregates, generate the positive outcome of decreasing the current total-budget deficit by roughly $150 billion. An accounting convention that rewards such a reform is dangerously misleading.

ii. Distorting Effects on Reforms That Change Trust Fund Operations

A separate problem with total-budget aggregates is that they generate substantial budgetary consequences for relatively modest reforms in the operation of the trust funds. The perceived budgetary consequences of certain reforms are so great, in fact, that some proposals are summarily rejected as politically infeasible or too costly.

A good example is a proposal to invest some portion of trust fund assets in the stock market.[130] There is much to be said for and against such a reform, but the merits of the idea are not of concern here. Under a total-budget approach, moving trust fund assets into the stock market has a significant budgetary cost. If the trust funds were to move $500 billion into the capital markets over the next five years—a not implausible figure—the full amount would be deducted from total-budget aggregates, as these funds would no longer be invested in intragovernment accounts.[131] The accounting effect would be to exacerbate total-budget deficits over the five-year-period, which would be a distortion in that the wealth of the trust funds would not be diminished. And from a macroeconomic perspective, the transaction is largely a wash, albeit one with potential distributional consequences.[132]

A similar kind of distortion could follow from changes in the structure of the trust funds themselves. Suppose that Congress were to decide that the Social Security program should be more independent of the executive branch so as to ensure, for example, that the collection of trust fund taxes or the

---

sion of state and local workers in the Social Security system, see ALICIA H. MUNNELL, THE IMPACT OF MANDATORY SOCIAL SECURITY COVERAGE OF STATE AND LOCAL WORKERS: A MULTI-STATE REVIEW (AARP Public Policy Institute, Paper No. 2000-11, Aug. 2000) (based on an analysis limited to the seventy-five-year time frame, reporting that the inclusion of state and local workers would reduce the system's actuarial deficit by ten percent)

[130] *See, e.g.*, 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 25–62.

[131] A similar budgetary effect would occur if trust fund assets were transferred to individual accounts. *See* CONGRESSIONAL BUDGET OFFICE PAPER, THE BUDGETARY TREATMENT OF PERSONAL RETIREMENT ACCOUNTS (Mar. 2000).

[132] For an overview of the economic effects of stock market investments, see Peter A. Diamond, *The Economics of Social Security Reform*, *in* FRAMING THE SOCIAL SECURITY DEBATE: VALUES, POLITICS, AND ECONOMICS (R. Douglas Arnold et al. eds., 1998).

investment of trust fund assets would not be used to manipulate budgetary aggregates.[133] If the management of trust assets were moved to truly independent trusts bound by traditional trust principles, the trusts could well lose governmental status for budget-scoring purposes, with the result that the trusts would be entirely off-budget entities even for purposes of budgetary aggregates. Because this change in classification would remove Social Security's cash flow surpluses from total-budget aggregates, there would be a large apparent cost to the reform, a cost with little relation to the reform's actual significance.

iii. Generating Reform Proposals with Ulterior Motives

Another problem with total-budget aggregates is their tendency to generate reform proposals designed primarily to have effects on other parts of the budget. Perhaps the best example of this phenomenon is former Senator Moynihan's perennial suggestion that Social Security trust fund taxes be cut to eliminate their distorting effect on total-budget aggregates.[134] Moynihan, who was one of Congress's leading experts on Social Security and a strong supporter of the benefits it provides, perceived other members of Congress to be using the Social Security surplus—as reflected in total-budget aggregates—as a vehicle for cutting income taxes, thereby using regressive payroll taxes as a substitute for progressive income taxes. While his first-order preference likely would have been to move to off-budget treatment for Social Security and to keep payroll taxes and income taxes at their then-current levels, he believed this solution to be politically impossible. Therefore, forced to choose a tax cut, he preferred a cut in Social Security taxes over one in income taxes, notwithstanding the deleterious effect on the system's long-range solvency.[135] Whatever one thinks of the wisdom of Senator Moynihan's political judgments, his strategy revealed the twisted effect of our current reliance on total-budget aggregates as the principal measure of fiscal balance.

---

[133] Such a scenario is not entirely far-fetched. In a budgetary controversy a few years ago, when Congress's failure to increase the limit on the national debt was seen to give the legislative branch a tactical advantage over the executive, the Secretary of the Treasury delayed investments of other trust funds in government securities in order to keep the debt within legal limitations and thereby diminish the bargaining power of Congress. *See Use of Trust Funds to Avoid Default: Hearing Before the House Comm. on Banking and Fin. Servs.*, 104th Cong. (Dec. 13, 1995) (statement of Robert E. Rubin, Secretary of the Treasury) (describing past techniques to avoid debt limit). Increasing the independence of the Social Security trust funds is one way to prevent such behavior with respect to Social Security reserves in the future.

[134] *See* PATASHNIK, supra note 19, at 88–90.

[135] *See id.*

II. An Accrual Accounting System for Social Security

This Part sketches out an accrual accounting system for Social Security. It begins with a brief discussion of the assets and liabilities that might in theory be recognized under a system of accrual accounting. It then describes a set of GAAP-style accrual-based financial statements for Social Security, including rough estimates of what such a system of accounts would have looked like for the calendar year ending December 31, 2002.[136] Next it compares this system of accounting to other measures that academic writers and other experts sometimes use to assess the long-term solvency of Social Security and explains the principal differences between these measures and accrual accounting. It then discusses how a modified form of accrual accounting, incorporating some of the valuable attributes of alternative solvency measures, could be used to frame and enhance public debate over proposals to reform the Social Security system. The Part concludes with an exploration of how such a modified system of accrual accounting might be consolidated with federal budgetary aggregates.

A. *Possible Components of an Accrual-Based System of Accounting*

Before constructing a system of accrual accounting for Social Security, one must consider which liabilities and assets such an accounting system should recognize. For this purpose, it is useful to distinguish among three different groups of participants: retirees who have completed their working careers; workers who are in mid-career; and future workers, born and unborn, who have not entered the workforce. Box Three summarizes their positions with respect to both contributions to the Social Security system and statutory entitlements to receive benefits. Current retirees have paid most of their taxes and earned their full statutory entitlements to benefits.[137] Workers in mid-career have paid some, but not all, of their taxes and earned some, but not all of, their statutory entitlements to benefits. Future workers have not yet started to pay payroll taxes or earn statutory entitlements.

---

[136] I am extremely grateful to the Office of the Chief Actuary and, in particular, Stephen Goss, the Chief Actuary of the Social Security Administration, for making this information available to me.

[137] Retirees have paid most, but not all, of their taxes, because Social Security trust fund revenues include some income taxes paid on Social Security benefits. *See supra* text accompanying notes 26–31. Retirees pay some of these taxes after leaving the workforce and ceasing to pay payroll taxes.

BOX THREE
CONTRIBUTIONS AND STATUTORY ENTITLEMENTS OF SOCIAL SECURITY PARTICIPANTS

Participants	*Contributions*	*Statutory Entitlement to Benefits*
Retirees	Most Taxes Paid	All Entitlements Earned
Workers in Mid-Career	Some Taxes Paid Some Taxes to be Paid	Some Entitlements Earned Some Entitlements to be Earned
Future Workers		
–Over Next 75 years	All Taxes to be Paid.	All Entitlements to be Earned.
–After 75 years	All Taxes to be Paid.	All Entitlements to be Earned.

The basic presentation of the trust funds' current financial position in the Trustees Report (illustrated in Table One, showing a net increase in assets of $165.4 billion in 2002) recognizes only a limited number of the elements shown in Box Three. Specifically, this basic presentation reflects all taxes that retirees and mid-career workers paid during the year. Past tax payments are also reflected in the trust funds' accumulated reserves, which accrue interest during the year.[138] In terms of benefits, the basic presentation of the trustees' report explicitly recognizes only benefit payments that are paid during the course of the year.[139] Accordingly, the basic presentation reflects only a fraction of the statutory entitlements that retirees have earned to date, and essentially none of the statutory entitlements earned to date by mid-career workers.[140] These recognition principles largely follow the logic of cash-flow accounting, under which the transfer of cash is the essential requirement of income or expense recognition.[141]

Were one to model accounting of Social Security on principles of accrual accounting developed under GAAP for the private sector, the system's financial statements would have to recognize substantially more liabilities for the systems' obligations to pay future benefits. Under GAAP, liabilities are

[138] As explained in the 2003 Trustees Report, the current value of the trust funds equals "the accumulated value of past OASDI taxes [plus accrued interest] less cost." 2003 TRUSTEES REPORT, *supra* note 8, at 62.

[139] Implicitly, past benefits payments are also recognized in the sense that they have been offset against past tax contributions.

[140] I use the qualifier "essentially" because some mid-career workers will have been paid temporary disability benefits and perhaps other ancillary benefits.

[141] The basic presentation is not fully consistent with cash-flow accounting because it recognizes the accrual of interest on reserves even though there is no corresponding cash payment.

probable future economic sacrifices based on past transactions or events. Inherent in the definition is a degree of subjectivity, as there is room for disagreement regarding the probability that benefits promised under the current statutory formula will be paid in the future. The Social Security Act currently requires those payments, but Congress is constitutionally free to alter benefit formulas by statutory amendment. For reasons explained above,[142] the likelihood of full payment of Social Security benefits for current retirees seems sufficiently high to warrant recognition of liabilities for all future benefits of retirees under GAAP. Moreover, the probability of payments to mid-career workers is sufficiently likely to warrant some recognition of obligations to these participants under GAAP, with a larger fraction of benefits recognized for older mid-career workers and a smaller fraction of benefits for younger workers. A GAAP-style accrual accounting system would recognize the system's obligations to mid-career participants in a gradual manner over the course of their working life, reflecting the growing probability that benefits will be paid in accordance with applicable statutory standards as workers approach retirement age. Box Three and subsequent discussions refer to a benefit as "earned" if there is a sufficient probability that the benefit would be recognized as a liability under a GAAP-style accrual accounting system.

In the field of social insurance, financial analysts have not followed GAAP accounting. Indeed, the dominant presentations of the financial posture of the Social Security system are the ones outlined and critiqued in Part I of this Article, under which benefits are not recognized until paid. Within limited circles of experts, however, accrual-like concepts have been employed to estimate the present value of the system's total obligations.[143] Following the logic of cash-flow projections emphasized in the trustees' reports, these alternative measures typically focus on the amount by which the present value of the system's future benefit payments exceeds the present value of future revenues and current reserves.[144] Unlike GAAP-style principles of accrual accounting, these measures do not seek to assess the likelihood that these future taxes will be contributed or future benefits paid. Rather, these

---

[142] *See supra* notes 64–65 and accompanying text.

[143] *See, e.g.*, Goss, *supra* note 43; GOKHALE & SMETTERS, *supra* note 104, 25–28; 2003 TECHNICAL PANEL ON ASSUMPTIONS & METHODS, REPORT TO THE SOCIAL SECURITY ADVISORY BOARD 87–88 (Oct. 2003) [hereinafter 2003 TECHNICAL PANEL REPORT], *available at* http://www.ssab.gov/2003TechnicalPanelRept.pdf (last visited Nov. 22, 2003). *See also infra* text accompanying notes 260–271 (discussion of international comparisons of the implicit debt of public pension plans).

[144] *See, e.g.*, Goss, *supra* note 43. *See also infra* text accompanying notes 241–257 (discussion of FASAB requirements for social insurance). Accrual accounting also employs present-value concepts for liabilities, such as retirement benefits, that are to be paid in the future. The present value of the benefit is recognized at the time the obligation is incurred, and then each year the obligation accrues an interest charge until the obligation equals the full amount of the benefit at the time of payment. The illustrative income statements presented below distinguish between the initial accrual of benefit obligations and the annual accrual of interest on previously accrued benefits. See *infra* Part II.B.2.

measures recognize the present value of all future taxes and projected benefits of covered participants under current statutory rules. These measures are thus broader than GAAP-style accrual accounting.[145] On the liability side, these alternative approaches recognize the present value of both benefits that have been earned to date and benefits that will be earned in the future. On the assets side, they recognize the present value of contributions to be made in the future plus, in some instances, residual reserves representing the excess contributions over benefit payments that have been made in the past.[146]

By convention, there are two populations for which these broader measures are calculated: the closed-group and the open-group. The closed-group is the group of people already within the Social Security System: current retirees and mid-career workers (typically defined as everyone currently fifteen years or older).[147] The open-group is the closed-group plus all future workers.[148] In the past, Social Security's Office of the Chief Actuary has typically calculated the open-group system over a seventy-five-year time horizon,[149] but recently independent experts have pushed for the measure to be calculated over an infinite horizon, thereby taking into account all future contributions to the trust funds and all future benefit payments.[150]

B. *A GAAP-Style Accrual Accounting System for Social Security*

The bare bones of an accrual accounting system for Social Security are straightforward. They should, at a minimum, include an annual balance sheet, which lists the total assets and liabilities of an institution as of a specified date, and an income statement, which gives operating results for a specific period, such as one year.[151] This Part explains how these elements can be roughly approximated from publicly available information and then presents an estimated GAAP-style balance sheet and income statement for Social Security for the calendar year 2002.

---

[145] Accrual accounting, thus, occupies an area between the trustees reports' basic presentation, illustrated by Table One, which recognizes a narrower set of taxes and benefits, and the alternative measures of long-term solvency, which reflect a broader range of taxes and benefits.

[146] *See* 2003 TECHNICAL PANEL REPORT, *supra* note 143, 87–89.

[147] *Id. at* 87.

[148] *See, e.g.*, Goss, *supra* note 43, at 32.

[149] *See id.*

[150] *See, e.g.*, GOKHALE & SMETTERS, *supra* note 104, at 44.

[151] A balance sheet offers a picture of an organization's total financial resources (assets) and obligations (liabilities) at a specific point in time. If its assets exceed its liabilities, the organization has a surplus; if its liabilities exceed its assets, it has a deficit. An income statement illuminates the extent to which an organization's operations increased or decreased its net assets during the period in question. Profits increase an organization's surplus or decrease its deficit. Losses decrease an organization's surplus or increase its deficit.

1. Balance Sheet

The chief difficulty in constructing a GAAP-style balance sheet for Social Security is deriving an estimate of the system's accrued liabilities—that is, the present value of all benefits to be paid to retirees and the portion of benefits the mid-career workers have earned to date. While the Social Security Administration does not publish its own estimate of the system's accrued liabilities, its Office of the Chief Actuary does report a statistic from which accrued liabilities can be derived. Each spring, the Chief Actuary calculates an estimate of the system's "maximum transition cost" as of the beginning of the year.[152] This measure, estimated to be $12.6 trillion as of December 31, 2002, represents an estimate of the amount by which the system's accrued liabilities—that is, its benefits earned to date for both retirees and mid-career workers—exceed the system's accumulated reserves. The Chief Actuary's methodology for accruing benefits for mid-career workers under the maximum transition cost measure is relatively conservative. Projected benefits are accrued on a forty-year straight line basis reflecting the typical working life of participants.[153] While one could propose alternative formulas for accruing benefits to mid-career participants,[154] the Chief Actuary's approach is plausible and close enough to other reasonable systems of accrual to provide useful illustrations for purposes of this Article. Were Social Security to move formally to a system of accrual accounting, more care would need to be given to the precise accrual formula used. But for current purposes, the principles of accrual implicit in the Chief Actuary's maximum transition cost estimate are acceptable.

[152] What this measure reflects is the amount that the federal government would have to pay if the Social Security system were to be shut down at the point of measurement. The maximum transition cost reveals how much money it would take beyond the amounts currently held in the trust fund reserves to pay all future payments of participants and beneficiaries accrued to date. In policy debates, this measure is sometimes described as the cost of completely shutting down the Social Security system. The measure is sometimes rejected as irrelevant to public discussion of Social Security finances because it is widely agreed that there is no political support for entirely ending Social Security. For my purposes, however, the maximum transition cost is a most useful statistic because it is quite similar to the measures of accrued liabilities that GAAP requires of private pensions. The maximum transition cost makes no allowance for future revenues nor does it consider benefits that would accrue in the future. It simply represents the difference between the total value of Social Security benefits that have earned to date and the current level of trust fund reserves.

[153] *See, e.g.*, Goss, *supra* note 43, at 34.

[154] *See* Box Four.

BOX FOUR
HOW SHOULD SOCIAL SECURITY BENEFITS ACCRUE?

Accrual accounting for Social Security would spread the costs of retirement benefits over the working lives of participants. Under such a system, retirement benefits should be fully accrued when workers retire, but there are a range of methods under which the growth of benefits could be recognized over the working lives of participants.

The actual statutory requirements of the Social Security Act, for example, could be used to determine the accrual formula. Under this approach, the trust funds would recognize retirement benefit obligations when participants have earned the legal right to receive those benefits under current statutory requirements. This approach, however, would produce a number of anomalies. Forty quarters—that is, ten years—of participation is required before participants are entitled to retirement benefits. So under this formula, workers would not, strictly speaking, earn any benefits until they had worked for ten years. On the other hand, because Social Security benefit formulas are highly progressive, participants earn relatively large amounts of benefits once they pass the ten-year vesting threshold and relatively smaller amounts of benefits in their later years of work.

The Chief Actuary's estimate of maximum termination cost uses a forty-year straight line method of accrual, in which projected benefits are allocated evenly over an assumed forty-year working life for participants. The illustrations of accrual accounting presented in this article are based on this methodology. Compared to the Social Security statutory formula, the forty-year straight-line approach accrues benefits at a faster rate during the first ten years of participation and at a slower rate thereafter. While not tracking statutory requirements, this accrual formula provides a defensible allocation of costs over time. But one could also imagine accrual formulas with shorter periods of accrual or more complex accrual formulas.

For purposes of developing a sensible formula of accruing Social Security benefits, a number of considerations would be relevant. In addition to the statutory formula and the actual experience of participants, one would also want to recognize that benefits accrued under this formula might become more difficult for Congress to change than benefits that have yet to accrue. In addition, special rules of realization might be appropriate for disability and other non-retirement benefits paid out from the trust funds.

With the Chief Actuary's estimates of maximum transition costs, it is a relatively straightforward matter to sketch out a GAAP-style accrual-based balance sheet. The system's only material assets are the reserves of the trust funds, which at year-end 2002 equaled about $1.4 trillion. The system's accrued liabilities equal the sum of its current reserves ($1.4 trillion) and the Maximum Transition Cost ($12.6 trillion), and therefore reached $14.0 trillion as of December 31, 2002.[155] Based on these estimates, Figure Four presents a balance sheet as of December 31, 2002, with one slight embellishment. Liabilities have been segmented into classes based on the age cohort of the participants with whom the accrued benefits are associated. Based on information reported in the Social Security Administration's own annual report (as opposed to the trustees' annual reports discussed above), about one-third of the system's current liabilities or roughly $4.6 trillion of claims are estimated to represent accrued obligations to participants at or above the age of sixty-two,[156] the age at which Social Security retirement benefits are generally available.[157] In other words, the Social Security system owes current retirees accrued benefits with a present value equal to more than three times the system's current reserves. The remaining accrued liabilities—roughly $9.4 trillion in accrued claims—consist of obligations to younger participants. A full-blown system of accrual accounting could easily distinguish among other classes of claimants by, for example, separating out accrued liabilities by generational cohorts.[158]

[155] This estimate probably understates the level of accrued liabilities that would be reported if the financial statements of Social Security were constructed in accordance with the actual GAAP rule for private pension plans. The Office of the Chief Actuary's forty-year accrual period is longer than the thirty-five-year period under which benefits accrue under the Social Security Act. *See* Social Security Handbook, *How are the Average Monthly Earnings (AME) or the Average Indexed Monthly Earnings (AIME) Computed?* (Last revised Mar. 2001), *available at* http://www.ssa.gov/OP_Home/handbook/handbook.07/handbook-0701.html (explanation of how benefits are calculated). *See also* STEUERLE & BAKIJA, *supra* note 60, at 76. It is also longer than comparable periods of accrual for purposes of private pension plans. *See* DAN M. MCGILL ET AL., FUNDAMENTALS OF PRIVATE PENSION 213–14 (7th ed. 1996). Shorter periods of accrual tend to raise the net present value of accrued liabilities.

[156] *See* SOC. SEC. ADMIN., FY 2002 PERFORMANCE AND FINANCIAL REPORT 91 (2002) (estimating the actuarial present value of estimated benefits for participants 62 years of age and older to be $4,401 billion as of January 1, 2002, which was 32.9% of the system's total projected benefits at the time: $13,374 billion).

[157] *See* Social Security Handbook, *Entitlement to Retirement Insurance Benefit* (Last Revised Mar. 2001), *available at* http://www.ssa.gov/OP_Home/handbook/handbook.03/handbook-0301.html (entitlement to retirement insurance benefit).

[158] These distinctions could be useful in explaining the differing impacts of various reform proposals.

FIGURE FOUR

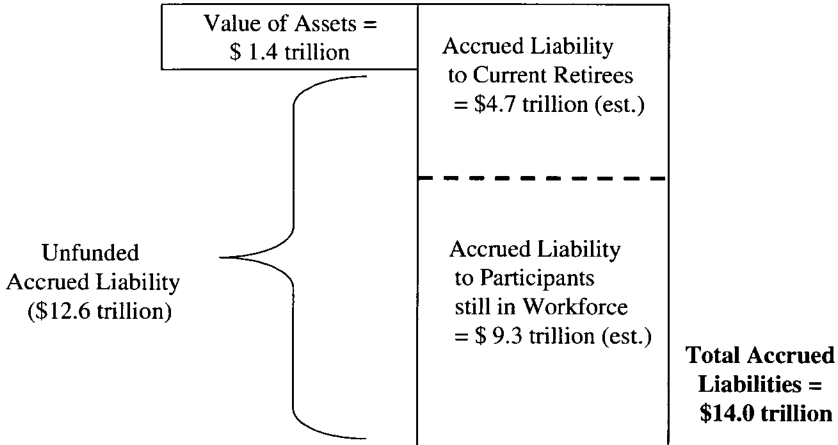


Figure Four conveys an important message, and one strikingly at odds with the basic presentation found in the trustees' reports. It indicates that the accrued commitments of the Social Security system to current retirees and mid-career workers ($14 trillion) is ten times the value of accumulated reserves ($1.4 trillion). The difference between these two figures, $12.6 trillion, is the amount of funding required to honor Social Security benefits that have been earned to date, using the definition of accrual implicit in the maximum-transition-cost estimate. Were Social Security to conform to GAAP accounting requirements, the system's unfunded accrued liability would have a magnitude approximately equal to the one depicted in Figure Four.

2. Income Statement

TABLE SIX

Income Statement for Trust Funds (est.)
(Jan. 1, 2002 through Dec. 31, 2002)

Revenues	
Payroll Taxes [known]	$ 532.5 billion
Income from Taxation [known]	$ 14.2 billion
Interest on Trust Fund Assets [known]	$ 80.4 billion
	$ 627.1 billion
Expenses	
Administrative Expenses [known]	($ 4.2 billion)
Other Costs [known]	($ 3.7 billion)
Interest Charge [derived]	($ 786.9 billion)
Net Accrual of Liabilities [derived]	($ 299.8 billion)
	($ 1,094.6 billion)
Profit (Loss) from Operations	**($ 467.5 billion)**

Faced with a balance sheet reporting Social Security's substantial unfunded accrued liabilities as of year end 2002, one might naturally ask whether the system's overall solvency has improved or worsened in the course of the past year. An income statement for the past year is the accounting format best suited to answer this question. Table Six presents an accrual-based income statement for the Social Security system for the past year, comparable to the balance sheet present in Figure Seven.

Under accrual accounting, an entity's income during a period is equal to the difference between its net worth at the beginning of the period and its net worth at the end of the period (adjusting for dividends and capital contributions, which are not at issue here). As mentioned, the Social Security system's net worth at the end of 2002 was negative $12.6 trillion, and the system's net worth a year earlier was negative $12.2 trillion.[159] Thus, simple accounting

[159] This is the Chief Actuary's estimate of the maximum transition cost as of December 31, 2001.

logic dictates that the system suffered a loss of some $467.5 billion during the course of the year.[160]

With this estimate of the system's annual loss for the 2002 calendar year, one can construct a rough approximation of the system's income statement as well. The system's revenues for the year are known and consist of payroll taxes ($532.5 billion), tax receipts ($14.2 billion), and the yield on assets in the trust fund ($80.4 billion), for total revenues of $627.1 billion. Estimating the system's expenses on an accrual basis is a bit more complex but possible. Two expenses of the system are known: the administrative expense of operating the system during the year ($4.2 billion) and a limited number of other expenses ($3.7 billion).[161] However, because the system lost $467.5 billion in 2002, it must have incurred some $1,086.7 billion in additional expenses during the year.[162] Conceptually, these expenses must be divided principally between the accrual of new benefit promises and the accrued interest on previously accrued benefit promises. Because the amount of pension promises that were outstanding at the beginning of 2002 and the amount of pension promises that were liquidated in the course of the year are known, it is possible to estimate the interest cost of the Social Security system as $786.9 billion in 2002.[163] This means that the amount of newly accrued benefit promises must have been on the order of $299.8 billion.[164]

---

[160] To be more precise, the trust funds' negative net worth was $12,629.3 billion at the end of 2002 and $12,161.8 billion at the end of 2001.

[161] The other expenses consist principally of annual contributions to another pension system, one for retired railway workers.

[162] The $467.5 billion in annual loss is explained above as the amount by which the system's maximum transition cost (or negative net worth) increased in 2002. This implies that the system's expenses must have exceeded its revenues by $467.5 billion. Total revenues have been calculated above to be $627.1 billion. This means that total expenses were $1,094.6 billion (the difference between positive $627.1 and negative $467.5). So far, I have only identified $7.9 billion in expenses in 2002 ($4.2 billion in administrative expenses plus $3.7 billion in other costs). Therefore, $1,086.7 billion of expenses remain to be identified.

[163] At the beginning of 2002, the trust funds had $13,374.6 billion of total accrued liabilities ($1,212.5 billion of which were supported with cash reserves and the remaining $12,161.8 billion of which were unfunded liabilities). According to the 2003 Trustees Report, $454 billion in benefits were paid during the course of 2002, implying that $12,920.5 billion of the accrued liabilities remained outstanding. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 27. A 6.09% interest rate—the ultimate valuation interest rate used in the estimates of Office of the Chief Actuary—suggests an interest charge of $786.9 billion for the year.

[164] This figure represents the difference between the amount of expenses and the $786.9 billion in interest charge derived in the preceding footnote. As a residual category, this net accrual of liabilities incorporates a number of factors (such as changes in a variety of technical estimates) and thus can fluctuate considerably from year to year. A better estimate of the annual rate of accrual of net liabilities can be obtained by averaging the net accruals over a number of years. *See* Table Seven (presenting five years of income statements).

TABLE SEVEN

Estimated Income Statement for Trust Funds

Five Years: 1998 to 2002

(billions of dollars)

Revenues	*1998*	*1999*	*2000*	*2001*	*2002*
Payroll Tax	$430.2	$459.6	$492.5	$516.4	$532.5
Income from Taxation	$9.7	$11.6	$12.3	$12.7	$13.8
Interest on Trust Fund Assets	$49.3	$55.5	$64.5	$72.9	$80.4
Other	$0.0	$0.0	($0.8)	$0.0	$0.4
Total Revenues	$489.2	$526.6	$568.4	$602.0	$627.1
Expenses					
Administrative Expenses	($3.5)	($3.3)	($3.8)	($3.7)	($4.2)
Other Costs	($3.8)	($3.8)	($3.7)	($3.3)	($3.7)
Interest Cost on Previously Accrued Liabilities	($626.6)	($674.9)	($724.3)	($750.6)	($786.9)
Net Accrual of New Liabilities	($513.7)	($503.7)	($713.8)	($299.2)	($299.8)
Total Expenses	($1,147.6)	($1,185.8)	($1,445.5)	($1,056.8)	($1,094.6)
Net Profit (loss)	($658.4)	($659.2)	($877.1)	($454.8)	($467.5)

Based on the same methodology, Table Seven presents a series of income statements for the trust funds for the past five years. This series of statements reveals that the $467.5 billion loss in 2002 was not an anomaly. Indeed, when this loss is compared to the system's results for the past half decade, the 2002 performance was a relatively good year. On average, the Social Security system lost $623.7 billion a year between 1998 and 2002 on an accrual accounting basis.

TABLE EIGHT

Unfunded Accrued Liabilities of Trust Funds

1992–2002 (year end)

	1997	1998	1999	2000	2001	2001
Current Dollars (billions)	$9,512	$10,170	$10,829	$11,707	$12,161	$12,629
Percent of GDP	116.09%	117.51%	118.48%	120.50%	121.23%	122.18%

Another way to present the financial position of the Social Security trust funds over the past five years is to chart the system's unfunded accrued liability over that period. Table Eight offers such a presentation, first providing the level of unfunded deficit at year end in current dollars and then presenting the deficit figure as a percentage of GDP.[165] Under both measures, the system's financial posture has clearly deteriorated. In the five years, between year-end 1997 and year-end 2002, the unfunded accrued liabilities of the system increased by more than $3 trillion from $9,512 billion at year-end 1997 to $12,629 billion at year-end 2002. Even expressed as a percentage of GDP, the deficit increased from 116.1% to 122.2% of the GDP.

Subsequent discussions address the difficult normative question of whether this magnitude of unfunded accrued liabilities should be viewed as problematic. For current purposes, it is important to note the stark contrast between the direction Social Security appears to be going under an accrual-based accounting system and that indicated by the trustees' reports. Under the latter approach, the system has been adding approximately $150 billion to its cash reserves for each of the past five years. Under a GAAP-style accrual accounting, the system's annual returns have been at least $600 billion lower, incurring losses in excess of $450 billion a year.

C. Comparing Accrual-Based Financial Statements with Other Measures of Long-Term Solvency

Academic writers and experts sometimes use other summary statistics to estimate the long-term solvency of the Social Security system. Because these measures sometimes factor into expert discussion and have recently been reported as supplementary information in the trustees' reports,[166] they are important enough to merit comparison with the GAAP-style accrual-based system. This Part engages in such a comparison and concludes by demonstrating how some of the advantages reflected in these alternative measures might be grafted onto a modified accrual-based system.

1. Closed-Group Unfunded Obligation

One common measure of the Social Security system's long-term solvency is the closed-group unfunded obligation or closed-group liability.[167]

---

[165] By construction, the decline in the system's unfunded accrued liabilities each year is the same as the trust funds' operating losses for the year. So, the difference between the system's accrued liability between year-end 1997 ($9,512 billion) and year-end 1998 ($10,170 billion) was $658 billion, the same as its net loss for 1998 as reported in Table Seven.

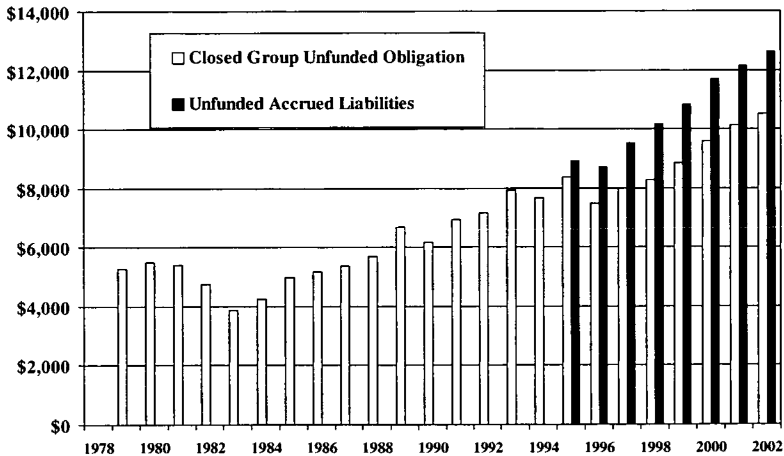
[166] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63–65.

[167] Under generally accepted accounting principles for government entities, the Federal Accounting Standards Advisory Board requires that this measure, and its component parts, be included as supplementary information. *See infra* text accompanying notes 241–259. This information now appears in the Financial Report of the United States and the Social Security Administration's annual financial statements. *Id.* In the 2003 Trustees Report, the closed-group

This measure equals the present value of future contributions that mid-career participants will make to the system, plus the current reserves in the trust funds, minus the present value of all payments to be made to current retirees and mid-career participants in the future.[168] This measure assumes that no new participants are allowed to enter the system, and that current participants continue in the system under the current rules for the rest of their and their beneficiaries' lives.[169] In other words, the system's closed-group liability represents what the net obligations due under the current system would be if the system were closed to new entrants. For this reason, the measure is sometimes referred to as the closed-group transition cost.[170] This measure would be appropriate if one were trying to evaluate the solvency of a system under which all current participants were to be grandfathered into the existing system and all new workers were to be steered into a new system.

FIGURE FIVE

Closed-Group Unfunded Obligation: 1978 to 2002
(Billions of Dollars)



unfunded obligation was also reported in a new section of the report entitled *Additional Measures of OASDI Unfunded Obligations*. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63.

[168] *See* 2003 TECHNICAL PANEL REPORT, *supra* note 143, at 88.

[169] *Id.*

[170] *See* Goss, *supra* note 43, at 33 tbl.3.

Figure Five shows the Office of the Chief Actuary's estimate of the closed-group unfunded obligation of the Social Security system over the past few decades. Also included are estimates of the system's accrued unfunded liabilities for the years in which such estimates are available. As of December 31, 2002, the system's closed-group unfunded obligation was $10.5 trillion. This is how much the government would have to raise today to finance fully a grandfathered Social Security system for current retirees and mid-career workers. Notice that this figure is roughly $2 trillion less than the system's unfunded accrued liability of the same date. This difference represents the amount by which the present value of future taxes to be paid by mid-career workers (and some retirees) exceeds the present value of additional entitlements that mid-career workers will accrue in the future.[171] The system's overall deficit shrinks a bit if all current participants are required to stay in the system for the rest of their lives. Put slightly differently, one could say that this $2 trillion represents the excess taxes that current participants are going to pay into the Social Security system for the privilege of earning additional benefits from the system over the balance of their working lives.

2. *Open-Group Unfunded Obligation: Seventy-Five-Year Horizon*

Another summary statistic representing the long-term solvency of the trust funds—and the measure most commonly cited by the Social Security Administration[172]—is the open-group unfunded obligation estimated over a seventy-five-year horizon or, for short, the seventy-five-year open-group liability.[173] Like the closed-group measure, this open-group measure reflects the difference between the present value of future tax revenues and the present value of future benefits.[174] The open-group measure is, however, broader than the closed-group measure in that it includes taxes paid by and benefits paid to future participants insofar as such payments are made over the next seventy-five years.[175] The seventy-five year open-group liability is closely

---

[171] One could decompose this figure into the net present value of future taxes and the net present value of future benefits. As of December 31, 2001, the present value of future contributions from current participants was in excess of $13 trillion, and the present value of benefits to be earned by current participants in the future, over and above the benefits accrued to date, was approximately $11 trillion. *See* SOC. SEC. ADMIN., PERFORMANCE AND ACCOUNTABILITY REPORT: FISCAL YEAR 2002, at 77 (2002) [hereinafter 2002 PERFORMANCE REPORT], *available at* http://www.ssa.gov/finance/2002/fy02PAR.pdf.

[172] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 3, 61–62. Occasionally, this measure of solvency is even picked up by the general press. *See, e.g.*, Cocco, *supra* note 80, at A23 ("The Social Security trustees report that in today's dollars, a transfer of $3.5 trillion from general government revenues, if made now, would enable the retirement system to pay the Baby Boomers their promised benefits.").

[173] A more complete description of the open-group unfunded obligation, as well as of the other measures of accrued liability discussed in this Part, may be found in Goss, *supra* note 43, at 31–33.

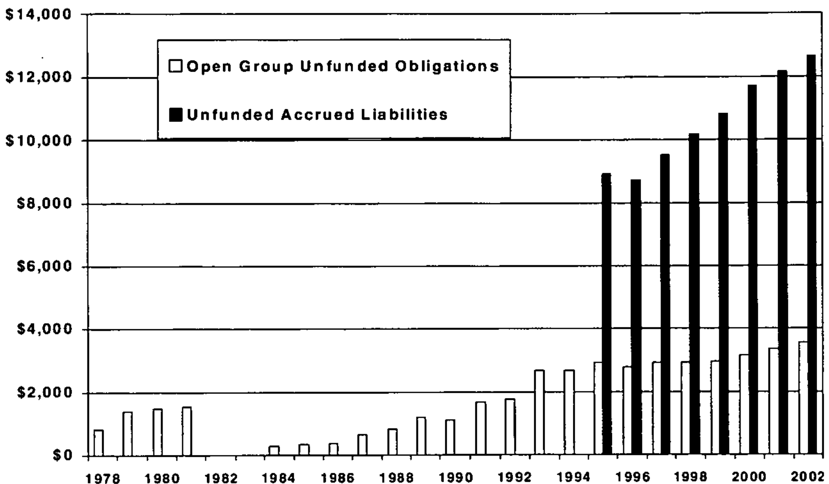
[174] *See* 2003 TECHNICAL PANEL REPORT, *supra* note 143, at 87.

[175] *See* Goss, *supra* note 43, at 33 tbl.3. To a limited degree, the seventy-five-year open-

related to the trust fund actuarial deficits previously discussed. The actuarial deficit (negative 1.92% of payroll as of year-end 2002) is the amount by which payroll taxes would have to be increased over the next seventy-five years for the system to meet its projected obligations. The seventy-five-year open-group liability is the net present value of these additional taxes.

FIGURE SIX

Open-Group Unfunded Obligations: 1978 to 2002 (Billions of Dollars)



As of December 31, 2002, the system's seventy-five-year open-group liability was estimated to be negative $3.5 trillion.[176] Figure Six reports the open-group obligations over the past few decades. Notice that around 1983, the seventy-five-year open-group liability fell to zero. This is because, after

---

group liability measure is also narrower than the closed-group measure. First, some current participants—i.e., the very young mid-career participants—will receive benefits more than seventy-five years in the future. These payments factor into the closed-group liability measure but not the seventy-five-year open-group liability measure. Also, the current reserves of the trust funds are typically deducted from the closed-group liability measure on the assumption that they would be used to finance the benefits of current participants if these participants were grandfathered under current law. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63 tbl.IV.B8. Because the open-group measure is generally considered to be an estimate used to evaluate the solvency of the system on an on-going basis, the reserves are not fully available to finance benefits over the next seventy-five years. *See supra* text accompanying notes 153–159. At least a portion of these benefits must be set aside to maintain minimum acceptable trust fund ratios.

[176] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 61.

the 1983 reforms, the system was projected to meet its obligation for the relevant period of reference: seventy-five years.[177] After 1983, the open-group unfunded liability began to grow. This reflects the cliff effect[178] as well as unrelated changes in actuarial assumptions.[179] After 1983, the system's seventy-five-year open-group liability increases, because the period of analysis includes an increasing number of years in which cash payments to the system will substantially exceed cash inflows.

Figure Six also includes estimates of the system's unfunded accrued liabilities for the years in which these estimates are available. Notice the substantial difference between the unfunded accrued liabilities and the seventy-five-year open-group liability. As of year-end 2002, the system's unfunded accrued liabilities ($12.6 trillion) exceeded its seventy-five-year open-group liability ($3.5 trillion) by about $9.1 trillion. Because we know from our analysis of closed-group liability that current participants will contribute excess taxes of about $2 trillion over the balance of their lives (predominantly within the next seventy-five years), the remaining $7.1 trillion must come from the excess of the present value of taxes to be paid by future participants over benefits to be paid to future participants in the next seventy-five years. Even with these excess taxes, the seventy-five-year open-group liability is negative $3.5 trillion. The foregoing discussion reveals that the relatively small size of the seventy-five-year open-group liability measure is largely explained by excess tax revenues to be contributed by future participants over the next seventy-five years.

3. *Open-Group Unfunded Obligation: Infinite Horizon*

The seventy-five-year open-group liability measure is, perhaps, an understandable extension of traditional Social Security accounting. The trustees' reports' long-term projections go out only seventy-five years, the same time frame as the open-group liability measure. For the reasons outlined in Part I, however, limiting analysis to cash flows over seventy-five years is inherently misleading. It ignores all of the benefits future workers should receive at the end of the seventy-five year horizon and it produces the cliff effect. For this reason, many academic writers prefer a long-term balance measure that looks beyond the seventy-five-year horizon.[180] The infinite-horizon open-group liability measure offers such a tool. It continues the open-group methodology indefinitely, factoring in all future tax revenues and all future benefits.[181] The Office of the Chief Actuary has not yet developed a

---

[177] *See* Goss, *supra* note 43, at 33 tbl.3 (reporting positive seventy-five-year open-group surplus at year-end 1983).

[178] *See supra* note 75 and accompanying text; *see also infra* Part III.B.4.

[179] *Cf.* 2003 TRUSTEES REPORT, *supra* note 8, at app. B (reviewing changes in actuarial balance since 1983).

[180] *See generally* GOKHALE & SMETTERS, *supra* note 104.

[181] *See* 2003 TECHNICAL PANEL REPORT, *supra* note 143, at 87–88.

times series of infinite-horizon open-group liability estimates. The 2003 Trustees Report did, however, for the first time, include a single estimate of the system's infinite-horizon open-group liability as of December 31, 2002.[182] The estimate was $10.5 trillion—exactly the same as the Chief Actuary's estimate of the system's closed-group liability.[183] This equality implies that the present value of benefits to be paid to future participants is equal to the present value of the future taxes that these participants will pay. Other academic experts have estimated that the present value of taxes to be paid by future participants is $3.5 trillion greater than the present value of their benefits, implying an infinite-horizon open-group liability of about $7 trillion.[184] The distance between these estimates comes from differences in demographic projections about future life expectancies and economic assumptions about discount rates.[185] But the main point is that both estimates indicate that the infinite-horizon open-group liability is much larger than the Chief Actuary's estimate of the seventy-five-year open-group liability.

4. *GAAP-Style Accrual Accounting vs. Alternative Measures of Long-Term Solvency*

Putting aside superficial differences in presentation format,[186] there are two important theoretical differences between GAAP-style accrual accounts and alternative measures of long-term solvency. First, the GAAP-style measures distinguish between benefits that have been earned (or accrued) to date and benefits to be earned (or accrued) in the future, whereas the alternative measures factor in all benefits to covered participants, regardless of when they are earned. Second, the alternative measures factor in taxes to be paid in

---

[182] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63.

[183] *Id.*

[184] *See* GOKHALE & SMETTERS, *supra* note 104, at 25–28.

[185] *Id.*

[186] Although the alternative measures are typically expressed in terms of a single statistic (the closed-group liability or infinite-horizon open-group obligation), one could decompose these measures into a series of assets—current reserves plus the net present value of future tax revenues from various groups of participants—and liabilities—such as the present value of future benefits to various groups of participants. As explained below, FASAB's requirements for social insurance require these elements to be reported in the Required Supplementary Stewardship Information section of financial statements for government entities that seek to comply with GAAP for federal entities, although FASAB does not require these elements to be included as balance sheet entries. *See infra* Part IV.A.2. The trustees' reports do not attempt to comply with FASAB requirements. In the summer of 2003, FASAB amended its rules governing the placement of this information, calling for a more prominent display of the information in a Statement of Social Insurance, or SOSI. *See* RECLASSIFICATION OF STEWARDSHIP RESPONSIBILITIES AND ELIMINATING THE CURRENT SERVICES ASSESSMENT, Statement of Federal Financial Accounting Standards No. 25 (Fed. Accounting Standards Advisory Bd. July 2003), *available at* http://www.fasab.gov/pdf/sffas-25.pdf (last visited Nov. 21, 2003) [hereinafter RECLASSIFICATION STATEMENT]. Unlike the prior supplementary information, the SOSI is considered to be a basic financial statement and subject to audit requirements. It remains to be seen how this change will affect the Trustees Report and other public disclosures regarding Social Security.

the future, whereas GAAP-style accrual accounts do not. These differences reflect the goals that the two approaches are designed to achieve, and each approach has its advantages.

a. Distinguishing Benefits Earned-to-Date from Total Benefits

Derived from traditional long-term cash-flow analyses of payments, the alternative measures are concerned with estimating the present value of all payments projected to be made for the relevant time horizon to covered participants, assuming the continuation of current statutory benefit and tax formulae.[187] Therefore, there is no reason for them to distinguish between benefits earned before or after the date of the forecast. The rules of accrual under GAAP, however, recognize liabilities only when they arise out of past transactions or events and have a sufficient probability of being paid. For obligations to make payments arising from earnings records stretched over multiple decades, GAAP-style accounting demands differentiation between benefits earned to date and benefits earned in the future.

The key question, of course, is which approach is more appropriate. In contrast to alternative measures, GAAP-style recognition principles privilege benefits earned to date. Is this privileging appropriate for the financial statements of Social Security?

The statutory requirements for private pension plans may offer some guidance. ERISA prohibits employers from reducing pension benefits that have accrued; however, employers are free to change or even eliminate the accrual of additional benefits in the future.[188] The justification of this restriction is that workers should be entitled to rely on an employer's promises to pay benefits that have accrued. While Congress remains constitutionally free to impose retroactive reductions in the benefits of Social Security, the rules governing private employers may still provide useful normative guidance. At least for purposes of accounting recognition—where a collective estimate of the trust funds' obligations is made—the government arguably should be held to at least as high a standard for benefit obligations as it imposes on private employers for their pension promises.

Another justification for recognizing only earned benefits is the underlying logic of GAAP recognition rules. Earned benefits are the ones that, in all likelihood, will result in future payment, and GAAP-style financial statements recognize such commitments because their principal purpose is to demonstrate the relationship between an entity's current economic resources and its current obligations.[189] Arguably, the financial statements of Social Security should also highlight the difference between the system's current

---

[187] *See supra* note 144 and accompanying text.

[188] Employee Retirement Income Security Act § 204(g), 29 U.S.C. § 1054(g) (2000).

[189] *See* GERALD I. WHITE ET AL., THE ANALYSIS AND USE OF FINANCIAL STATEMENTS 73 (2d ed. 1997).

assets and the value of all benefits the payment of which is sufficiently probable to warrant recognition under GAAP for private firms. Under this principle, whenever the commitments become probable obligations, that moment is the moment when they should appear on the financial statements of Social Security and be included in other federal budgetary aggregates. To the extent that financial statements are intended to help monitor and control the creation of financial obligations, this feature of accrual accounting has undeniable appeal.

Finally, there is the issue of transitional costs. In reform discussions, particularly those involving the suggestion of the partial privatization of Social Security benefits, analysts sometimes overlook the accrued claims of current workers and retirees.[190] Requiring the trust funds to produce financial statements that recognize these accrued liabilities would make it much less likely that these important obligations would be ignored.

b. The Inclusion of Benefits-To-Be-Earned and Future Taxes

The case against privileging earned benefits over to-be-earned benefits and for relying on the broader alternative measures described above has two separate strains: one political and one financial.

The political argument rejects the proposition that benefits to be earned in the future are any less binding than benefits that have been earned to date.[191] Often advanced by defenders of a strong social welfare safety net, this view regards the government's obligation to pay future generations' obligations as indistinguishable from its commitment to pay the next check to a current retiree.[192] Sensing perhaps that distinguishing between earned benefits and benefits-to-be-earned will begin to unravel the traditional Social Security system, proponents of this view prefer alternative measures of long-term solvency that treat all benefit payments the same.[193] Irrespective of one's views of the merits of the political prediction regarding the implications of adopting a

---

[190] *See, e.g.*, MICHAEL TANNER, THE BETTER DEAL: ESTIMATING RATES OF RETURN UNDER A SYSTEM OF INDIVIDUAL ACCOUNTS 15–16 (Cato Institute Social Security Paper No.31, 2003) (erroneously assuming the seventy-five-year open-group actuarial deficits as reported in the Trustees Report to be a reasonable estimate of transition costs and then discounting the need to account for transition costs on the grounds that these costs should be covered through a reduction in federal spending), *available at* http://www.socialsecurity.org/pubs/ssps/ssp31.pdf.

[191] For a good example of this view, see Peter A. Diamond & Peter R. Orszag, *Assessing the Plans Proposed by the President's Commission to Strengthen Social Security Reform*, 96 TAX NOTES 703, 705–06 (2002) (defending the use of "scheduled benefits" as the appropriate baseline for comparing reform proposals).

[192] *Id.*

[193] As far as I know, this precise objection to accrual accounting has not been made in print. Indeed, defenders of cash-flow accounting sometimes object to accrual accounting on the grounds that it might make it more difficult to reduce accrued benefits. *See* Peter A. Diamond & Peter R. Orszag, *Accrual Accounting for Social Security*, 41 HARV. J. ON LEGIS. 173 (2004). The same analysts, however, then insist on scheduled benefits—which include accrued as well as to-be-accrued benefits—to be the appropriate baseline for analyzing reform proposals. *See* Diamond & Orszag, *supra* note 191, at 705–06.

system of accrual accounting for Social Security, one could still object to this line of argument as unpersuasive in its claim that the government's obligation to make payments to future generations under current statutory formulae are appropriately considered to be equivalent to the government's obligations either to current retirees or to mid-career workers nearing retirement age.

The financial argument against privileging earned benefits proceeds from the claim that the logic of GAAP-style accrual accounts does not apply to government programs such as Social Security.[194] With private employers, GAAP is critically concerned about unfunded accrued liabilities for fear that unfunded promises will not be paid.[195] With public pension programs, on the other hand, the amount of unfunded accrued liabilities is not as important because the government's power to tax can cover the shortfall or alter benefits. Accordingly, the alternative measures of long-term solvency are more appropriate for public pension plans because they factor in the economic resources—tax contributions in the future—that are intended to finance benefits earned to date and those earned in the future.[196] Under this reasoning, the closed-group and infinite-horizon group liability measures are better presentations of the system's financial posture than GAAP-style accounts.[197]

In contrast to the political argument, this financial argument is well-founded and important. By reflecting future excess tax revenues, the alternative measures of long-term solvency do have some advantages. The question then is how GAAP-style accounts might be modified to incorporate these advantages while, at the same time, retaining the clarity of GAAP accrual accounting, which highlights the value of accrued liabilities.

5. *Modified Accrual Accounting*

Augmenting GAAP-style accrual accounting statements of Social Security to recognize as a "quasi" asset an amount equal to the present value of excess revenues to be contributed by system participants over the additional benefits that they will accrue during the balance of their working lives would resolve this dilemma. Such a system does not diminish the privileged status of earned benefits as only benefits accrued to date would appear on the liabilities side of the Social Security trust fund balance sheet, but it would reveal how much of those accrued liabilities is matched by the present value of excess contributions. The difference between the system's accrued liabilities and the sum of its reserves and excess tax contributions could be de-

---

[194] *See* Diamond & Orszag, *supra* note 193.

[195] *See* EMPLOYERS' ACCOUNTING FOR PENSIONS, *supra* note 66.

[196] If the goal here is to calculate the net value of future taxes, alternative measures must account for both future taxes and future benefit payments.

[197] The seventy-five-year open-group measure is not defensible on this ground because it does not provide an accurate measure of net tax contributions, truncated, as it is, after seventy-five years.

nominated the system's implicit debt: the amount of additional government resources required to honor the system's accrued liabilities over and above current reserves and excess future tax revenues already committed under current law.

FIGURE SEVEN

Modified Accrual Balance Sheet of Social Security Trust Funds
(Dec. 31, 2002)

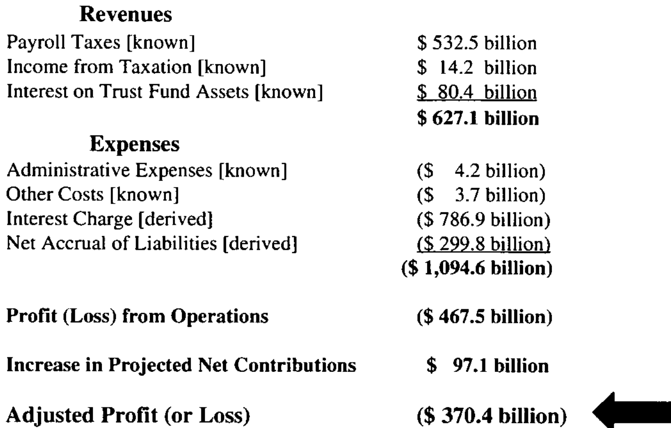
Assets	Liabilities
Value of Reserves = $ 1,378 billion	Accrued Liability to Current Retirees = $4.7 trillion (est.)
Net Future Tax Contributions of Current Participants = $2,119 billion	
Implicit Trust Fund Debt: ($10.5 trillion)	Accrued Liability to Participants still in Workforce = $ 9.3 trillion (est.)
	Total Accrued Liabilities = $14.0 trillion

Figure Seven presents a modified accrual balance sheet for December 31, 2002. The balance sheet's assets consist of the value of current reserves ($1.4 trillion) plus the present value of excess future tax contributions of current participants ($2.1 trillion). The amount by which the system's accrued liabilities ($14.0 trillion) exceed the sum of these assets represents the trust funds' implicit debt as of December 31, 2002, and equals $10.5 trillion. This figure is the same as the closed-group liability of year-end 2002, as only the excess tax contributions of current participants, and not future participants, were included. From this balance sheet, one may readily ascertain the system's unfunded accrued liabilities: the $12.6 trillion difference between accrued liabilities ($14 trillion) and current reserves ($1.4 trillion), which is the deficit featured in a GAAP-style accrual accounting statement. This format would also reveal the amount by which accrued liabilities are expected to be reduced by excess future taxes.

TABLE NINE

Modified Income Statement for Trust Funds (est.)

(Jan. 1, 2002 through Dec. 31, 2002)



Revenues	
Payroll Taxes [known]	$ 532.5 billion
Income from Taxation [known]	$ 14.2 billion
Interest on Trust Fund Assets [known]	$ 80.4 billion
	$ 627.1 billion
Expenses	
Administrative Expenses [known]	($ 4.2 billion)
Other Costs [known]	($ 3.7 billion)
Interest Charge [derived]	($ 786.9 billion)
Net Accrual of Liabilities [derived]	($ 299.8 billion)
	($ 1,094.6 billion)
Profit (Loss) from Operations	**($ 467.5 billion)**
Increase in Projected Net Contributions	**$ 97.1 billion**
Adjusted Profit (or Loss)	**($ 370.4 billion)**

Following similar logic, one can also construct a modified income statement for the trust funds in 2002, as is done in Table Nine. The only difference between this income statement and the GAAP-style income statement presented in Table Six is the inclusion of income items reflecting the increase in projected net tax contributions ($97.1 billion) resulting principally from the addition of a new annual cohort of participants—those turning fifteen in 2002. Even with the recognition of this additional future net tax revenue, the trust funds still showed a $370.4 billion loss in 2002.

One could imagine amplifying these modified financial statements further, adding in the net tax contributions of future participants and aligning the fiscal imbalance with the infinite-horizon open-group measures. For largely pragmatic reasons, that additional step will not be taken here. First, on the practical level, we lack time series for the infinite-horizon open-group measure, and so it is not possible to construct income statements or multiple-year estimates of fiscal imbalance. Also, on a practical level, the Chief Actuary's most recent estimate of the infinite-horizon closed-group liability is the same as the closed-group liability, so there does not seem to be much to be gained from using another measure that will generate roughly the same results.[198] Finally, projections of infinite-horizon open-group liability are inherently more speculative than closed-group estimates, which are themselves more speculative than estimates of unfunded accrued liabilities. While all

[198] *See supra* notes 182–183 and accompanying text.

estimates of long-term solvency entail some degree of speculation and approximation, the infinite-horizon open-group measure is at the extreme.

D. Accrual Accounting and the Unified Federal Budget

Imagine, for a moment, that future trustees' reports of the Social Security system were revised along the lines outlined above, using a modified system of accrual accounting as the principal presentation format. For the foreseeable future, the system would report annual losses on the order of several hundred billion dollars and the implicit debt of the trust funds would annually mount, both in absolute terms and probably as a fraction of current GDP. Both of these facts—annual losses and mounting implicit debt—would present major challenges for budgetary officials when the time came to aggregate the finances of Social Security with other federal budgetary accounts.

Two basic approaches could, however, accomplish this aggregation of accounts, and both would have important implications for budgetary politics. The first is a fully consolidated approach; the second a fully unconsolidated approach. The consolidated and unconsolidated approaches represent more radical alternatives than do the total-budget and on-budget aggregates that are currently used when Social Security cash-flow surpluses are combined with or kept separate from on-budget accounts. They also offer more accurate presentations of the true impact of Social Security on public finances.

There is considerable uncertainty as to how politicians would react to changes in accounting systems. Some argue that reported levels of budgetary aggregates have little effect on politicians' substantive choices.[199] In addition, it is possible that even if trustees' reports adopted a modified system, politicians would still rely on the system's current cash-flow surpluses for discussing federal budgetary aggregates. Nevertheless, it is plausible that changing the public's understanding of the financial status of the Social Security trust funds might substantially impact the manner in which the trust funds are combined with federal budgetary aggregates. This Part's goal is to explore what that impact might be.

1. Full Consolidation Within the Federal Budget

One approach to combining accrual-based Social Security with the broader federal budget would be to establish a system of full consolidation. Under this technique, the profit or loss of the Social Security system each year would be included in the unified budget. Under this approach, Social Security would have put a $370.4 billion drag on the unified budget in 2002 (as

---

[199] For a discussion of the literature on this point, see Kent A. Smetters, *Thinking About Social Security's Trust Fund, in* PROSPECTS FOR SOCIAL SECURITY REFORM 201, 203–05 (Olivia S. Mitchell et al. eds., 1999).

opposed to the $160.3 billion positive effect under current cash-flow accounting).[200] Thus, the consolidated federal deficit for 2002 would have been negative $688.6 billion (rather than negative $157 billion as reported).[201] While this deterioration of more than $500 billion is jarring, it is an accurate reflection of the amount by which the federal government's express promises to public debt holders and implicit unfunded promises to Social Security beneficiaries increased over the course of the year. To give readers a sense of the magnitude of this change, Table Ten presents restated budgetary aggregates for the federal government from 1998 through 2002.

TABLE TEN

Restated Annual Budgets With Fully Consolidated Trust Funds
Five Years: 1998 through 2002
(By Fiscal Year, in Billions of Dollars)

	1998	1999	2000	2001	2002	Total 1998–2002
On-Budget Surplus or Deficit (-)	−30.0	1.9	86.6	−33.4	−317.5	−292.4
Social Security Profit or Loss (-)	−287.1	−549.7	−761.4	−535.2	−370.4	−2,503.9
Postal Service Outlays	0.2	− 1.0	−2.0	−2.3	−0.7	−5.8
Consolidated Surplus or Deficit (−)	−316.9	−548.8	−676.8	−570.9	−688.6	−2,802.1
Consolidated Surplus or Deficit (−) as a percentage of GDP	−3.66%	−6.00%	−6.97%	−5.69%	−6.66%	*n.a.*

Source: Congressional Budget Office, The Budget and Economic Outlook: Fiscal Years 2004-2013, at 148 (Jan. 2003) (Table F-1) (for historical data for on-budget deficits and surpluses and Post Office Outlays).

If one extended the logic of consolidation to the federal government's balance sheet, one might also present the growth of the trust funds' implicit obligations alongside the growth of the government's debt held by the public. Once Social Security's accrued liabilities are recognized on the government's balance sheets, it is logical to include these obligations if one opts for a consolidated presentation of public debt. Such a presentation is offered in Figures Eight and Nine, first with debt figures measured in current dollars

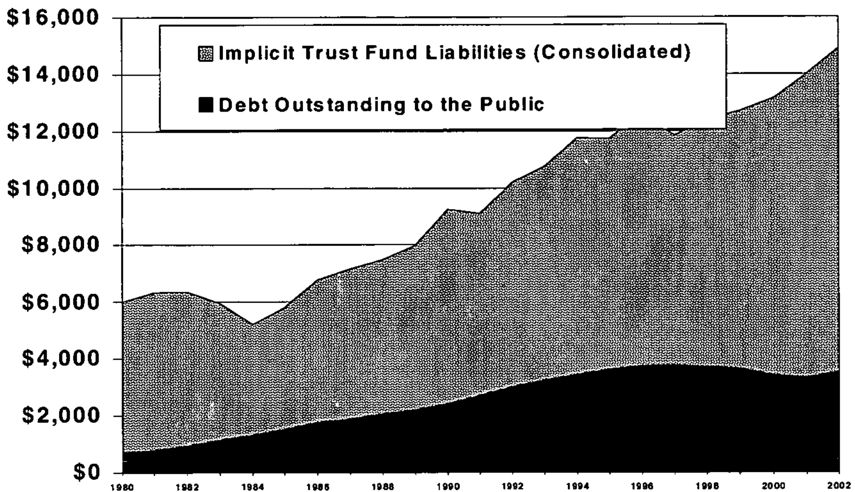
[200] *See* CBO JANUARY 2003 REPORT, *supra* note 74, at 148. In this Part, measures of profit and loss will be based on the modified accrual accounting approach. Were the losses based on a GAAP-style accrual accounting system, without current recognition of future excess tax contributions, losses would be approximately $100 billion higher each year.

[201] *See* Table Ten. If the 2002 Social Security performance under this approach were consolidated with the on-budget performance of the federal government in Fiscal Year 2003, the consolidated deficit would have been roughly $930 billion. *See* Jackson, *supra* note *.

and next as measured as a percentage of the GDP. These figures reveal two important points. First, the magnitude of Social Security's implicit trust fund obligations is much greater than federal debt outstanding to the general public.[202] Second, over the past two decades, the growth in accrued liabilities of Social Security greatly surpassed increases in publicly held debt, and more than offset the small reductions in public debt outstanding from 1998 to 2001, when the total-budget accounts were in surplus.

FIGURE EIGHT

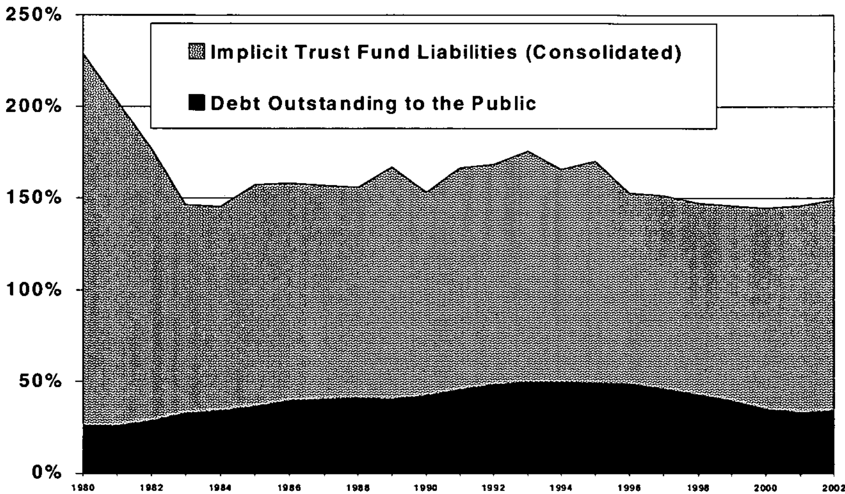
Federal Obligations on a Fully Consolidated Basis:
Public Debt Plus Implicit Trust Fund Liabilities: 1980–2002
(billions of dollars)



[202] Under principles of consolidation, inter-government transactions are netted out. *See generally* BUDGET CONCEPTS REPORT, *supra* note 103. For that reason, trust fund reserves are excluded from the calculation of public debt and are not counted against trust fund accrued liabilities. Thus, the implicit trust fund obligations (consolidated) presented in Figures Eight and Nine are equal to the closed-group liability plus trust fund reserves.

FIGURE NINE

Federal Obligations on a Fully Consolidated Basis:
Public Debt Plus Implicit Trust Fund Liabilities: 1980–2002
(Percent of GDP)



2. *An Unconsolidated Approach to Social Security Accounting*

Politicians, undoubtedly, would find full budgetary consolidation of Social Security, under the terms proposed in the preceding paragraphs, an unpalatable option. Indeed, it could be politically costly to explain to the public why the levels of reported federal deficits are so much higher than previously stated. Some might also object to combining explicit and implicit debt to consolidate aggregates.[203] Accordingly, a more plausible and, in many respects, preferable solution would be to use a fully unconsolidated approach. Treating Social Security as an unconsolidated entity is, in a sense, an extension of the off-budget treatment that the system purportedly, but in fact only episodically, enjoys today.[204]

---

[203] In my opinion, this combination is not obviously inappropriate, as the accrued liabilities presented in Figures Eight and Nine are, by definition, obligations that the federal government is, in all probability, going to honor. One third of these liabilities are owed to retirees, and are therefore politically sacrosanct. The balance are due to mid-career workers, but have been recognized under an accrual formula that reflects the probability that they will be honored. *See supra* Box Four.

[204] *See supra* text accompanying notes 88–101.

Before going further, one must understand how this fully unconsoli-dated treatment would work.

a. Budgetary Aggregates

If Social Security trust funds were treated on a fully unconsolidated basis, the annual operations of the trust funds would not be included in the federal budgetary aggregates. This approach would be similar to the Clinton Administration's efforts in 1999 and 2000 to focus attention on what is conventionally referred to as on-budget figures.[205] There would be less incentive, however, for politicians to revert to using a consolidated approach, because as described above, the effect of consolidation would be to detract substantially from total budget aggregates.

b. Interest Payments and Other General Revenue Contributions

Under an unconsolidated approach, transactions between Social Security and other federal budget accounts would be treated in the same way as transactions between the federal government and unrelated third parties. As a result, all payments to the Social Security trust funds, including interest payments on federal bonds held by Social Security, plus any contributions to Social Security from general revenues, would count as expenses of the federal government in the year they were paid.[206] In contrast to the rhetoric of lockboxes, this approach would clarify the extent to which general revenues were actually used to enhance Social Security's solvency.

c. Measures of Public Debt Outstanding

Under a fully unconsolidated presentation of Social Security finances, bonds held by the trust funds would be included in the calculation of total federal debt outstanding. Not only would this reflect the political reality that the federal government will undoubtedly honor its commitment to redeem these bonds, but the approach also presents a more accurate picture of the federal government's future burden of explicit debt.[207] Figures Ten and

---

[205] *See* ELMENDORF ET AL., *supra* note 95, at 41–44.

[206] An open question is how income taxes currently allocated to the Social Security trust funds should be treated. As described above, the trust funds' revenues include a portion of the income taxes imposed on certain Social Security benefits. Although these payments are not typically characterized as the allocation of general revenues to the trust funds, that is, arguably, what they are.

[207] When Social Security was first included in the unified budget in the late 1960s, the commission recommending the change contemplated that debt issued to the trust funds would be reported in an aggregate measure of gross federal debt. *See* BUDGET CONCEPTS REPORT, supra note 103, at 85. Over time, public accounting has focused nearly exclusively on federal debt held by the general public and has rarely given measures of gross federal debt much prominence. *See, e.g.*, JANUARY 2002 CBO OUTLOOK, *supra* note 98, at xv summary tbl.2.

Eleven provide a graphic presentation of the relative size of these debt holdings over the past twenty years, and reveal the increasingly important share of public debt held by the Social Security trust funds. Figure Ten presents the series in current dollars, whereas Figure Eleven presents the series as a percentage of GDP.

FIGURE TEN

Total Federal Debt on an Unconsolidated Basis:

Public Debt Outstanding (with Bonds held by Trust Funds: 1980–2001)
(billions of dollars)

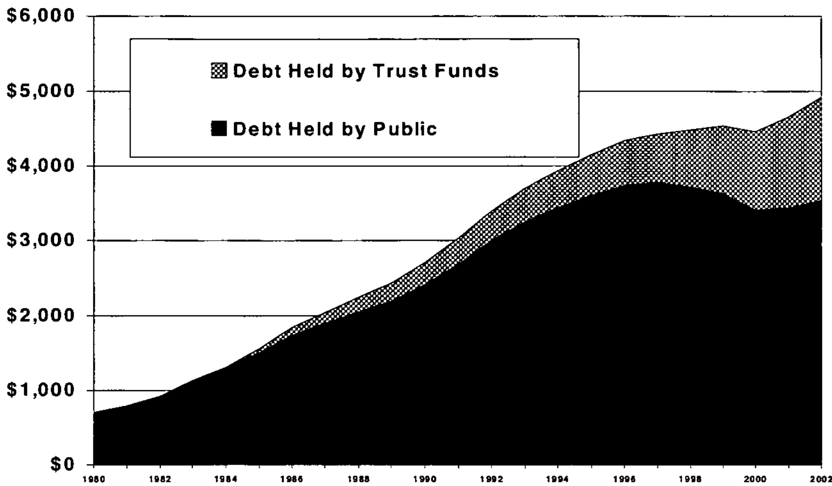
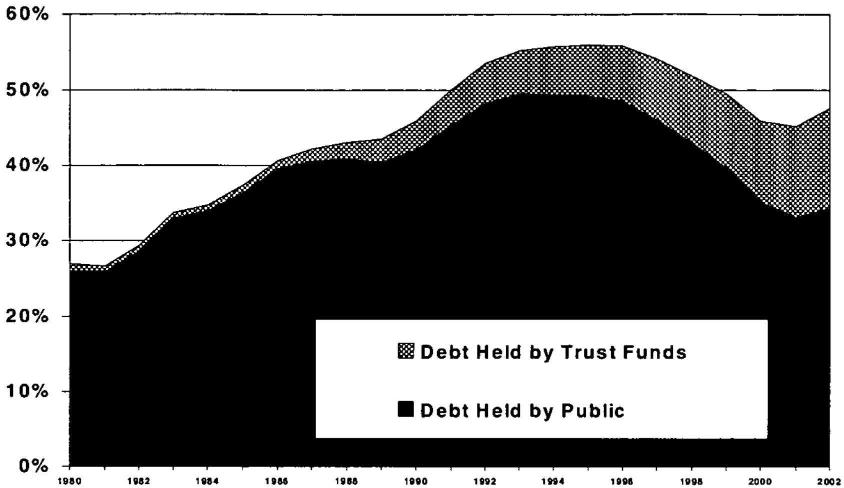


FIGURE ELEVEN

Total Federal Debt on an Unconsolidated Basis:

Public Debt Outstanding (with Bonds held by Trust Funds: 1980–2001)
(percentage of GDP)



d. Separate Financial Statements for Trust Funds

Finally, the Social Security trust funds would maintain their own separate financial statements—both balance sheets and income statements—prepared on the basis of modified accrual accounting. An important element of these financial statements would be the prominent reporting of annual increases in the system's implicit debt. Table Eleven presents the summary statistics that Social Security financial statements, presented on a fully unconsolidated basis, would provide.

TABLE ELEVEN

Key Statistics for Social Security Financial Statements Reported on a Fully Unconsolidated Basis 1998–2002

	1998	1999	2000	2001	2002
Adjusted Profit or Loss ($ billions)	–$ 287.1	–$ 549.8	–$ 761.4	–$ 535.2	–$ 370.4
Adjusted Profit or Loss (% of GDP)	–3.3 %	–6.0 %	–7.8 %	–5.3 %	–3.6 %
Implicit Debt ($ billions)	$ 8,294	$ 8,844	$ 9,605	$ 10,140	$ 10,511
Implicit Debt (% of GDP)	95.8%	96.8 %	98.8 %	101.1 %	101.7 %

III. ACCRUAL ACCOUNTING AND SOCIAL SECURITY REFORM

Beyond presenting a clearer picture of the true financial posture of the Social Security trust funds, modified accrual accounting statements of the sort outlined above could have a profound impact on the public debate over Social Security reform. This Part sketches out the potentially salutary ways in which an accrual-based accounting system might affect public policy in this area. For purposes of this Part, the reader should assume that the modified system of accrual accounting is the dominant format for presenting Social Security finances. If the bleak financial presentations of Figure Twelve, rather than the current trustees' reports, framed public discussions of Social Security finances, what would be the effect on the debate over Social Security reform?

FIGURE TWELVE

TWO BASIC PRESENTATIONS OF SOCIAL SECURITY FINANCES UNDER MODIFIED ACCRUAL ACCOUNTING

FIGURE SEVEN

Modified Accrual Balance Sheet of Social Security Trust Funds
(Dec. 31, 2002)

Assets	Liabilities
Value of Reserves = $ 1,378 billion	Accrued Liability to Current Retirees = $4.7 trillion (est.)
Net Future Tax Contributions of Current Participants = $2,119 billion	
Implicit Trust Fund Debt: ($10.5 trillion)	Accrued Liability to Participants still in Workforce = $ 9.3 trillion (est.)
	Total Accrued Liabilities = $14.0 trillion

TABLE NINE

Modified Income Statement for Trust Funds (est.)
(Jan. 1, 2002 through Dec. 31, 2002)

Revenues	
Payroll Taxes [known]	$ 532.5 billion
Income from Taxation [known]	$ 14.2 billion
Interest on Trust Fund Assets [known]	$ 80.4 billion
	$ 627.1 billion
Expenses	
Administrative Expenses [known]	($ 4.2 billion)
Other Costs [known]	($ 3.7 billion)
Interest Charge [derived]	($ 786.9 billion)
Net Accrual of Liabilities [derived]	($ 299.8 billion)
	($ 1,094.6 billion)
Profit (Loss) from Operations	**($ 467.5 billion)**
Increase in Projected Net Contributions	**$ 97.1 billion**
Adjusted Profit (or Loss)	**($ 370.4 billion)**

A. Clarifying the True Nature of Social Security Finances

1. Debunking the Debilitating Myth of Current Surpluses

Perhaps the most important impact of restating the financial posture of Social Security on the basis of accrual accounting would be to alert the general public to the true financial posture of the program. Public debate would be dramatically different if it were framed by reports of annual losses on the order of several hundred billion dollars a year rather than trustees' reports locating the trust funds' difficulties several decades in the future. In addition, publicizing both the trust funds' mounting accrued liabilities—in excess of $14 trillion—and the $10.5 trillion by which those liabilities exceed both current reserves and the present value of future excess taxes of current participants would likely have a profound effect on public discourse, by stiffening the resolve of politicians to address the system's problems expeditiously and the willingness of the general public to accept at least some modicum of pain in reform proposals.

2. Focusing Public Attention on the Trust Funds' Mounting Implicit Debt

By recognizing the trust funds' current obligations, a modified accrual accounting system would also focus public attention on the implicit debt that the Social Security system is imposing on future generations. Absent changes in benefit formulae or other structural reforms, future generations will have to bear these obligations through higher direct contributions to the Social Security trust funds or general revenue support.

FIGURE THIRTEEN

Implicit Trust Fund Debt & Public Debt: Current Dollars
(1980 to 2002 Actual; 2003–2012 Projected; billions of dollars)

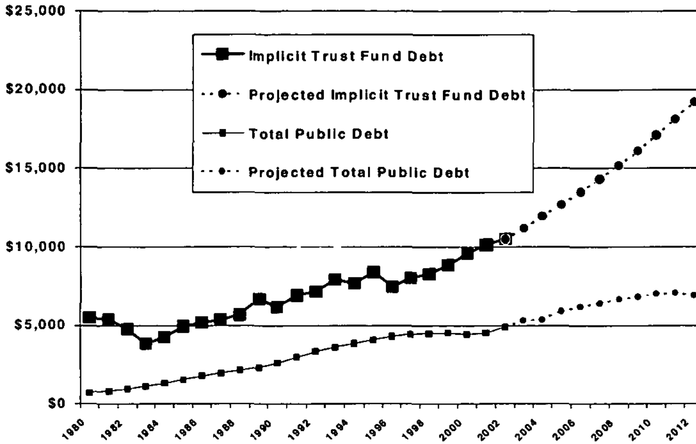
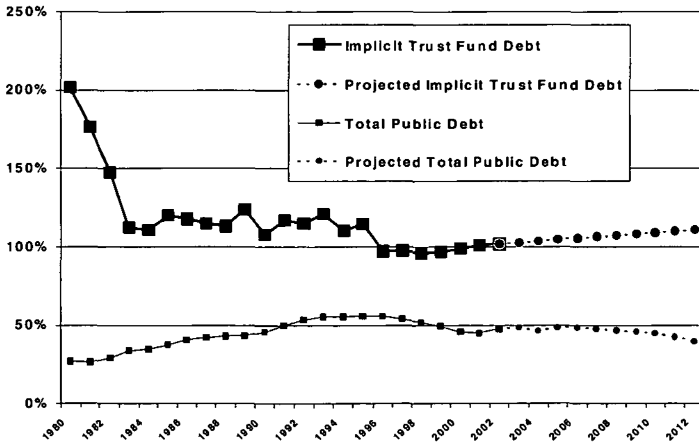


FIGURE FOURTEEN

Implicit Trust Fund Debt & Public Debt: Percent of GDP
(1980 to 2002 Actual; 2003–2012 Projected)



For example, revised trustees' reports under this new regime would presumably include projections of the future trends in implicit trust fund debt.

Figures Thirteen and Fourteen project a plausible path of growth of implicit trust fund debt in the next decade, both in absolute terms and as a percentage of GDP.[208] Also included in these figures are levels of total public debt—that is, debt held by the general public plus debt held by the Social Security trust funds. Once Social Security's finances are restated in accordance with modified principles of accrual accounting, such comparisons between implicit trust fund debt and explicit public debt will be both inevitable and useful, as the general public will quickly learn that mounting financial commitments of Social Security are both larger and growing faster than the public debt, even once restated to include debt held by the trust funds.

3. *Developing Appropriate Targets for the Implicit Debt of Social Security*

A further advantage of highlighting the size of Social Security's implicit debt and that debt's relationship to explicit federal debt is the possibility that it would prompt a sensible national debate over the appropriate level of these obligations. While economists often speak in terms of acceptable level of explicit federal debt relative to the size of the economy,[209] similar discussions about the appropriate levels of other kinds of governmental obligations, particularly public pension obligations—at least in the United States—are rare.[210] Once the implicit debt of the Social Security system is quantified through accrual accounting, one might reasonably expect such discussions to begin. While this Article will not engage in such speculation,[211] it is useful to imagine how discussions of Social Security reform might proceed were such a consensus to emerge. Suppose, for example, that economists generally agreed that the implicit debt associated with Social Security should be limited to no more than 50% of GDP. Suppose further that a bipartisan consensus were reached that Social Security reform legislation should attempt to get the system's implicit debt to at least 75% of GDP over the next ten years—roughly halfway to the 50% target from the current level of Social Security implicit debt, which is slightly more than 100% of

---

[208] For purposes of this projection, the critical unknown is the future rate of growth of implicit trust fund debt. These figures are based on the assumption that the debt will exceed GDP growth by the same percentage it has exceeded GDP growth (seventeen percent) over the past five years. In earlier periods this relationship does not always hold, but the difference is difficult to interpret because of changes in actuarial assumptions and, in some years, program design. *See* CONG. BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: FISCAL YEARS 2004–2013, at xvi, 43 (2003) (providing rates of GDP and public debt growth).

[209] *See, e.g.*, CBO AUGUST 2003 UPDATE, *supra* note 100, at 2 tbl.1-1 (reporting debt held by the public as a percentage of GDP).

[210] As discussed below, economists do, however, routinely estimate and criticize the unfunded pension obligations of other countries. *See infra* Part IV.B.1.

[211] Acceptable levels of implicit debt for Social Security would likely depend on a number of factors, including the size of other entitlement programs (most notably Medicare and other health benefits for the elderly), projected growth rates of the economy and population, and a host of other technical factors.

the GDP. Such an agreement would provide a baseline for evaluating various Social Security reform proposals.

FIGURE FIFTEEN

Setting a Path for Reducing Social Security's Implicit Debt
1979 to 2002 Actual; 2003–2012 Projected
(billions of current dollars)

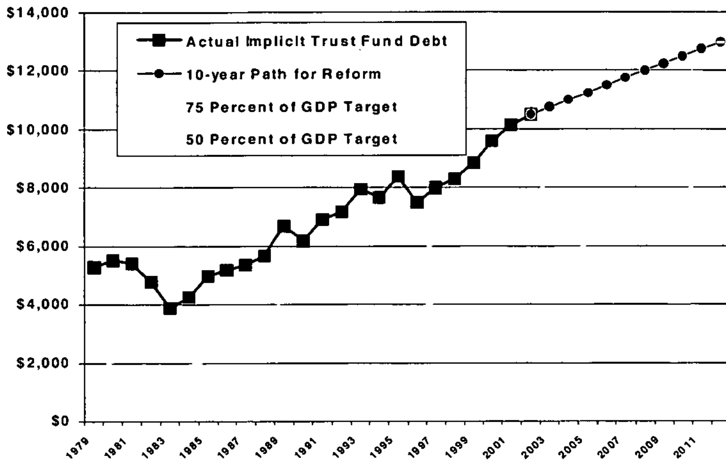


Figure Fifteen illustrates this hypothetical path for reform. The line with circles reflects a ten-year path for reform necessary to achieve this goal of reducing implicit trust fund debt to 75% of the GDP at the end of ten years. If this framework for reform were adopted, one of the criteria for evaluating reform proposals would be to assess whether they reduced the system's implicit debt within the timeframe specified. Essentially, one could use figures similar to Figure Fifteen to map the projected financial impact of competing reform proposals.

B. Enhancing the Quality of Public Debate over Reform Proposals

As compared with the seventy-five-year actuarial deficit (negative 1.92% of payroll) that currently frames most reform discussions, modified accrual accounting statements and a target path for reducing implicit trust fund debt as a percentage of GDP would offer a far preferable framework for public discussions.

A modified system of accrual accounting would make considerably more transparent how various elements of reform proposals would affect the implicit trust fund debt. Consider the five basic ways in which the financial condition of Social Security might be improved. (1) Reductions in benefits to be accrued in the future would appear initially as increases in the "quasi asset"

representing excess tax contributions on the trust fund balance sheet, and later as smaller increases in the system's accrued liabilities as those benefits were accrued. (2) Higher payroll taxes to be imposed in the future would also appear first as increases in the "quasi asset" representing excess future tax contributions and later as greater trust fund reserves once those taxes were paid into the system. (3) Direct contributions of general revenues would be recognized in the year in which the contributions were made, presumably as increases in trust fund reserves. (4) Similarly, higher returns from trust fund reserves—through stock market or other investments—would be recognized in the year those returns were earned. (5) To the extent that reform plans included reductions in benefits accrued to date for retirees or mid-career workers, those reductions would be immediately recognized as reductions in the system's accrued liabilities. Each of these five mechanisms for restoring Social Security solvency would have a clear and distinctive impact on accrual-based balance sheets and income statements.

1. Exposing Short-Range Cash-Flow Effects

One certain advantage of accrual accounting is that it would expose the limitations of reform proposals that rely principally on short-range cash-flow effects. For example, bringing new participants into the Social Security system always has a positive cash-flow effect in the short term because the new participants initially pay large amounts of taxes and receive few benefits. On a modified accrual-accounting basis, and in terms of economic reality, the gains from such proposals are typically much smaller and may in fact be negative, depending on the new entrants and the kinds of benefits they will enjoy. For example, pools of new low-income workers can detract from the system's solvency in the long term, even though their participation might have a positive cash-flow effect in the short term. Whatever the long-term impact, under accrual accounting, there would be no short-term advantage in terms of either trust fund finances or federal aggregates to pursuing such strategies.

2. Recognizing the Value of Reforms with Future Cash-Flow Effects

Conversely, some reform proposals have a positive economic effect on the trust funds' solvency, but their benefits are understated in the current accounting system because their cash-flow effects occur in the future, even beyond the seventy-five-year long-term perspective. A good example of this phenomenon can be found in the proposals of the President's Commission on Social Security Reform. A principal feature of the Commission's proposals was the imposition of an offset in traditional Social Security benefits for contributions made to individual accounts.[212] Under accrual accounting, such offsets would reduce the trust funds' accrued liabilities in the year of the

[212] *See* BUSH COMMISSION REPORT, *supra* note 2, at 90.

contributions, whereas under traditional Social Security accounting, they would not be recognized until partially beyond the seventy-five-year projection period.

3. *Flagging Reform Proposals That Expand Accrued Liabilities*

A further advantage of accrual accounting is its ability to flag reform proposals that include retroactive benefit enhancements.[213] Though there is much to recommend these proposals on the merits and though the ultimate solution to the problems of Social Security will almost certainly include a combination of sticks and carrots, the current accounting system makes it too easy to slip in expensive improvements because the yardstick for evaluating reform proposals is their long-range impact on cash-flow solvency. An accrual accounting system would highlight such proposals—at least when applied on a retroactive basis—as immediate increases in the system's accrued liabilities. An advantage of accrual accounting is that it would reflect the true costs of such reforms and facilitate a more informed debate on their merits.

4. *Eliminating the Cliff Effect*

A further advantage of accrual accounting is that it would eliminate the cliff effect, which plagues reform proposals evaluated solely on the basis of the actuarial deficit. The cliff effect arises because the trust funds' tax base is much lower than projected expenditures at the end of the seventy-five-year projection period.[214] A reform proposal can eliminate the seventy-five-year actuarial deficit for the current measurement period, only to have insolvency return to the system a few years later, as the period of analysis shifts forward to include additional years of cash-flow imbalance. Proposals that bring the Social Security system's implicit debt to a sustainable ratio to the GDP will be much more likely to achieve long-range balance than will proposals that eliminate only the seventy-five-year actuarial deficit.

5. *Identifying Back-loaded Reform Proposals*

Another way in which accrual accounting would improve the quality of debate over reform proposals would be by unmasking reform proposals that are substantially back-loaded. A good case in point is one of the reform options that the 1996 Advisory Council proposed. While the proposal purported to eliminate the long-range actuarial deficit (then estimated at 2.17% of payroll), the elimination was accomplished in part through a new tax to be imposed five decades in the future.[215] If this proposal were evaluated through

---

[213] Examples include improved benefits for elderly widows and various efforts to increase the level of benefits for participants who have had lower wages.

[214] *See supra* note 75 and accompanying text.

[215] *See* 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 90.

the lens of accrual accounting and pro forma income statements for the near term, it would be readily apparent that these future taxes did nothing to reduce the system's accrued liabilities. The only impact would be on the amount of the system's "quasi-asset" for excess future tax contributions and the amount of the impact would likely be smaller than the Advisory Council estimated, as the impact of the increased taxes could be offset to some degree by increases in benefits to be accrued in the future.

6. *Flagging the Amount of Excess Future Taxes*

A final advantage of the modified system of accrual accounting is its capacity to demonstrate the extent to which the system's accrued liabilities are being funded with excess taxes to be charged to participants in the future. While there is nothing inherently wrong with levying on participants' taxes that exceed the value of benefits to be accrued in the future, there is presumably some limit to the extent to which a social insurance system should rely on such excess taxes, lest public support for the system be undermined in the future. A reform proposal that relied on higher payroll taxes in the future might substantially reduce the Social Security system's implicit debt, but it might be a proposal with significant long-term political weakness. A modified accrual accounting system would expose concerns of this sort.[216]

C. *Altering the Optics of Reform Proposals*

In addition to the direct benefits of accrual accounting described above, this mode of analysis could bring a number of additional—albeit somewhat more subjective—benefits to reform debates.

1. *Distinguishing Accrued Rights of the Elderly and Other Participants*

An essential element of accrual accounting would be the recognition of the benefits that participants have accrued to date. These liabilities would be recognized on an annual basis on the system's income statement and reflected on the system's balance sheet at the end of each year. In a sense, accrued benefits are privileged over benefits that accrue in the future.[217] If accrued benefits were presented in this way, long-term participants in the system—particularly retirees and near retirees—might be persuaded that their interests

---

[216] This again raises an interesting question of public finance: the extent to which accrued liabilities of social insurance programs should be funded by excess payroll tax contributions as opposed to other potential sources of support, such as general revenues. For an interesting discussion of this issue, see generally Peter A. Diamond, Social Security, The Government Budget, and National Savings (2003) (unpublished manuscript, on file with author).

[217] For private pension plans, federal statutes prohibit the elimination of all accrued interest in pensions and thus also privilege accrued benefits over benefits that will accrue in the future. *See supra* text accompanying note 188.

would not be adversely affected by reform proposals. One could imagine policy analysts or politicians using Social Security balance sheets as a tool for persuading nervous constituencies that their expectations for retirement income will not be compromised. In other words, analysts would be able to distinguish between the fully accrued rights of retirees and the only partially accrued rights of other participants. The confusion of the interests of these two groups often complicates current discussion of reform proposals. Accrual accounting offers a possible path out of this quagmire.

2. *Differentiating Prospective and Retroactive Effects*

A related point is the capacity of accrual accounting to offer a new way to differentiate between the retrospective and prospective effects of reform proposals. Given the current system, changes usually apply to both previously accrued and to-be-accrued benefits. For example, the 1983 change in retirement age applied to covered workers both for benefits accrued before 1983 and for those that would accrue thereafter.[218] Similarly, enhancements in benefit provisions, such as linkage of benefit increases to productivity improvements in the 1970s, also apply both retroactively and prospectively.[219] One of the advantages of accrual accounting is that it offers reformers more options in the way they structure changes to the system, and having more options may be necessary to develop a politically viable reform package.

a. *The Possibility of Prospective Benefit Cuts*

Consider if benefit cuts were imposed, not on an all-or-nothing basis, but rather only on benefits that would accrue in the future. For example, imagine that, rather than switching over fully to cost-of-living indexation of initial benefit levels, a reform proposal were structured to preserve the existing productivity-adjusted formulas for accrued benefits and establish cost-of-living indexing only for benefits that accrue in the future. Such a reform package could be characterized as honoring existing commitments and expectations while still offering immediate improvements in the system's financial posture through the reduction of the rate of accrual of benefits in all future years. This option is both preferable to and more equitable than the current practice of imposing benefit cuts with effective dates that are postponed for years but then imposed on a fully retroactive basis.

---

[218] *See* Robert L. Clark, *Liabilities, Debts, Revenues, and Expenditures: Accounting for the Actuarial Balance of Social Security*, 41 HARV. J. ON LEGIS. 161 (2004). Although the change did not go into effect until many years after 1983, once it went into effect for a particular age cohort, the change applied to all of that cohort's benefits.

[219] *See* PRESIDENT'S PROPOSAL, *supra* note 69, at 2–4.

b. More Modest Methods of Enhancing Benefits

A similar point could be made about benefit enhancements. Accrual accounting would invite benefit improvements to be imposed only with respect to benefits accrued in the future. This option would allow politicians a means to address the significant weaknesses in the Social Security system without imposing on it the substantial costs of fully retroactive application.

3. Changing Expectations of the General Public

The general point that runs through all of these examples is that accrual accounting offers a way to change the expectations of the general public in a subtle, but important, way. Through a complicated combination of factors, American workers have been encouraged to think of themselves as being entitled to the level of retirement benefits authorized under current law, even though their retirement may not occur for many years or even many decades.[220] The implicit understanding seems to be that, as long as participants make their payroll-tax contributions as currently structured, they are entitled to receive their benefits as scheduled. The current system of Social Security accounting reinforces this understanding, because it is designed to measure the balance between projected benefits and projected taxes. This linkage is even more apparent in the personal statements that the Social Security Administration sends out to all participants, estimating their projected level of future benefits on the assumption that they maintain their current level of employment until retirement. Lobbyists and politicians regularly encourage this understanding by opposing any change in promised levels of benefits. Accrual accounting offers a different perspective on Social Security benefits. It privileges a subset of benefits—those that have already accrued—and invites participants to ratchet downward their sense of entitlement to benefits that have not yet accrued.

D. Accrual Accounting and Individual Accounts

The reforms proposed in this Article could also affect the policy debate surrounding the creation of individual accounts in a number of ways, although the overall direction of the effects is ambiguous.

1. Posing the Problem of Transition Costs

Perhaps the most significant effect would be to make the magnitude of unfunded accrued liabilities of the trust funds more prominent. Some proponents of individual accounts emphasize the higher returns that participants could realize on these accounts without addressing the question of how the un-

---

[220] *See supra* text accompanying notes 63–65.

funded liabilities of the current system would be addressed. Indeed, one of the problems with the public debate over individual accounts is the difficulty of comparing the rates of return under traditional Social Security, which address a portion of the system's unfunded accrued liabilities, and the rates of return on individual account proposals, which leave the question of unfunded transition costs unresolved. If individual account plans were vetted through the screen of accrual accounting, the issue of accrued unfunded liabilities would be unavoidable.

2. *Eliminating Budgetary Anomalies*

Accrual accounting would also eliminate a number of budgetary accounting anomalies that can cast individual account proposals and certain other reforms in an unfavorable light. Under current accounting rules, funds transferred to individual accounts are treated as expenditures, as are investments of trust fund assets in any sort of financial asset other than government bonds. These conventions mean that the creation of individual accounts and proposals to invest trust fund assets in the stock market all give rise to the appearance of budgetary costs. Given current budgetary imperatives, these anomalies make it difficult for politicians to adopt such reforms, regardless of their substantive merit. If the Social Security trust funds were accounted for on the fully unconsolidated basis outlined above, anomalies of this sort would disappear. Neither the creation of individual accounts nor the transfer of trust fund assets into the stock market would have a budgetary effect.

3. *Comparing Traditional Benefits and Individual Accounts*

A final benefit of accrual accounting would be the enhancement of the comparability of traditional Social Security retirement benefits and individual accounts. A problem in the current debate over individual accounts is the difficulty that both experts and the general public face when comparing the relative merits of these two benefit structures. Individual accounts are usually valued in terms of the current amount of assets in a particular individual's account. Traditional benefits, by contrast, are typically described as future payments, often made on a monthly basis, beginning at a point in the future, for example, the participant's sixty-fifth birthday, and continuing for the remainder of the combined lives of the participant and the participant's spouse with various adjustments for inflation as well as spousal and survivors' benefits. To a large degree, proponents of individual accounts are asking the public to accept the first kind of benefit as a substitute for the second. Comparing these two is extraordinarily difficult. If we began to evaluate Social Security benefits on an accrual basis, however, restating the benefits of individuals as the net present value of expected benefits for each individual would be a fairly simple step. The widespread disclosure of such individual restatements—perhaps as part of the annual statements sent to each participant—could

greatly improve the public's understanding of traditional Social Security benefits and strengthen public debate over the relative merits of individual accounts.

E. Social Security Reform, the Federal Fisc, and National Savings

Adopting a system of accrual accounting for Social Security and implementing this system on a fully unconsolidated basis could also have important benefits for federal budgetary politics and, by implication, for national savings. While the gains to be had in this area are highly speculative, there are good reasons to believe that the impact would be positive.

1. Budgetary Effects of Fully Unconsolidated Treatment

An initial impact of moving the Social Security trust funds to a fully unconsolidated basis would be to shift the attention of politicians and analysts to what is now referred to as the on-budget surplus or deficit. Unlike the current budgetary treatment, which encourages politicians to resort to total-budget aggregates, eliding distinctions between on-budget and off-budget accounts, this Article's proposed reforms would discourage the combination of annual trust fund results, because on an accrual-accounting basis the trust funds experience substantial annual losses. By keeping national attention on on-budget aggregates, this proposal would increase the likelihood that the on-budget accounts remain in or near balance.[221] In effect, this would mean that cash-flow surpluses on Social Security trust funds would be used to decrease the amount of public debt held by the general public and thereby increase national savings.

2. Budgetary Effects of General Revenue Contributions

This Article's proposals would also untangle the budgetary confusions that surround proposals to allocate additional general revenues to the Social Security system. As explained above, one of the great confusions of the lockbox debates of 1999 and 2000 was the question of how to characterize and account for general revenues committed to Social Security.[222] Under this Article's proposals, their accounting would be quite simple. Any contribution to the trust funds would be treated as an on-budget expense in the year the contribution was made. Were on-budget accounts otherwise in balance for that year, the contribution would move the on-budget accounts into deficit. If Congress wished to make such a contribution without borrowing additional

---

[221] There is considerable uncertainty whether changes in budgetary targets actually affect budgetary decisions, and simply focusing attention on on-budget aggregates does not mean that the political branches will keep these budgetary accounts in balance. *See supra* Part III.A.3. Moving from total budgetary aggregates to on-budget aggregates, however, should tend to reduce, if not eliminate, federal deficits.

[222] *See supra* note 113 and accompanying text.

funds from the general public, new general revenues would need to be raised or other expenses reduced. This treatment would, in my view, impose appropriate discipline on the federal budget process and discourage the federal government from honoring its commitments to Social Security simply through the issuance of new debt to the general public.

3. Tracking Gross Public Debt and Unfunded Trust Fund Liabilities

A final advantage of this Article's proposal is that it would encourage the presentation of a more accurate picture of the overall size of federal obligations. The approach would highlight the gross amount of public debt—that is, the combination of debt held by the general public and debt held by the trust funds. This figure would be a more accurate representation of the debt burden that taxpayers will have to shoulder in the future and therefore the one that politicians and analysts should monitor. Of course, accrual accounting would also highlight the implicit trust fund debt of the Social Security system, another form of public obligation passed on to future generations and one with significant implications for overall national savings. Bringing these combined measures of public obligation into the limelight would further enhance public debate on the issues.

F. Substantive Implications of Accrual Accounting

Some readers will no doubt be curious about the substantive implications of restating Social Security finances on the basis of accrual accounting. The impact here is ambiguous. Consider, for example, the hot-button issue of privatization. Defenders of traditional Social Security benefits recoil from estimates of the system's accrued liabilities out of fear that reports of unfunded obligations in excess of $10 trillion would weaken political support for the system. On the other hand, supporters of individual accounts resist references to accrued liabilities because they would highlight the transition costs of moving to a fully privatized Social Security system. It is not clear which of these groups would fare better under an accrual-accounting system, but the general public would be in a better position to evaluate the merits of the traditional system as compared to those of a system of privatized accounts if the comparison were made in terms of accrual accounting.

Another important substantive dimension of Social Security is the system's role in redistributing wealth from rich to poor and from young to old. Some fear that accrual accounting would more clearly expose Social Security's redistributive elements.[223] Conceivably, such information could diminish public support for Social Security and its redistributive role. Recent studies,

[223] This problem would become more acute if participant benefits were also reported on an accrual-accounting basis. Each participant would then see the net expected tax (or net expected subsidy) of his or her annual participation in Social Security.

however, suggest that many younger workers believe that they will never receive any Social Security benefits.[224] Presumably, these participants think that they get no benefit from their annual Social Security contributions. A fully implemented system of accrual accounting would reveal to these individuals that their annual contributions are generating real expected benefits in terms of retirement security, as well as life insurance and disability benefits. Perhaps for these individuals, accrual accounting would increase support for the system.

Finally, objections to accrual accounting on the basis of substantive concerns are, in the end, deeply undemocratic. Such objections reduce to arguments that we should not present the finances of the Social Security system in the most realistic manner because the general public would react badly. The moral weakness of such claims aside, intentional obfuscation of government finances cannot be the right way to build sustained public support for such an important social insurance program.

IV. Postscript on Normative Baselines

The accrual accounting proposal for Social Security outlined in this Article generates surprisingly impassioned responses in some circles.[225] The source of this passion is not always clear. Sometimes, the criticism seems to be based on an intuition that cash-flow accounting is the only appropriate method for recording government programs.[226] Other times, defenders of the status quo stress that Social Security is a social insurance program and assert that this characterization constitutes a complete explanation of current accounting practices.[227] This postscript responds to these reactions, and explains why this Article's recommendations are consistent with recent trends in government accounting, including recent reforms in federal accounting standards for social insurance programs.

---

[224] *See* Jacobs & Shapiro, *supra* note 64, at 355–56.

[225] *Compare* Pozen, *supra* note 16, *and* Elizabeth Garrett, *Accounting for the Federal Budget and Its Reform*, 41 Harv. J. on Legis. 187 (2004) *with* Clark, *supra* note 218.

[226] *See generally* Budget Concepts Report, *supra* note 103.

[227] *See* Research and Dev. Arrangements, Statement of Financial Accounting Standards No. 17, § 26 (Financial Accounting Standards Bd., 1999). Others argue that the pay-as-you-go nature of the Social Security system justifies the system's current accounting treatment. *See id.* at § 28. This claim is not particularly strong. First of all, the Social Security system is no longer operating on a purely pay-as-you-go basis. The combined trust funds are now pre-funded to the tune of $1.4 trillion and the level of pre-funding will continue to grow for a number of years. *See supra* Figure Three. More important, the central problem with Social Security finances is that, even with this substantial amount of pre-funding, the system is promising benefits that will not be sustainable from projected revenues in the future. Given demographic trends, the pay-as-you-go aspect of Social Security financing is more a cause of the system's problems than a justification of its current accounting practices.

A separate, but equally insubstantial, objection is that accrual accounting is appropriate only for fully funded public pension systems. Nothing about accrual accounting necessitates any particular level of pre-funding. *See* Box Five. Indeed, Social Security would likely remain substantially underfunded on an accrual-accounting basis. This Article's argument is that the level of under-funding should be kept apparent.

BOX FIVE

THE NORMATIVE IMPLICATIONS OF ACCRUAL ACCOUNTING: FUNDING AND PROFITABILITY

Once the financial statements of Social Security are restated in terms of accrual accounting, one might reasonably inquire whether this accounting format has normative implications for the system's funding or profitability. In particular, does accrual accounting imply that Social Security should be fully funded—that is, that the system's accrued liabilities should not exceed its assets—or that its annual income statement should ordinarily show a profit or at least the absence of substantial losses? The answer to both of these questions is no.

Although this is principally a question of public finance, there is no reason to suppose that a public pension system needs necessarily to be fully pre-funded or even that its accrued liabilities need not exceed the sum of current reserves plus the net present value of excess taxes for current participants—that is, that the system's implicit debt be zero. What is important, however, is that the system's level of implicit debt not be allowed to grow in an unbounded manner in comparison to overall growth in the economy. Ideally, that level of this implicit debt should stay within some target ratio to GDP. Perhaps the current level of implicit Social Security debt, equal to slightly more than 100% of GDP is appropriate, although I expect that most experts would prefer a lower target, such 75% or 50% of GDP, targets illustrated in Figure Fifteen.

The appropriate annual level of profit or loss for a Social Security system following principles of accrual accounting follows from the target level of implicit debt for the system. If the system were at the target level, then it would be acceptable for the system to report annual losses as long as those losses did not cause the system's implicit debt to grow faster than the overall economy. If, on the other hand, the system's implicit debt were greater than target levels, better performance on annual income statements would be appropriate. Even under these conditions, however, annual profitability would not be required as the implicit debt to GDP ratio of the trust funds will decline as long as the system's implicit debt grows more slowly than the overall economy. For example, the hypothetical reform path illustrated in Figure Fifteen implies adjusted annual losses for the system on the order of $200 billion over the next decade. Even with this level of annual loss, the system's implicit debt will gradually decline as a percentage of GDP over the coming decade.

It then presents several additional normative perspectives that support the claim that accrual accounting is the most nearly accurate way to present the finances of the Social Security system. First, it reviews how the problem of unfunded public pension plans is generally discussed in

the international context. When multinational organizations, such as the International Monetary Fund (IMF), and other analysts examine the financial posture of public pension programs in other countries, they generally speak in terms of accrued liabilities as well as the sort of long-term cash-flow projections that dominate the trustees' reports.[228] While various measures of accrued liabilities are employed, accrual accounting is the principal metric against which the solvency of public pension schemes is tested. Second, this Part offers a brief sketch of how public policy analysts value Social Security benefits in a variety of other academic settings. When trying to assess the economic impact of Social Security, economists regularly use a form of accrual accounting to estimate the value of benefits.[229] If these experts believe that accrual accounting is the best way to estimate the value of participants' benefits in Social Security, then they should see that accrual accounting is also the best way to estimate the liabilities associated with the obligations of the trust funds to pay those benefits.[230]

A. Accrual Accounting and the Federal Government

Although cash-flow accounting is the norm in the federal budget, the notion that elements of accrual accounting might be grafted onto federal accounting standards is not new,[231] and examples of accrual accounting in current federal budgeting and accounting standards are increasingly common.

1. Accrual Accounting in General

One prominent illustration of accrual accounting in the federal budget is the Federal Credit Reform Act of 1990 (FCRA),[232] which established accrual accounting for a wide range of federal programs, including loan and credit-guarantee programs.[233] With the passage of the FCRA, Congress recognized

---

[228] *See infra* Part IV.B.1.

[229] *See infra* Part IV.B.2.

[230] For a description of the accounting treatment of defined-benefit pension plans, see Howell Jackson, A Comparison of Social Security Benefits and Private Pension Plans (Sept. 3, 2002) (unpublished manuscript, on file with author). As explained in that paper, the structure of private pension benefits is more similar to that of traditional Social Security benefits than is often appreciated, and the accounting rules for private pension plans offer a good model for fleshing out this Article's accrual accounting proposal.

[231] For example, the 1967 presidential commission on budget concepts—the same group that initially recommended that Social Security be brought on-budget—also advocated introducing accrual accounting techniques to certain governmental operations. *See* BUDGET CONCEPTS REPORT, *supra* note 103, at ch. 4.

[232] 2 U.S.C. § 661c (2000).

[233] For an overview of the FCRA, see ACCOUNTING & INFO. MGMT. SUBDIV., GEN. ACCOUNTING OFFICE, CREDIT REFORM: GREAT EFFORT NEEDED TO OVERCOME PERSISTENT COST ESTIMATION PROBLEMS 1–5 (1998); *see also* Michael R. Pompeo, *Accrual Accounting for Federal Credit Programs: An Evaluation of the Federal Credit Reform Act of 1990*, 66 TAX NOTES, 257, 257–58 (1995).

that, for some kinds of government programs, cash-flow accounting offers a misleading picture of true government costs. For example, with a loan program, cash outflows in the year the loans are made tend to overstate the cost to the federal government because many loans will be repaid in future years. Conversely, guarantee programs may have no cash outflows in the year a guarantee is made, or even positive cash flows, if a guarantee fee is charged, even though the issuance of a guarantee can represent a significant liability for the government because payments may have to be made in the future when the guarantee comes due. The FCRA requires the government to recognize the expected cost of government credit programs in the year in which the obligations are incurred.[234] Thus, the FCRA mandates accrual, as opposed to cash-flow, accounting for an important segment of the federal budget.

Federal insurance programs, such as deposit insurance or flood insurance, are expressly exempted from the coverage of the FCRA.[235] Nevertheless, the accounting challenges of public insurance programs are quite similar to those of credit programs. When underwriting insurance, the government receives payments in the current period in exchange for a commitment to shoulder costs in the future. Accrual accounting is a natural way to account for these obligations as they arise. As a result, over the past fourteen years, a number of government studies have called for the extension of accrual accounting concepts to this area. For example, in a 1997 report, the GAO explored the extension of accrual accounting treatments to federal insurance programs, such as federal deposit insurance and other insurance programs run by the federal government, but not to Social Security.[236] The GAO generally endorsed such an expansion, with the caveats that, in many areas, the development of accrual accounting systems would be complex and that, as an initial matter, supplemental reporting of risk estimates should be undertaken.[237] Representatives of the GAO recently reiterated their support for this expansion of accrual accounting,[238] and a proposal to account for the retirement benefits of federal employees on an accrual accounting basis is currently being debated in Washington.[239]

Accrual accounting is not inherently inappropriate for government programs.[240]

---

[234] *See* 2 U.S.C. §§ 661a–661f (2000).

[235] *Id.* at § 661e (2000).

[236] *See* BUDGETING FOR FEDERAL INSURANCE PROGRAMS, *supra* note 73, at 4–7.

[237] *Id.*

[238] *See* GEN. ACCOUNTING OFFICE, BUDGET PROCESS: EXTENDING BUDGET CONTROLS (2002) (testimony of Susan J. Irving).

[239] *See generally* CONG. BUDGET OFFICE, THE PRESIDENT'S PROPOSAL TO ACCRUE RETIREMENT COSTS FOR FEDERAL EMPLOYEES (June 2002).

[240] The federal government's adoption of accrual accounting methods is part of a broader trend toward accrual accounting in governments around the world. *See generally* Int'l Fed'n of Accountants Pub. Sector Comm., *Guidelines for Government Financial Reporting* (July 1998). As discussed in the IFAC report, accrual accounting is an increasingly prominent alternative form of presentation for public accounting systems with numerous advantages over traditional

2. Accrual Accounting for Social Insurance

As a social insurance program, Social Security has typically been excluded from past proposals for accrual accounting within the federal government. Government accountants, however, have addressed accounting statements for social insurance in a recent initiative. This initiative came from the Federal Accounting Standards Advisory Board (FASAB)—the board responsible for developing generally accepted accounting practices for federal entities.[241] In the mid-1990s, FASAB began an extensive review of the appropriate accounting treatment of social insurance, paying particular attention to Social Security, the federal government's largest social insurance program. The process included a series of exposure drafts and public comments, culminating in the August 1999 Statement on Social Insurance,[242] which established important new standards of disclosure for social insurance programs, including Social Security.

Although little known outside of government accounting circles, the FASAB Statement on Social Insurance represents an extensive and sophisticated consideration of the special problems of accounting for social insurance. While some participants in the FASAB process opposed any use of accrual accounting in financial statements for social insurance programs,[243] others argued for reforms that are quite similar in spirit to, albeit less extensive than, this Article's proposal.[244] The arguments advanced in favor of recognizing some form of accrued liability on the balance sheets of Social Security were quite similar to the ones made in this Article,[245] and they stressed the fact that existing financial statements are "inherently misleading" because they "fail to quantify the size of the promise that is continuously being made and on which people are being told they can rely."[246] Both to be consistent with generally accepted accounting standards for the private sector

---

cash-flow accounting. Over the past decade or two, accounting reforms in New Zealand, the United Kingdom, and Australia have incorporated elements of accrual accounting into their governmental financial statements, some of which extend to public pension systems. *See* GEN. ACCOUNTING OFFICE, ACCRUAL BUDGETING: EXPERIENCE OF OTHER NATIONS AND IMPLICATIONS FOR THE UNITED STATES 10 (2000) [hereinafter GAO ACCRUAL BUDGETING REPORT] (concluding that "for some activities, such as credit and pension programs, cash-based measurement is incomplete and potentially misleading").

[241] *See* Federal Accounting Standards Advisory Bd., Memorandum of Understanding among the General Accounting Office, the Department of the Treasury, and the Office of Management and Budget on Federal Government Accounting Standards and a Federal Accounting Standards Advisory Board (May 7, 2003), *available at* http://www.fasab.gov/pdf/mou05222003.pdf.

[242] *See* ACCOUNTING FOR SOCIAL INSURANCE, Statement of Recommended Accounting Standards No. 17 (Fed. Accounting Standards Advisory Bd., Aug. 1999) [hereinafter ACCOUNTING FOR SOCIAL INSURANCE], *available at* http://www.fasab.gov/pdf/17_ss.pdf (last visited Oct. 14, 2003).

[243] *Id.* at paras. 65–72.

[244] *Id.* at paras. 73–79.

[245] *See id.* at para. 74 ("[A]n accounting liability should be recognized at an earlier point than when payments are due and payable.").

[246] *Id.* at para. 79.

and to give the public a more nearly accurate picture of the financial posture of the Social Security program, proponents argued, Social Security financial statements should include some actuarial estimate of the future value of benefits.

After extensive debate, FASAB crafted a compromise between a form of accrual accounting and current practice. The final FASAB Statement on Social Insurance mandates that financial statements of government insurance programs include an elaborate system of supplementary information, known as Required Supplementary Stewardship Information (RSSI), which includes specific disclosures about the actuarial value of future benefits as well as a substantial amount of additional material about program sustainability.[247] With respect to Social Security, FASAB rules require disclosure of the actuarial present value of all future benefits payable to participants eligible to receive retirement benefits (those sixty-two years of age or older).[248] The Social Security RSSI also must include separate disclosures of the actuarial present value of future benefits to be paid and taxes to be received from those currently in the system (those fifteen to sixty-one years old) and those not yet in the system but projected to join the system over the next seventy-five years.[249] Taken together, these numbers sum to what was described above as the open-group liability of the system, and with the component line-item entries, users can calculate a number quite similar to the closed-group liability measure.[250] In the summer of 2003, FASAB revised its rules governing social insurance to place the information in a new Statement on Social Insurance, which will have greater prominence than the earlier RSSI format.[251]

As the FASAB establishes generally accepted accounting standards for government entities, its statement on social insurance is binding on the Social Security Administration. Indeed, if one consults the annual financial statements of the Social Security Administration—as opposed to the trustees' annual reports—one can find the supplementary statements of net present values of future benefits and taxes clearly disclosed along the lines FASAB requires.[252] For example, the most recent report estimates the present value of benefits payable to participants sixty-two years old or older to have been $4.4 trillion as of January 2, 2002.[253] Curiously, budgetary presentations of the OMB, though purporting to comply with FASAB standards, do

---

[247] *Id.* at paras. 24–27.

[248] *Id.* at para. 27(3).

[249] *Id.*

[250] The closed-group liability number is based on a seventy-five-year projection, as opposed to the one-hundred-year projection typically used by the Office of the Chief Actuary.

[251] *See* RECLASSIFICATION STATEMENT, *supra* note 186, at 2.

[252] *See* 2002 PERFORMANCE REPORT, *supra* note 171, at 75–78. Starting with the 2003 report, the trustees have begun to report alternative long-term measures of insolvency, but not the line-by-line components that FASAB requires. *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63.

[253] *See* 2002 PERFORMANCE REPORT, supra note 171, at 77.

not include RSSI information for Social Security or net present values of benefits and taxes.[254]

While the FASAB Statement on Social Insurance clearly does not go as far as the reforms this Article proposes, its provisions indicate that applying accrual accounting concepts to Social Security is not an entirely heretical idea. In the debate leading up to the adoption of the statement, industry experts made arguments similar to the ones advanced in this Article.[255] When FASAB began the rule-making process that ultimately moved the supplementary information to an earlier section of agency financial statements, it emphasized the importance of these disclosures.[256] Reading between the lines, one senses that the Board's unwillingness to advance reforms even closer to those proposed herein is dictated more by political resistance in Washington than by a conviction that a more nearly complete system of accrual accounting for Social Security would not be appropriate.

More important for purposes of this Article's argument is that the FASAB statement is based on a normative vision of the purpose of governmental accounting practices that is quite similar to the one described here. Drawing on its prior concept release on the Objectives of Federal Financial Reporting, the Board emphasized the relevance of the nation's financial condition for the financial statements of governmental entities underwriting social insurance. The Board called for "[F]ederal financial reporting [to] provide information that helps the reader to determine whether the government's financial position improved or deteriorated over the period."[257]

This Article's argument is that the current accounting presentation of Social Security does a very poor job in these dimensions and that the proposed alternative approach would represent a substantial improvement. In particular, the current financial statements of the trust funds are ineffective in explaining whether the system's financial position improved or deteriorated during a particular year.[258] Moreover, the absence of any measure of the system's accrued liabilities makes it quite difficult to tell whether budgetary resources will be able to sustain the current level of promised benefits.

---

[254] *See, e.g.*, OFFICE OF MGMT. & BUDGET, ANALYTICAL PERSPECTIVES: FISCAL YEAR 2003, at 32–33 (2002).

[255] A key difference between this Article's proposals and the most analogous proposals advocated in the development of the FASAB statement is the Article's recommended distinction between benefits accrued to date and benefits that will accrue in the future. While the FASAB approach employs an accrued-liability concept for participants who are already eligible for retirement, it uses a net present value computation for future benefits and payments for other participants—in effect, lumping accrued benefits with those to be accrued in the future and then deducting future taxes from those amounts. As explained above, there are a number of advantages to distinguishing accrued benefits from those that will accrue in the future. *See supra* Part III.C.2. Among other things, this distinction reflects the greater moral obligation of the government to honor accrued benefits.

[256] *See* RECLASSIFICATION STATEMENT, *supra* note 186, at 3–4.

[257] *See* ACCOUNTING FOR SOCIAL INSURANCE, *supra* note 242, at para. 7.

[258] *See supra* Part I.A.2.a.

In conclusion, Social Security's status as a social insurance program is not a strong argument against reflecting the system's mounting liabilities in its financial statements. Indeed, under generally accepted accounting standards for federal entities, these liabilities must currently be reported in supplementary notes, and a number of commentators have correctly argued that these liabilities should also be reflected on the balance sheets of social insurance programs.[259] In other words, the experts on government accounting have adopted an accounting standard that is substantially different from the one that the trustees follow in their annual reports and that dominates public debate over Social Security financing. There is strong intellectual support within the accounting community for the kinds of reforms proposed in this Article.

B. The Estimate of Public Pension Plan Obligations in Other Contexts

Another way to approach the question of how the obligations of the Social Security system should best be estimated is to consider how the issue is addressed in other contexts. As explained below, public policy analysts—principally economists—often need to calculate the size of public-pension-plan obligations. This Part reviews two prominent illustrations. The first is the public-finance literature dealing with the size of public-pension-plan obligations in various countries—often, but not exclusively, developing countries. The second context is a separate set of economic writings in which economists attempt to value Social Security benefits of workers, typically either to assess the impact of Social Security on other forms of savings or to present a full picture of household wealth. In both contexts, the standard approach is to estimate the present value of benefits to be paid in the future. Sometimes the calculation presents a net-present-value figure that includes both accrued and to-be-accrued benefits minus taxes to be paid. Increasingly, however, analysts are employing a benefits-accrued-to-date formulation similar to unfunded accrued liabilities measures that GAAP-style accrual accounting would highlight.

While these analogies are not directly related to financial accounting or government budgetary issues, the methodologies employed in these other areas are relevant to this Article's arguments. In all of the following examples, analysts are trying to estimate the economic reality of public pension obligations—whether in terms of the obligation they impose on the public fisc or in terms of the amount of value they add to individual wealth. In these contexts, analysts invariably resort to net-present-value calculations, often limited to benefits accrued to date. Within the literature, these measures have emerged as the most appropriate way to value Social Security obligations. That the traditional system of accounting for Social Security financing neglects comparable measures is noteworthy and deeply troubling.

---

[259] *See* ACCOUNTING FOR SOCIAL INSURANCE, *supra* note 242, at para. 78.

1. Measures of National Public-Pension-Plan Obligations

Over the past decade, as the crisis of Social Security financing has emerged as a major issue of public policy in the United States, a similar debate has emerged in a number of other countries, where the problem of unfunded public pension plans and often the private-pension-plan system is even more severe than in the United States. Many countries with the most seriously underfunded pension plans are in the developing world; other industrialized nations are not, however, devoid of serious deficits. Since the mid-1990s, a number of economists have written about the problem of underfunded public pension plans and have offered a variety of ways to address distressed systems, ranging from privatization to more incremental solutions. What is important about this literature is not the specific reforms advocated but, rather, the manner in which analysts quantify the magnitude of underfunded pension obligations in various countries and then present cross-country comparisons of pension underfunding.

For many years, the leading article on this subject was one written by two economists with the International Monetary Fund: Sheetal K. Chand and Albert Jaeger. Their 1996 working paper proposed techniques for estimating the public-pension-plan liabilities and then applied the techniques to the public systems of eight industrialized nations, including the United States.[260] The approach begins with a measure of accrued pension obligations for all pensioners and current workers.[261] These accrued liabilities, in Chand and Jaeger's terminology, constitute "recognition bonds"—the amount it would cost a government to terminate its public pension program (analogous to the Social Security system's maximum termination cost).[262] The Chand and Jaeger framework continues by adding benefits that will accrue in the future (to generate a "gross pension liability" figure) and then netting off projected contributions (to generate a "net pension liability" figure analogous to the open-group liability number discussed above). The balance of the Chand and Jaeger analysis relies principally on the net-pension-liability figure to evaluate various reform proposals.

As a seminal work on public pension financing, the Chand and Jaeger paper is an important precedent for evaluating the appropriateness of the traditional approach to Social Security financial statements. While elements of the Chand and Jaeger framework overlap with aspects of long-range financing estimations included in the annual trustees' reports, their organizing principles are strikingly different. Chand and Jaeger's approach is squarely grounded in present value calculations of future benefits, and it begins with a measure of the net present value of benefits accrued to date of the

---

[260] *See generally* SHEETAL K. CHAND & ALBERT JAEGER, AGING POPULATIONS AND PUBLIC PENSION SCHEMES (IMF Occasional Paper No. 147, 1996).

[261] *Id.* at 36.

[262] *Id.*

sort this Article proposes. Subsequent studies of public pension systems routinely adopt a similar perspective.[263] The trustees' reports traditionally have been wholly devoid of net present value calculations, and even the 2003 report includes these measures in only a limited manner, located some sixty pages into the document and entirely absent from introductory materials.[264] Not surprisingly, general press coverage of the 2003 trustees' report made no mention of these measures of insolvency.[265]

Another important aspect of the public-finance literature on unfunded pension obligations is the practice of comparing the size of these liabilities to the GDP of particular countries and then making explicit comparisons of this ratio with the ratio of traditional public debt to GDP. For example, in the Chand and Jaeger paper, the level of accrued liabilities of the U.S. Social Security system is reported as 108.3% of GDP in 1990 (31.7% representing obligations to retirees and 76.6% representing accrued obligations to workers).[266] Elsewhere, the paper aggregates the U.S. government's net public pension plan obligations in 1990 (25.7% of GDP) with its other public debt (63.3% of GDP) to come up with a "combined" net debt liability (89.0% of GDP). Other writers in the literature use a similar approach.[267] As explained

---

[263] A good example of this literature is ROBERT HOLZMANN, FINANCING THE TRANSITION TO MULTI-PILLAR (World Bank Soc. Prot. Discussion Paper No. 9809, 1998) (estimating the implicit pension debt of OECD member countries). Other examples abound. *See, e.g.*, SERGIO CLAVIJO, FISCAL EFFECTS OF THE 1993 COLUMBIAN PENSION REFORM (Working Paper of the Int'l Monetary Fund, IMF. Doc. WP/98/158, 1998); EDWARD WHITEHOUSE, PENSION REFORM IN BRITAIN (World Bank Soc. Prot. Discussion Paper No. 9810, 1998); CHEIKH KANE ET AL., BRAZIL: SOCIAL INSURANCE AND PRIVATE PENSIONS (World Bank Rep. No. 12336-BR, 1995); CHINA PENSION SYSTEM REFORM 125–29 (World Bank Rep. No. 15121-CHA, 1996). *See also* IMF FISCAL AFFAIRS DEP'T, DRAFT MANUAL ON FISCAL TRANSPARENCY 47 (Oct. 19, 1998) (endorsing measurement of "*unfunded public pension liabilities* alongside public debt in assessing sustainability" of current fiscal policies).

To be fair, other studies of the pension liabilities in the international context employ cash-flow analyses similar to the ones included in the trustees' reports. *See* DAVE TURNER ET AL., THE MACROECONOMIC IMPLICATIONS OF AGEING IN A GLOBAL CONTEXT 7 (Ageing Working Papers, OECD Doc. AWP 1.2 Eng, 1998). *See also* LOUISE FOX & EDWARD PALMER, LATVIAN PENSION REFORM 29–32 (World Bank Soc. Prot. Discussion Paper No. 9922, 1999); MANFRED KOCH & CHRISTIAN THIMANN, FROM GENEROSITY TO SUSTAIN-ABILITY: THE AUSTRIAN PENSION SYSTEM AND OPTIONS FOR ITS REFORM 16–23 (Working Paper of the Int'l Monetary Fund, IMF Doc. WP/97/10, 1997). Even Chand and Jaeger have written papers that employ cash-flow analyses. *See* SHEETAL I. CHAND & ALBERT JAEGER, REFORM OPTIONS FOR PAY-AS-YOU-GO PUBLIC PENSION SYSTEMS 12–13 (World Bank Soc. Prot. Discussion Paper No. 9927, 2000).

[264] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 61–63.

[265] *See supra* text accompanying notes 79–81.

[266] Chand & Jaeger, *supra* note 260, at 27 tbl.16.

[267] A recent paper by a group of World Bank economists on pension reform in China, for example, discusses the implicit pension debt of a wide range of countries measured as a percentage of their GDP. *See* YAN WANG ET AL., IMPLICIT PENSION DEBT, TRANSITION COST, OPTIONS AND IMPACT OF CHINA'S PENSION REFORM 9–13 (World Bank Policy Research Working Paper No. 2555, 2000) (estimating implicit pension debt of the United States at 113% of GDP). *See also* RICHARD DISNEY, OECD PUBLIC PENSION PROGRAMS IN CRISIS: AN EVALUATION OF REFORM OPTIONS 4–11 (World Bank Soc. Prot. Discussion Paper No. 9921, 1999); ROBERT HOLZMANN ET AL., PENSION SYSTEMS IN EAST ASIAN AND THE PACIFIC: CHALLENGES AND OPPORTUNITIES 30–32 (World Bank Soc. Prot. Discussion Paper No. 0014, 2000) (estimating implicit pension debt as a percentage of GDP in Korea, the Philippines, and Thailand).

above, one of the principal advantages of recognizing the actual present value of unfunded public pension obligations is that doing so facilitates comparisons of this sort and, presumably, better-informed discussions of public policy. With the traditional approach to Social Security accounting, such comparisons are extremely difficult. The system's failure to report net public pension liabilities as a percentage of GDP is, at a minimum, curious, because when analysts want to compare the financial status of the public pension systems of other countries to that of the United States, this is the measure typically employed.

Within the community of public economists, the principal issue of disagreement is not whether a country's unfunded pension liability should be measured on a present-value basis but rather which kind of present-value measure should be used. As explained above, the initial Chand and Jaeger paper used a net-pension-liability measure, decomposed into accrued and to-be-accrued components. This is reminiscent of the supplementary materials that FASAB requires, but it more sharply distinguishes accrued obligations to current workers and is substantially more informative than the single open-group liability figure that U.S. officials tend to report when required to estimate Social Security's unfunded obligations.[268] Within the community of policy analysts, however, there are those who think that the Chand and Jaeger net-pension-liability measure should be dropped as the principal summary statistic for measuring a country's unfunded public pension obligations and replaced with a figure that estimates the country's unfunded accrued obligations to date—that is, the measure of unfunded accrued liabilities highlighted in GAAP-style accrual accounting.

A forceful advocate of this view is the U.K.'s Richard Disney, who writes about the measures of pension liabilities in the context of the European Union, where the debt burden of member states is an important issue.[269] He summarizes his argument as follows:

> European governments should move from an *ad hoc* combination of cash flow accounting and projected liabilities in measuring the sustainability of pension schemes to a proper accrual basis, as is now taking place in other components of the government budget . . . . On an accrual basis, the budgetary report should provide a calculation of the change in accrued pension liabilities as a result of the government's receipt of pension contributions, net of pension payments, during the budget period. This should be supplemented by, but not confused with, additional analyses including actuarial confirmation that, at current contribution rates, *current* pension expenditure is indeed covered by current contribution re-

[268] *See supra* Parts II.C.2–.3.

[269] *See* Richard Disney, *How Should We Measure Pension Liabilities in EU Countries?*, *in* PENSIONS: MORE INFORMATION, LESS IDEOLOGY 95–111 (Tito Boeri et al. eds., 2001).

> ceipts, and by actuarial projections of future liabilities with, perhaps, some calculation of the contribution rates required to finance these prospective liabilities.
>
> By cumulating accrued liabilities up to the end of the current accounting period, the government would also arrive at the measure of the implicit or current accrued liabilities of the pension scheme. This measure is not the same as the prospective liability arising from the continuation of the pension scheme into the future. An attraction of providing a measure of implicit 'debt' along these lines is that it provides an exact measure of the current termination liability of the existing unfunded scheme. One reason why governments are so reluctant to consider greater pre-funding of pensions, especially if it involves a greater degree of private provision, is that funded reforms of this type make implicit debt explicit . . . .[270]

The views expressed in this excerpt closely track those of this Article, particularly in their recognition that the failure to acknowledge the magnitude of accrued pension obligations expressly biases consideration of certain reform proposals, such as pre-funded individual accounts.[271]

In short, the financial statements of the Social Security trust funds deviate from emerging international standards for the evaluation of the solvency of public pension plans. In addition, an important camp in these debates, epitomized by Disney, advocates solvency measures strikingly similar to the ones this Article proposes.

2. *Economic Estimates of the Value of Social Security Benefits*

Yet another way to estimate the extent of Social Security obligations is to consider the value of benefits due to participants. After all, every liability of the Social Security system reflects an asset of a participant or beneficiary. If the system truly had no outstanding liabilities—the fiction upon which the financial statements of the system now rest—then the interests of participants and beneficiaries must have no true value. This is clearly not the case. As a matter of political reality, Americans have a strong sense of entitlement to their Social Security benefits.[272] More importantly, when economists model consumer behavior, they routinely classify Social Security benefits as assets

---

[270] *Id.* at 102.

[271] In more recent work, many other economists are using measures of accrued liabilities as the appropriate estimate of implicit pension debt. *See, e.g.*, YAN WANG ET AL., *supra* note 267, at 9. For a recent overview endorsing an accrued liability definition of implicit pension debt for purposes of international comparison, see generally ROBERT HOLZMANN ET AL., IMPLICIT PENSION DEBT: ISSUES, MEASUREMENT AND SCOPE IN INTERNATIONAL PERSPECTIVE (World Bank Pension Reform Primer, 2001).

[272] *See supra* text accompanying notes 63–65.

and estimate their value based on the present value of expected benefits.[273] If the economists are right that Social Security benefits are best characterized as financial assets of individual participants, then the obligations of the Social Security system to honor those benefits are best characterized as liabilities of the system.

Examples of economic valuations of Social Security benefits for individuals abound, so this Part simply summarizes some prominent examples. In all cases, the goal of the analyst was to estimate the true economic effect of retirement benefits and the analyst employed some sort of discounted-value technique, typically using projections based on actual or assumed contributions to the Social Security system.[274]

a. Life Cycle Savings Literature

A familiar illustration of this approach to valuing Social Security benefits is a series of papers in which economists attempt to estimate the effect of Social Security benefits on individual savings. Many economists believe that public pension programs such as Social Security reduce other forms of savings and a number of economists have done empirical work exploring the relationship between Social Security benefits and savings. In a recent survey of literature on the subject, the CBO described the basic research methodology for the largest group of these studies:

> Most studies begin by estimating the total value of Social Security benefits that a person is expected to claim less the taxes to be paid, adjust for the length of time before the benefits will be received (or taxes paid) and the probability that the recipient will survive—the "present value" of benefits minus taxes. That sum is referred to as Social Security wealth. Then, using regression analysis, a researcher tests whether the private wealth held by people is related

---

[273] *See infra* Part IV.B.2.

[274] To be fair, other economic studies use discounted cash-flow systems to estimate the unfunded liability of the Social Security system or to calculate the implicit rate of return on Social Security contributions. For examples of economists' estimates of the system's unfunded liabilities, see *supra* note 127. For a well-known example of economists' using discounted cash-flow analyses to estimate rates of return on Social Security contributions, see John Geanakoplos et al., *Social Security Money's Worth*, *in* PROSPECTS FOR SOCIAL SECURITY REFORM 79–151 (Olivia S. Mitchell et al. eds., 1999). *See also* 1994–1996 ADVISORY COUNCIL REPORT, *supra* note 2, at 15–21 (making extensive use of money's-worth analysis for various reform options); STEUERLE & BAKIJA, *supra* note 60, at 106–15 (estimating net transfers from Social Security for various income groups and generational cohorts). Lawrence Kotlikoff's work on generational accounting is similarly spirited in its comparison of the net present value of Social Security contributions and benefits by generational cohort. *See, e.g.*, LAURENCE J. KOTLIKOFF, GENERATIONAL ACCOUNTING: KNOWING WHO PAYS, AND WHEN, FOR WHAT WE SPEND ch. 5 (1992).

to their Social Security wealth, controlling for other factors such as age and income.[275]

This approach is telling. When trying to estimate the significance of Social Security for individual behavior, economists routinely invoke discounted cash flow analyses to estimate the Social Security wealth of individual recipients. Their regression models are exploring the relationship between this asset—Social Security wealth—and other forms of private savings.

If one were to analogize this methodology to the earlier discussion of Social Security's unfunded obligations, the approach used in these studies is conceptually similar to the closed-group-liability measure, which is also the measure of implicit trust debt reflected in the modified accrual accounting approach. The study is limited to current participants in the system, and combines the present value of accrued and to-be-accrued benefits and then deducts the present value of to-be-contributed pay-roll taxes. If one were to aggregate this measure of Social Security wealth over all current participants and beneficiaries—that is current worker, retirees, and their beneficiaries—one should in theory generate a level of assets equal to the trust funds' closed-group liability or implicit trust fund debt ($10.5 trillion as of December 31, 2002).[276] If, as this literature implies, Social Security wealth is an appropriate way to estimate the value of Social Security benefits to current participants and beneficiaries, then why is a comparable methodology not appropriate for the liabilities of the Social Security trust funds?

b. Household Wealth Literature

A separate body of economic literature explores the composition of individual wealth in the United States. A prominent example of this literature is the 1992 Health and Retirement Study of a nationally representative sample of households, which includes considerable information on Social Security benefits.[277] The study presents several different valuations of the Social Security wealth of study participants. All of the study's valuations represent the present value of expected Social Security benefits discounted by an appropriate interest rate and adjusted for the life expectancies of beneficiaries. The study's first valuation technique limits itself to the present value of benefits that individuals have accrued as of the date of the study.[278] Two subsequent estimates represent the value of the participants' total projected

---

[275] *See* CONG. BUDGET OFFICE, SOCIAL SECURITY AND PRIVATE SAVING: A REVIEW OF EMPIRICAL EVIDENCE 10 (July 1998) (reviewing thirty empirical studies of the effects of Social Security on savings).

[276] *See* 2003 TRUSTEES REPORT, *supra* note 8, at 63.

[277] *See generally* Olivia S. Mitchell et al., *Social Security Earnings and Projected Benefits*, *in* FORECASTING RETIREMENT NEEDS AND RETIREMENT WEALTH (Olivia S. Mitchell et al. eds., 2000) (describing study's methodology for valuing Social Security benefits).

[278] *Id.* at 345–50.

benefits—either to the age of sixty-two or to the individuals' normal retirement age.

When economists want to explore the wealth of individual households—for example, to consider variation in wealth across income levels—they routinely rely on data sources such as the 1992 Health and Retirement Study, and they routinely include measures of Social Security wealth as important components of overall household wealth. Typically, researchers take one of two approaches to estimate the value of Social Security wealth.

First, some economists use a measure of Social Security wealth that is similar to the one used in the life-cycle savings literature—that is, a measure based on the present value of accrued and to-be-accrued benefits minus the present value of to-be-paid payroll taxes.[279] As explained above, this approach is analytically similar to the closed-group liability figure and the implicit trust fund debt measures of modified accrual accounting. It includes within the definition of Social Security wealth the present value of accrued benefits plus the net present value of benefits expected to accrue in the future and payroll taxes to be paid in the future.[280]

A second approach is to estimate Social Security wealth based solely on accrued benefits to date.[281] These estimates of the value of Social Security benefits are analogous to GAAP-style accrual accounting and are sometimes used to estimate both asset values for accrued Social Security benefits and also annual Social Security accruals as a component of overall household income.[282]

These surveys on household wealth are noteworthy in two respects. First, they demonstrate that economists clearly regard Social Security benefits as financial assets that are appropriately combined with other assets, such as private pensions, bank accounting, and home ownership. Second, the technique these economists use to value Social Security benefits is highly comparable to the valuation techniques this Article recommends for valuing Social Security trust funds liabilities. A review of these and many other economic studies[283] reveal that it is common practice for economists to attribute real eco-

---

[279] *See, e.g.*, ARTHUR B. KENNICKELL & ANNIKA E. SUNDEN, PENSIONS, SOCIAL SECURITY, AND THE DISTRIBUTION OF WEALTH 7–9 (Fed. Reserve Bd., Finance and Economics Discussion Paper No. 97-55, Nov. 6, 1997) (analysis based on Survey of Consumer Finances, a study sponsored by the Federal Reserve Board and Internal Revenue Service triennially).

[280] In the Kennickell and Sunden paper, only employee contributions are deducted (as opposed to combined employer and employee contributions). *See id.* at 9, n.11.

[281] *See, e.g.*, Alan L. Gustman et al., *Pension and Social Security Wealth in the Health and Retirement Study*, *in* WEALTH, WORK AND HEALTH: INNOVATIONS IN SURVEY MEASUREMENT IN SOCIAL SCIENCES 150 (Robert Willis ed., 1999) (using Health and Retirement Study data to estimate Social Security wealth).

[282] *See id.* at 162 tbl.2.

[283] There are numerous other cases where economists quantify Social Security benefits in a similar manner. For example, studies that explore the effect of Social Security on retirement decisions often quantify the effect of Social Security by estimating the amount by which the accrued value of Social Security benefits increases over the course of a year. *See, e.g.*, COURTNEY COILE & JONATHAN GRUBER, SOCIAL SECURITY AND RETIREMENT 14 (Boston Coll. Ctr. For Ret. Research, Working Paper No. 2000-11, 2000) ("We initially follow the literature and

nomic value to Social Security benefits and to use net present value techniques to estimate those values. Again, the question for defenders of current Social Security accounting practices is why similar techniques are not used to estimate the liabilities of the Social Security trust funds.

---

focus on accrual, the change in [Social Security Wealth] resulting from an additional year of work.").

Other economic writing looks at the relationship between the accrual of Social Security benefits and the accumulation of pension wealth. *See, e.g.*, CHRISTOPHER M. BONE, ACTUARIAL PERSPECTIVES ON IMPLICATIONS OF SOCIAL SECURITY REFORM FOR EMPLOYER SPONSORED PENSION PLANS (Pension Research Council, Working Paper No. 97-01, 1997). *See also* PETER R. ORSZAG, SHOULD A LUMP-SUM PAYMENT REPLACE SOCIAL SECURITY'S DELAYED RETIREMENT CREDIT? (Boston Coll. Ctr. for Ret. Research, Issue Brief No. 6, Apr. 2001) (estimating the value of delayed retirement by taking the discounted present value of additional benefits accrued).

COMMENT

LIABILITIES, DEBTS, REVENUES, AND EXPENDITURES: ACCOUNTING FOR THE ACTUARIAL BALANCE OF SOCIAL SECURITY

ROBERT L. CLARK*

Determining the financial status of the Old Age Survivors and Disability Insurance program (OASDI or Social Security) is complex and difficult. Understanding the current and future financial status of Social Security, however, is necessary for modifying and reforming national retirement policies. A number of methods illustrate the current and projected future actuarial balance of Social Security. In his interesting article,[1] Professor Jackson describes some of the methods, selects his favorite, and argues that it is the best method of presenting the financial status of Social Security. He makes a series of important observations, but also ignores some key issues associated with unbiased reporting of the financial status of Social Security. This Comment begins by describing the range of possible financial measures of Social Security's health and examining the role of each measure. The next Part presents a basic critique of Jackson's article and its analysis of the 2003 Trustees Report of the OASDI Trust Funds.

I. ACCOUNTING FOR SOCIAL SECURITY: MEASURING ITS FINANCIAL STATUS

Policymakers, the press, and the public need to know the current and projected future financial status of Social Security. Optimal retirement policies should be based on quality research, unbiased projections, and sound accounting practices. It is extremely important to know:

(1) whether the Social Security system is actuarially solvent using current benefit formulas and tax rates and what solvency means,

---

* Professor, College of Management, North Carolina State University. Chairman of the 2003 Technical Panel on Assumptions and Methods that reviewed and evaluated the long-range projections of the Trustees of Old Age Survivors and Disability Insurance as presented in their annual reports. Ph.D., Duke University, 1974; M.A., Duke University, 1972; B.A., Millsaps College, 1971.

[1] *See* Howell E. Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

(2) what level of taxation is needed to pay promised benefits if the basic structure of Social Security is maintained, or, alternatively, how much benefits would need to be reduced so current tax rates would provide sufficient revenues,

(3) whether the trust funds are increasing or decreasing in value and when these funds will be completely depleted, and

(4) what is the magnitude of implied liabilities accrued by workers to date and how much additional money would be needed to pay off all promised benefits if OASDI were terminated today.

Traditionally, the Trustees of OASDI have focused on determining the seventy-five-year actuarial balance of the system and the size and trend of assets in the Trust Funds.[2] Jackson asserts that accrual accounting of promised benefits based on participation to date is the best method of summarizing the current financial status of Social Security.[3] In essence, he argues that the best accounting method for OASDI is based on the termination of Social Security and the needed revenues to fund (over and above existing assets held in the Trust Funds) the present value of all promised benefits. Another method of determining the present value of Social Security net liabilities is to assume that OASDI will continue to exist, to determine the present value of paying all present and future benefits, and finally, to compare this to the existing Trust Funds and the present value of all future scheduled tax collections. This method assumes Social Security will be maintained and tries to determine the magnitude of any projected shortfall in revenues.

Jackson aims his criticism at the 2003 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds (hereinafter, "2003 Trustees Report") and its measures of financial status.[4] Early drafts of his paper were written prior to the publication of the report. The report incorporates numerous changes in the evaluation and measurement of the financial status of Social Security.[5] The 2003 Trustees Report is a significant improvement over previous reports, and as such, sharply reduces the relevance of many of

---

[2] *Compare* 2002 BD. OF TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. [hereinafter 2002 TRUSTEES REPORT], *with* 2003 BD. OF TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. DISABILITY AND INS. TRUST FUNDS ANN. REP. [hereinafter 2003 TRUSTEES REPORT].

[3] *See* Jackson, *supra* note 1, at Part II.C.5.

[4] *See id.* at Part I.A.

[5] *See* 2003 TRUSTEES REPORT, *supra* note 2, at 42, 72–74. The report presents estimates of the present value of the seventy-five-year open-group unfunded obligations, the present value of the closed-group unfunded obligations, and the present value of the infinite-time period open group unfunded obligations. *See* 2003 TRUSTEES REPORT, *supra* note 2, at 61–63. In addition, the report indicates the amount that benefits would need to be reduced for expenditures to equal revenues obtained from currently scheduled payroll tax rates. *See* 2003 TRUSTEES REPORT, *supra* note 2, at 3, 8.

Jackson's criticisms. Jackson's paper would have more relevant policy implications if it acknowledged the material presented in the report and recognized the important changes it has made in accounting for Social Security. In particular, he should compare his preferred method of accrual accounting to the present-value calculations in the report rather than devote so much attention to the methods that dominated the pre-2003 reports.

Prior to 2003, the Trustees Reports focused almost exclusively on three concepts of the financial status of Social Security: the actuarial balance, the year that the Trust Funds will be depleted, and the change in the size of the Trust Funds. The actuarial balance concept is a useful indicator of the financial status of an ongoing Social Security system. It indicates the amount payroll taxes would need to be raised today (and continue at this higher rate over the next seventy-five years) to fund the system during this period. The pre-2003 Trustees Reports gave this measure by far the most attention. The 2002 Trustees Report indicated that the actuarial balance over the seventy-five-year projection period was –1.87% of taxable payroll.[6] This measure shows Congress how much additional revenue as a percent of payroll is needed over a seventy-five-year period if benefits are held constant. The 2003 Report showed an actuarial balance of –1.92% of taxable payroll, indicating that the financial status of the system had deteriorated slightly.[7]

The year that the Trust Funds have zero assets is another important measure of the financial status of OASDI. Law requires that Social Security benefits be paid from revenues generated by the payroll tax or from monies from the Trust Funds.[8] If the balance in the Trust Funds is zero and payroll tax revenues are insufficient then full benefits cannot be paid. The imminent exhaustion of the Trust Funds was a looming crisis in the early 1980s.[9] The immediate prospect of insufficient funds stimulated Congress to enact substantial changes in OASDI proposed by the Greenspan Commission in 1983.[10] Thus, the date of exhaustion of all assets in the Trust Funds has important policy implications and may affect reform decisions. The 2003 Trustees Report estimates that the OASDI Trust Funds will be exhausted in 2042.[11]

The annual change in the value of the Trust Funds also provides important information that indicates whether the Funds are increasing or

---

[6] *See* 2002 TRUSTEES REPORT, *supra* note 2, at 3.

[7] *See* 2003 TRUSTEES REPORT, *supra* note 2, at 2.

[8] *See* 42 U.S.C. § 401(a)(3)-(4) (2000). Section 401 of the Social Security Act appropriates into the Trust Funds amounts equal to the payroll and self-employment taxes reported to the Internal Revenue Service. *Id. at* § 401(a).

[9] For a discussion of the financial status of OASDI in the early 1980s, see generally GREENSPAN COMM'N, REPORT OF THE NATIONAL COMMISSION ON SOCIAL SECURITY REFORM (1983), *available at* http://www.ssa.gov/history/reports/gspan.html.

[10] *See id.* at 1.

[11] *See* 2003 TRUSTEES REPORT, *supra* note 2, at 3.

decreasing in assets. In 2002, the Trust Funds added $165.4 billion in assets.[12] If viewed in isolation, this increase in value does give the impression of a system that is financially secure. Jackson is concerned especially that this measure can be misleading.[13] He is correct on this point, and thus it is imperative that other measures be included in any evaluation of the financial status of Social Security.

Although each of these items is important to understanding the financial status of Social Security, considered alone they present an incomplete and somewhat misleading picture. The actuarial balance concept focuses attention almost exclusively on the need for tax increases. While this is one possible response to the current underfunding of OASDI, policy makers might decide to reduce future benefits instead of raising taxes. Additionally, the seventy-five-year actuarial balance concept does not provide any information on either the amount of money needed if the system were to be terminated and all accrued liabilities paid or the amount of money needed from general revenue funds to cover the shortfall in revenues. Neither does the seventy-five-year actuarial balance provide the necessary information to assess the merits of reform proposals that require upfront transitional costs and contain positive revenue effects outside the projection period.

It is against this backdrop that Jackson developed his analysis. Many of his criticisms apply most directly to the pre-2003 Trustees Reports. In particular, he argues that describing changes in the Trust Funds presents a misleading picture when there are short-term annual improvements but large long-term annual deficits.[14] Additionally, he criticizes the limitation of a fixed-term projection period and identifies the need for an indicator of the present value of outstanding liabilities.[15] The world of Social Security accounting changed significantly in the last year, and unfortunately, Jackson focuses too much attention on past practices without giving sufficient attention to the new accounting methods adopted by the Trustees in the 2003 report.

II. Accrual Accounting

Jackson argues that accrual accounting is the best and most informative method of illustrating the financial status of Social Security. Simply stated, under the accrual accounting method proposed by Jackson the current financial status of OASDI is the present value of all accrued benefits by workers and retirees minus the assets in the trust funds. This is the amount of additional money that would be needed if Social Security were

---

[12] *See id.* at 4.
[13] *See* Jackson, *supra* note 1, at introdution.
[14] *See id.* at Part I.A.2.a.
[15] *See id.* at Part I.A.2.a–b.

terminated today, no future taxes were paid, no additional benefits were earned, and all promised benefits based on past taxes were paid. It is also the maximum transition cost of changing to a fully funded system or the adoption of a system of individual accounts in which the government transfers sufficient assets to cover all accrued benefits earned to date.

I agree with Jackson that the accrual accounting measure of benefit liability is an important and useful indicator of the financial status of Social Security. Annual changes in this measure are another significant indicator of the financial status of the system. The fundamental question concerning accounting for Social Security is whether this is the only measure that should be reported or is even the best measure of the financial status of OASDI.

Jackson bases his conclusion on the premise that all promised benefits will be paid. He contends that there is a common belief among workers and their families that they accrue an entitlement—politically, if not constitutionally—to receive Social Security benefits at promised levels sometime after entering the workforce and prior to reaching retirement age.[16] While he acknowledges that retirement benefits are not legally guaranteed, he argues that they are politically guaranteed and thus should be viewed as a certain obligation or debt of the federal government.

Jackson completely ignores the fact that the last two major reforms of OASDI in 1977 and 1983 dramatically cut retirement benefits. In 1977, Congress altered the indexing formula used to calculate retirement benefits.[17] Though one can argue that this legislation was merely correcting a mistake introduced into the system by the 1972 amendments, the result was very real to millions of older Americans—their retirement benefits were substantially lower than they expected and much lower than those of persons who began receiving benefits only a few years earlier.[18] The impact of the 1977 legislation on the so-called "notch babies" and all future cohorts was a major reduction in benefits.[19] This experience directly contradicts the fundamental principle on which Jackson bases his argument that accrual accounting is the best method of assessing the financial status of Social Security.

The 1983 Social Security amendments postponed the cost of living adjustment for current retirees, and, more importantly, increased the

---

[16] *See id.*

[17] For a discussion of the 1977 Social Security Amendments, see John Snee & Marry Ross, *Social Security Amendments of 1977: Legislative History and Summary of Provisions*, SOC. SEC. BULL., Mar. 1978, at 3–20; *see also* Larry DeWitt, *Brief History*, *at* http://www.ssa.gov/history/briefhistory3.html (updated Mar. 2003).

[18] *See* SNEE & ROSS, *supra* note 17, at 13.

[19] *See* SYLVESTER SCHIEBER & JOHN SCHOVEN, THE REAL DEAL. 176–82 (1999). The benefit reductions imposed by the 1977 Amendments were referred to as "the benefit notch," and persons retiring who were directly affected by the reduction in benefits were called "notch babies."

normal retirement age for future retirees from sixty-five to sixty-seven.[20] Increases in the normal retirement age and the accompanying decreases in the proportion of the primary insurance amount (PIA) received at all ages is realistically equivalent to a reduction in retirement benefits. For example, individuals born in 1943 who start their retirement benefits at age sixty-five will receive only 93.3% of their PIA instead of the 100.0% of PIA that earlier cohorts of retirees received. Thus, these individuals who were aged forty at the time of the 1983 amendments had a permanent reduction in the future benefits. Once again, history does not conform to the basic tenet of Jackson's argument.

Projected revenues based on current law are insufficient to pay all promised benefits. Past modification of Social Security suggest that any future reform of OASDI will likely include a reduction in promised benefits. Jackson ignores historical experience when he concludes otherwise. He also ignores the recommendations of a series of presidential and congressional commissions, including the 1994–96 Social Security Advisory Council and the Bush Commission to Strengthen Social Security.[21] All of these review panels made recommendations for reductions in promised benefits as part of Social Security reform. In this way, Jackson's assertion that all accrued benefits will be paid is contradicted also by most reform proposals. Once the premise that all promised benefits will be paid is removed, Jackson's argument that the accrual accounting measure is the best indicator of the financial status of OASDI is substantially weakened. Thus, accrual accounting should be considered a useful and important piece of information that should be reported along with other measures to provide a more complete picture of the financial status of OASDI rather than the best measure of the financial status of Social Security.

III. Best Unbiased Estimates of Financial Status

The primary objective of the annual Trustees Report should be to present the best, most unbiased estimates of the financial status of OASDI. This goal demands projections of future revenues based on the current payroll taxes and future expenditures based on the current benefit formulas. Such projections must be long term and based on assumptions that reflect the best available information and research. Projections indicate annual differences between expenditures and revenues. After making

---

[20] *See* Social Security Amendments of 1983, Pub. L. No. 98-21, 97 Stat. 65 (1983).

[21] *See* 1994–1996 Soc. Sec. Advisory Council, Report On Social Security: Findings And Recommendations 25–33 (1997) [hereinafter 1994–1996 Advisory Council], *available at* http://www.ssa.gov/history/reports/adcouncil/report/toc.htm; President's Comm'n To Strengthen Soc. Sec., Strengthening Social Security And Creating Personal Wealth For All Americans 9–15 (2001), *available at* http://www.csss.gov/reports/Final_report.pdf.

the best and most unbiased projections, the next step is to determine how to summarize these results and present the data for evaluation. Prior to the 2003 Trustees Report, the Trustees reported the seventy-five-year actuarial balance and the year of depletion of the Trust Funds.

Jackson's strongest arguments for the merits of accrual accounting are based on a comparison to the methods used in the pre-2003 Trustees Reports. For example, he writes, "the trustees' reports are, in essence, statements of annual cash flows, comparing the system's annual cash receipts to its yearly payments for benefits and administrative expenses."[22] Additionally, he contends, "as long as inflow was adequate to meet outflow, the system was considered to be in balance."[23] These statements leave the impression that annual revenues and expenditures are compared and that OASDI is declared to be on sound financial footing if there are surpluses in the current year. This has not been the case for many years. All recent Trustees Reports have considered a seventy-five-year horizon of revenues and expenditures. Prior to 2003 Trustees Report, the reports presented the actuarial balance as the increase in payroll taxes needed to pay all promised benefits over this entire period. Thus, even though there are current annual "surpluses," the system was not considered to be in actuarial balance because of long-run deficits.

These measures presented a limited and somewhat biased assessment of the financial status of Social Security. In contrast, the 2003 Trustees Report presents a much more balanced assessment of the financial status of Social Security. First, throughout the report, the Trustees indicate the amount benefits would need to be reduced if revenues are not increased.[24] From this data the reader is able to consider the two basic options confronting policy makers (increasing taxes or reducing benefits). Policymakers and the public need to know the OASDI deficit in terms of both higher taxes and lower benefits. Together these measures present the range of possible policies that can be considered. Moreover, the explicit possibility of benefit reductions is acknowledged in this latest report.[25] Current Social Security accounting thus goes beyond the assumption adopted by Jackson.

Second, in 2003, the trustees for the first time present the present value of the seventy-five-year actuarial balance. This value indicates the amount of money that is needed today to provide sufficient funds to pay all promised benefits if the payroll tax and the benefit formula are unchanged. This concept provides an assessment of the additional funds needed today to finance the current system over the next seventy-five years. Jackson certainly overstates the facts by arguing that the reports do not

[22] Jackson, *supra* note 1, at introduction.
[23] *Id.* at Part I.
[24] *See* 2003 TRUSTEES REPORT, *supra* note 5, at 8.
[25] *See id.* at 3, 8.

recognize benefit obligations until the year of payment.[26] Since the long-range projections are based on benefit obligations and expected tax revenues, Jackson is incorrect. As reported in the 2003 Trustees Report, the present value of the seventy-five-year shortfall is $3.5 trillion.[27] This means that if $3.5 trillion were immediately transferred to the OASDI Trust Funds, all promised benefits could be paid with currently scheduled taxes.

Third, Jackson criticizes the use of the seventy-five-year projection of unfunded liabilities and points out the well-known "cliff" problem associated with projections of any specific length of time.[28] For years, analysts have recognized and discussed the implications of a fixed-term projection period.[29] Jackson correctly notes that an infusion of $3.5 trillion of new money to put the system in balance for seventy-five years would only temporarily solve the financial problem of insufficient revenues as a new deficit would emerge with each passing year.[30] He overlooks, however, the fact that in the 2003 Trustees Report the Trustees also presented an infinite time horizon estimate of the deficit. The infinite time horizon does not suffer from the "cliff" problem. The report estimates that the present value of the deficit associated with retaining the current system for an infinite time horizon is $10.5 trillion.[31]

These latter two concepts, the seventy-five-year actuarial balance and the infinite time horizon, are measures of the unfunded obligations of OASDI conditional on the system continuing to operate. They are important counterparts to the accrual accounting debt proposed by Jackson that illustrate the unfunded obligation conditional on the termination of the system. How should we decide which method should be the primary focus of attention? Jackson weakens his case by making a series of misstatements or overstatements concerning the current Trustees Report and the release of financial information for the system. Today, it would seem more likely that Social Security will be retained in some form, and so we should focus on the best methods of evaluating the financial status of the on-going system rather than focusing on the liabilities associated with terminating Social Security.

IV. Policy Confusion and the Ignorance of the Press

Beside his emphasis on the best measure of financial accounting for Social Security, Jackson is concerned with how the information is used by policy makers and the press. This is an important consideration, but even transparent accounting can be abused. Jackson describes a possible

---

[26] *See* Jackson, *supra* note 1, at Part I.A.2.a.
[27] *See* 2003 Trustees Report, *supra* note 5, at 3.
[28] *See* Jackson, *supra* note 1, Part I.A.2.b.
[29] *See, e.g.*, 1994-1996 Advisory Council, *supra* note 21, at 12.
[30] *See* Jackson, *supra* note 1, at Part I.A.2.b.
[31] *See* 2003 Trustees Report, *supra* note 5, at 61.

policy response to addressing the financial deficit using the current methods: the one-time transfer of sufficient monies to "solve" or eliminate the actuarial deficit.[32] He then concludes, "one of the unfortunate consequences of focusing on long-range actuarial deficits is that it encourages irresponsible recommendations of this sort."[33]

First, I am not sure what is irresponsible about providing additional money to cover a deficit. Second, Jackson's own proposal would create exactly the same incentive. Using accrual accounting, policymakers would be shown the amount of "debt" the system faces. There would still not be a tendency to try to solve the debt overhang with additional monies from the general fund, such as the Clinton proposals to allocate a portion of the projected federal surplus to OASDI.[34] If we relied entirely on accrual accounting as Jackson suggests, the underfunding of Social Security could be fully resolved by the federal government's issuing new government securities and transferring them to the OASDI Trust Funds. This solution would eliminate the Social Security unfunded obligation. In Jackson's analysis, it would have no effect on the government's financial status, because the Social Security debt is merely being replaced by government bonds and Jackson implicitly assigns the same standing to both types of government obligations.

In today's reform debates, transitional costs are well known. Any proposal for termination of the current OASDI program and replacing it with an individual account system must address the issue of transitional costs. This issue was made explicit by one of the proposals in the 1994–1996 Social Security Advisory Council.[35] Transitional costs and future revenue gains were also an important part of the discussion surrounding the recommendations of the Commission to Strengthen Social Security.[36] Jackson seems to have missed these debates as he writes, "In discussions of reform proposal, particularly those involving the partial privatization of Social Security benefits, analysts sometimes overlook the accrued claims of current workers and retirees."[37] While it is undoubtedly true that some people overlook these costs, no reform proposals will receive serious consideration without directly addressing these transition costs, and all recent proposals by leading advocates of reform have included explicit methods of dealing with the transition costs.

The concept of transition from one retirement plan to another raises additional questions. Specifically, one needs to consider which workers

---

[32] *See* Jackson, *supra* note 1, at Part I.A.2.b.

[33] *See id.* at Part I.A.2.c.

[34] *See generally* Douglas W. Elmendorf et al., *Fiscal Policy and Social Security Policy During the 1990s*, *in* AM. ECON. POL'Y IN THE 1990S 89, 106–09 (Jeffrey Frankel & Peter Orszag, eds., 2002).

[35] *See* 1994–1996 ADVISORY COUNCIL, *supra* note 21, at 25–33.

[36] *See* PRESIDENT'S COMM'N TO STRENGTHEN SOC. SEC., *supra* note 21, at 73–93, *available at* http://www.csss.gov/reports/Final_report.pdf.

[37] *See* Jackson, *supra* note 1, at Part II.C.4.a.

would be required to shift to the new plan, who would have the option of changing plans, and who would remain in the old plan. Jackson discusses this issue and concludes that somehow accrual accounting would make the answers to these questions easier to ascertain.[38] In fact, virtually all reform proposals begin by stating that current retirees will not be affected.[39]

V. FINAL ASSESSMENT

Future retirement policy must be based on a clear understanding of the cost of maintaining the current Social Security program and the cost of modifying or eliminating OASDI. The issue is what information the public and Congress need to develop an optimal retirement policy for the 21st century. In his thoughtful article, Jackson proposes that the primary method of presenting the financial status of Social Security be based on accrual accounting of existing liabilities and assets. His position is based on two basic principles: (1) participants in Social Security have an earned right to benefits based on taxes paid to date and (2) measurement of net liabilities is best done on a termination basis.[40]

Recognition of accrued liabilities is a useful and important aspect of evaluating the financial status of OASDI. It is not, however, necessarily the best method nor a method that is sufficient alone. First, promised benefits are not guaranteed. They have been reduced in the past and are very likely to be modified as part of any reform of Social Security. Second, OASDI is likely to be retained in some form. The current accounting approach is based on existing law and so to assess the financial status of Social Security under this assumption we should consider the projections of promised benefits along with scheduled taxes.

Every four years, the Social Security Advisory Board appoints a technical panel to review the assumptions and methods used in the annual Trustees Reports. I chaired the 2003 Panel that carefully reviewed the 2002 and 2003 reports. The Panel strongly endorsed the new methods for reporting the financial status of OASDI.[41] In particular, the Panel believed the inclusion of the present-value estimates and the addition of indications of the benefit reductions needed to bring the system into actuarial balance greatly improved the reports.[42] Moreover, the Panel believed that appropriate indicators of the financial status of OASDI as a continuing

---

[38] *See id.* at Part III.F.

[39] *See, e.g.*, 1994–1996 ADVISORY COUNCIL, *supra* note 21.

[40] *See id.* at Part I.A.2.a.

[41] *See* 2003 TECHNICAL PANEL ON ASSUMPTIONS AND METHODS, REP. TO THE SOC. SEC. ADVISORY BD. 87 (2003).

[42] *See id.*

program were superior to accrual accounting methods based on the termination liability.[43]

Since Jackson began his paper, the OASDI Trustees have made significant changes in their annual report. If Jackson were starting his paper today, it most likely would have a much different emphasis. The case for accrual accounting must be made in comparison to the new measures presented in the 2003 Trustees Report that are based on the continued operation of OASDI. Based on the changes in the report, the central issue is why accrual accounting should be preferred to the use of the present value of projected revenues minus expenditures of an ongoing retirement program. Unfortunately, Jackson spends too much time attacking the shortcomings of other accounting methods.

---

[43] *See id.*

COMMENT

ACCRUAL ACCOUNTING FOR SOCIAL SECURITY

PETER A. DIAMOND*
PETER R. ORSZAG**

Professor Jackson's article[1] represents a thoughtful piece of analysis. Regardless of what conclusions one reaches regarding his principal suggestions, his article raises important objections to the current system of accounting for Social Security and deserves careful study by Social Security and budget experts.

Jackson embraces the use of accrual accounting as the primary metric for evaluating Social Security's finances for both Social Security reform discussions and annual federal budgeting. Accrual accounting measures, which are already available to experts, would be a useful addition to the information provided regularly to the public and should be more easily available and more widely cited. They should not, however, replace the central measures now used in policy discussions. Including accrual measures of Social Security in the annual federal budget could cause small changes in assumptions to generate very large changes in budget outcomes. The resultant political pressure to alter projection assumptions would pose serious risks to the perceived legitimacy of the actuarial projections.

Jackson describes different accounting measures for the financial position of Social Security, and argues for making two of them central in popular and political discussions of Social Security and the federal budget. He also argues that the structure of a Social Security reform plan should reflect these preferred public accounting measures. Jackson's proposals would represent a far-reaching revision of federal accounting, and the opportunity to explore all its possible ramifications are beyond the scope

---

* Institute Professor and Professor of Economics, Massachusetts Institute of Technology. Ph.D., Massachusetts Institute of Technology, 1963; B.A., Yale, 1960.

** Joseph A. Pechman Senior Fellow in Economic Studies, Brookings Institution; Co-Director, Tax Policy Center. Ph.D., London School of Economics, 1997; M.Sc., London School of Economics, 1992; A.B., Princeton University, 1991. The views expressed here are the authors' alone and do not necessarily represent those of the staff, trustees, or officers of the institutions with which the authors are affiliated. The authors thank Henry Aaron, William Gale, and Howell Jackson for helpful comments.

[1] *See* Howell E. Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

of this Comment. This Comment therefore merely describes first reactions to some of Jackson's key points.

I. Change in Maximum Transition Cost

A great deal of information is available to interested experts about the current and projected financial position of Social Security.[2] Analysts are well aware of the different measures and can consider how any reform of the system would affect them. Jackson reviews these measures and effectively introduces a new measure by emphasizing the annual change in one of them. Before turning to the specific measure that Jackson embraces, it is worth noting that most of the alternative measures of Social Security's financial position are not easily accessible to the public. Jackson's suggestion that these measures be included in the most visible public reports is compelling; our disagreement focuses on whether any of these alternative measures should serve as the focal point for reform discussions.

Analysts have previously calculated the maximum transition cost, which reflects the cost of all accrued benefits to date minus the existing Trust Fund. Jackson calculates the annual change in this maximum transition cost, which no one to our knowledge had previously computed.[3] Jackson's calculation provides a measure of the growth in accrued benefits, adjusted for the realized cash flow experienced during the year.

The change in the maximum transition cost provides one way of measuring the cost of delaying reform. Analysts regularly argue that a delay in addressing the actuarial imbalance makes the problem more difficult to address—that the passage of time locks in place more of the benefits that could be reduced and forgoes more of the earnings that could be subject to higher taxation. Typically, analysts illustrate this point by highlighting the larger steps necessary to achieve any given definition of long-term balance if action were delayed. In 2001, The Social Security Advisory Board, for example, noted,

---

[2] The availability of this information distinguishes Social Security presentations from the corporate accounting scandals with which they are sometimes compared. That politicians and the media focus on some measures of Social Security solvency over others does not mean that the basic facts are hidden from view. In distinguishing the accounting appropriate for corporations and governments, it is also important to emphasize the distinction between accounting to inform lenders and accounting to inform taxpayers. With corporations, informing lenders (and potential stockholders) is central. Although lenders to government are also concerned with future obligations, the nature of the primary concern is different: it involves the willingness and ability to levy taxes to pay back debts. The fact that tax rates in any year are the same for workers in different cohorts increases the value of the traditional measure of actuarial balance relative to a cohort-based measure.

[3] *See* Jackson, *supra* note 1, at Part II.B.2.

> A reduction in replacement rates of about 5 percent beginning with individuals newly eligible in 2002 would solve 33 percent of the long-range financing problem. If a benefit formula change is delayed until 2020, the ultimate reduction in replacement rates would have to be larger, about 8 percent, to have the same impact on the long-range (75-year) actuarial deficit. By waiting until 2020, a larger reduction is needed because it applies for fewer years.[4]

Jackson's change in the maximum transition cost offers an alternative lens through which to quantify the cost of delay.

II. Reform and Accrued Benefits

Jackson links his measure with a normative suggestion for the structure of benefit reductions in a reform plan.[5] In particular, he argues that reforms should treat accrued benefits differently from benefits expected to accrue.[6] This conclusion seems problematic for several reasons. First, the level of "accrued" benefits can be defined in many different ways, as Jackson notes.[7] Requiring that a reform not affect accrued benefits would artificially elevate the importance of deciding which of these measures is the best one.[8] Second, even if agreement could be reached on the most accurate way of defining accrued benefits, the question would arise of whether changes in benefits would be fairer if reformers focus on the total level of benefits of different individuals or on the distinction between accrued and not-yet-accrued benefits. Paying attention to a reform's anticipated levels of benefits for different cohorts and the differences between those levels and previously scheduled benefits, along with the time lag until the change affects the actual payment of benefits, is a better focus for reformers.

One reason for preferring a focus on total benefits is that workers will likely pay more attention to that measure. A second reason has to do with fairness. Consider two individuals of the same age who end up with the same Average Indexed Monthly Earnings (AIME)[9] but different age-

---

[4] Soc. Sec. Advisory Bd., Why Action Should Be Taken Soon 17 (July 2001).

[5] Benefit reductions are measured, as they should be, relative to the benefits that would be generated by the current benefit formula.

[6] *See* Jackson, *supra* note 1, at Part II.B.4.

[7] *See* Jackson, *supra* note 1, at Part II.A.

[8] The recent uproar surrounding the transition from traditional defined benefit plans to cash balance plans in the private sector highlights the controversial nature of how "accrued" benefits are defined. *See, e.g.*, Mary Williams Walsh, *It May Be Time to Plumb Your Pension's Depths*, N.Y. Times, Dec. 15, 2002, at 8.

[9] For a description of the AIME calculation, see Staff of House Comm. on Ways and Means, 106th Cong., 2000 Green Book: Background Materials and Data on Programs Within the Jurisdiction of the Committee on Ways and Means 19–21 (Comm. Print 2000).

earnings patterns. One had relatively high earnings early in life, whereas the other had relatively high earnings later in life. A reform that continues to base benefits on AIME would treat the two the same. A reform that treated accrued benefits at the time of reform (somehow measured) differently from benefits based on later earnings would treat these two workers differently. Yet there seems no good reason to treat these two workers differently, assuming that AIME continues to be judged the appropriate basis for determining benefits.

Jackson argues that distinguishing accrued from not-yet-accrued benefits expands the opportunities for designing reforms.[10] That is unfortunately incorrect, and the practical effect could be precisely the opposite: Identifying benefits as "accrued" would undoubtedly accord those benefits a stronger degree of political protection against being reduced, thereby potentially limiting the degree of flexibility in designing a reform plan. In other words, a distinction between accrued and not-yet-accrued benefits changes the opportunities for reform, eliminating some options and adding others. The fundamental issue is therefore whether Congress would do a better job in allocating benefit reductions across different individuals with or without the constraint that accrued benefits be fully paid. The current approach is likely to work better because it relates directly to what beneficiaries really care about—the total of benefits to be received—and avoids the distinctions, as in the example above, that seem to have no good welfare basis. For example, would it have led to better legislation in 1983 to adjust the normal retirement age for part of benefits but not for all of benefits?[11]

Putting primary emphasis on the accrual measure that Jackson favors is thus potentially problematic for precisely the reason that he views as advantageous: that the natural tendency of highlighting that figure would be to encourage reform plans to distinguish between accrued and not-yet accrued benefits. To be clear, expanding the set of measures used to evaluate long-term solvency in Social Security would be beneficial. We do not, however, share Jackson's enthusiasm for making his accrual measure the central metric through which Social Security's financing and reform plans are viewed. The next Part discusses the role of solvency measures in the reform process and explains why the current seventy-five-year actuarial imbalance measure reflects a good balance among many competing demands.

---

[10] *See* Jackson, *supra* note 1, at Part III.B–D.

[11] *See generally* Social Security Amendments of 1983, Pub. L. No. 98-21, 97 Stat. 65 (1983).

III. Reform and Measures of Solvency

Currently, the measure that plays the largest role in Social Security public debate and reform proposals is the seventy-five-year actuarial deficit.[12] The Office of the Chief Actuary (OCACT) actually evaluates the system against many metrics, not only the seventy-five-year one. In particular, for OCACT to report that Social Security is in actuarial balance, its projection of the program's future financing must pass several tests, all of which must be met in order for the system to be deemed in actuarial balance. If short-run finances are insufficient to pay scheduled benefits, the system is clearly not in balance. Over longer periods, the reliability of the projections declines; as a result, if the projected long-run shortfall is modest enough, corrective action is not necessarily warranted and the system is deemed to be in balance.[13]

Noteworthy in this system is attention to *both* a present discounted value calculation and cash-flow considerations. It is very important that both cash-flow issues and present value issues be evaluated. Cash-flow concerns imply a presumption that Social Security will not be allowed to paper over its financial shortfalls by borrowing or by accounting gimmicks. For example, consideration only of a present value calculation would allow the use of distant (and politically implausible) large payroll tax increases to restore actuarial balance as part of a plan. There is of course no absolute protection against accounting gimmicks—they are always available, no matter what the accounting procedure. But gimmicks should be easily detectable, to lessen the likelihood and importance of their use.

The annual cash-flow projections and the current level of the Trust Fund are used to compute an "actuarial balance" figure. This measure reflects the degree to which the current Trust Fund and projected revenue over some period of time are sufficient to finance projected costs. When the projection shows insufficient resources to pay scheduled benefits, OCACT calculates the level of additional resources that would be sufficient to pay the benefits and also leave the Trust Fund with a projected balance equal to the following year's expenditures (a precautionary balance) at the end of the projection period.[14] The result is the "actu-

[12] This measure is equal to the present discounted value of annual cash flows over the next seventy-five years minus the existing Trust Fund plus the cost of ensuring the Trust Fund at the end of the seventy-fifth year is equal to the projected expenditures in the seventy-sixth year. *See* Peter A. Diamond & Peter R. Orszag, Saving Social Security: A Balanced Approach 30–34 (2004).

[13] Reports are made for Old-Age Survivors Insurance (OASI) and Disability Insurance (DI) programs separately as well as for the two together. We focus only on the latter. *See generally* 2003 Bd. of Trs. of the Fed. Old-Age and Surviviors Ins. and Disability Ins. Trust Funds Ann. Rep., H.R. Doc. No. 108-49, at 44–50, *available at* http://www.ssa.gov/OACT/TR/TRO3/tr03.index.html [hereinafter 2003 Trustees Report].

[14]

An actuarial balance of zero for any period would indicate that estimated cost for

arial deficit" or "actuarial imbalance"; the seventy-five-year actuarial imbalance is the single most visible number for evaluating Social Security's finances.

The actuarial imbalance is traditionally presented relative to taxable payroll.[15] That is, OCACT reports the actuarial imbalance as a share of taxable payroll over the next seventy-five years. One interpretation of this figure is that it represents the payroll tax increase that would be sufficient to finance benefits over the seventy-five-year horizon provided the increase began the following year and remained in force for the full seventy-five-year period.[16]

Jackson would replace the seventy-five-year actuarial balance (as the central measure in Social Security reform discussions) with a measure reflecting the closed-group deficit for those over fifteen years of age.[17] In other words, he would shift the discussion of the deficit in Social Security from an annual basis (i.e., seventy-five years) to a cohort basis (i.e., those over fifteen years of age).

Note that Jackson seems to share the view, inherent in the seventy-five-year actuarial imbalance measure, that an infinite horizon calculation should not be the focus of public discussion:[18] His exclusion of those fifteen years of age and younger reflects a view that some form of truncation is appropriate. A truncated measure is appropriate for the central measure of Social Security's financial condition, but the current measure is more attractive than Jackson's alternative. Because neither he nor we prefer an infinite horizon measure, the appendix provides arguments for preferring the seventy-five-year horizon to an infinite horizon.

The seventy-five-year measure not only defines the size of the deficit, but also provides a natural and understandable goal for reform: seventy-five-year balance.[19] A shift to the closed group measure that Jackson favors raises the question of what target for reform would result. As Jackson recognizes, embracing the closed-group measure does not imply

---

> the period could be met, on average, with a remaining trust fund balance at the end of the period equal to 100 percent of the following year's cost. A negative actuarial balance indicates that, over the next 75 years, the present value of income to the program plus the existing trust fund falls short of the present value of the cost of the program plus the cost of reaching a target trust fund balance of one year's cost by the end of the period.

Id. at 58.

[15] *See id.* at 45. (explaining "basic to the consideration of the long-range actuarial status of the trust funds are the concepts of income rate and cost rate, each of which is expressed as a percentage of taxable payroll").

[16] Technically, the payroll tax increase must be large enough to leave a Trust Fund equal to the following year's expenditures at the end of the projection period.

[17] *See* Jackson, *supra* note 1, at Part II.C.5.

[18] *See id.* at Part I.A.1.b.

[19] Those preferring a larger Trust Fund in order to encourage more national savings could argue for a target Trust Fund ratio in excess of one as part of the definition of "actuarial balance."

embracing a goal of zero balance for the closed-group deficit.[20] Eliminating the closed group deficit would in effect imply full funding.

Full funding is not an appropriate goal for reform at this point because it would impose an unfairly large burden on transition generations. Social Security was very generous to earlier cohorts.[21] The cost of that generosity means that the Trust Fund is smaller than it would have been if we had been less generous to earlier cohorts. The difference between what the Trust Fund would have been in the absence of allowing earlier cohorts to enjoy above-market returns on Social Security, and what the Trust Fund is today, reflects what we call Social Security's "legacy debt" from its past.[22] That legacy debt should, arguably, be spread over all future generations. Moving instead to full funding by some deadline would impose the full cost of financing the legacy debt on a limited number of generations, while sparing all future cohorts from sharing in that cost. In our view, it is difficult to justify imposing such a constraint on reform.

Jackson shares this view, arguing that some degree of partial funding is an appropriate choice for Congress.[23] But then what is the target for the closed-group balance and how is such a target set? The political process does not seem particularly adept at setting complex targets. Jackson considers the possibility of stabilizing the ratio of the implicit Social Security debt to Gross Domestic Product ("GDP"), which has some analytical attraction.[24] But stabilizing the ratio of public debt to GDP is a similarly attractive goal for the unified budget, and it has never caught hold with the public despite the fact that many members of the public appear to be concerned about the nation's fiscal imbalances. One drawback to using Jackson's proposed measure as the most prominent metric of Social Security finances is thus that it does not lend itself to a reasonable, natural target for reform. It may therefore dissipate political attention into arguing over the right target rather than agreeing to the steps for reaching that target. In particular, if an accrual measure were adopted, shifting the target for a reform might well prove to be politically easier than addressing a given shortfall through a combination of tax increases and benefit reductions.

In addition, it is important to note that both the closed-group and the seventy-five-year solvency measures potentially distort reform ideas—albeit in different ways—relative to some complete measure of their effects. Much of the problem arises from the fact that we tend to think of, and act on, benefit rules on a cohort basis but think of, and act on, tax rates and

---

[20] *See* Jackson, *supra* note 1, at Part III.A.3.

[21] *See, e.g.*, DEAN R. LEIMER, COHORT-SPECIFIC MEASURES OF LIFETIME NET SOCIAL SECURITY TRANSFERS 16 (Soc. Sec. Admin. ORES, Working Paper No. 59, Feb. 1994).

[22] *See* DIAMOND, *supra* note 12, at 37–39, 88–93, 182–83, 197 (discussing the legacy debt concept in the design of Social Security reform).

[23] *See* Jackson, *supra* note 1, at introduction.

[24] *See id.* at Part III.A.3.

taxable earnings levels on an annual basis.[25] Given this disjunction, a single measure cannot do justice to both concerns. For example, using a seventy-five-year actuarial balance measure makes an increase in the maximum taxable earnings base (that is, the earnings base subject to tax) appear to contribute more to financial restraint than it really does, because implied benefit increases beyond the seventy-five-year horizon are not included in the measure. On the other hand, consider a hypothetical system in which cash shortfalls begin in ten years, and assume an increase in the tax rate starting at that time. The tax increase generates revenue in years with a projected cash shortfall, and part of that revenue is excluded from the closed group measure since workers now younger than fifteen will be paying part of the future taxes. The relative distortions caused by the different metrics will depend on the specific projections and the nature of the reforms involved.

IV. Accrual Accounting and the Federal Budget

The main thrust of Jackson's article is a proposed modification to the focus of accounting for Social Security reform. The more radical portion of his analysis, though, involves also shifting the accounting system for Social Security in the context of the federal budget. Even for budgetary purposes, he considers replacing annual cash-flow accounting for Social Security with annual accrual accounting.[26] He argues that this change will cause policymakers to enact higher taxes or lower spending, or both, with this move desirable because accrual accounting will give a better picture of the change in long-run budgetary pressures.[27] Jackson argues that politicians are more likely to focus on the unified budget than on the non-Social Security budget when Social Security is in surplus and on the non-

---

[25] For example, the 1983 reforms changed the normal retirement age on a cohort basis and accelerated the year of scheduled tax rate increases (i.e., implemented changes on an annual basis). *See generally* Social Security Amendments of 1983, Pub. L. No. 98-21, 97 Stat. 65 (1983). It is noteworthy that Sweden has adopted an automatic adjustment mechanism to minimize (and possibly eliminate) the chance of running out of money. *See* Diamond, *supra* note 12, at 82–83. Similarly, one could think of a pure pay-as-you-go system that calculated relative benefit claims based on individual relative earnings histories and then set the overall benefit level to match the revenues available precisely. This system would be organized on an annual basis. But this would be obscured by cohort-based accrual accounting.

[26] *See* Jackson, *supra* note 1, at Part II.D.

[27] It is worth noting that for some issues, the unified cash budget is a better simple guide than the other measures under consideration. As noted by Jackson, the most obvious example is with regard to Treasury sales of bonds to the public. *See* Jackson, *supra* note 1, at Part I.B.2. More importantly, the unified budget is also critical for considering short-run macroeconomic policy. While Social Security analysts appropriately pay attention to long-run issues, many individual consumers appear to pay far more attention to cash flows, making the unified budget measure an important one, even if it does not accurately capture longer-run fiscal imbalances.

Social Security budget when Social Security is in deficit.[28] The focus of this Comment is on the Social Security measure, not the distinction between unified and other budget measures.

Although there is always a case for pragmatism, it is worth noting that the current divergence between the cash-accounting and accrual-accounting measures could readily go the other way in other times. If the goal is simply to apply pressure for fiscal discipline, and even assuming that increases in the perceived deficit in Social Security would raise such pressure, it is not clear that a permanent change in accounting is necessarily beneficial. In particular, if a well-designed Social Security reform plan that restores long-term solvency to Social Security is enacted, it is likely to be the case that cash-flow deficits will occur as the government draws down the Trust Fund that was accumulated precisely to help finance the retirement of the baby-boomers. The accrual measure that could accompany such a reform may, however, show different results, as benefits are paid to the already retired and new cohorts are accruing benefits more slowly as a result of reform. Our choice of budget concept should not depend solely on how it would influence policymakers in the current environment.

Typically, cash flows give a better picture of short-run macroeconomic effects, whereas accrual measures give a better picture of the change in long-run positions. Thus there is an important role for both. In terms of informing the public, a case could be made that the public is already fairly aware of short-run macroeconomic issues and it is therefore more important to focus on long-run issues, making accrual accounting look attractive. Such a case would be more persuasive if accrual accounting also seemed reasonable from other perspectives, but that is not the situation.

Jackson correctly notes that the unified budget already incorporates a very limited amount of accrual accounting.[29] The modest size of those budgetary categories, however, means that many of the difficult issues involved in accrual accounting do not rise to the level of generating political discussion and disagreement. Employing accrual accounting with regard to Social Security would qualitatively be a different matter. Furthermore, it is difficult to see why accrual accounting should be undertaken for Social Security, and not also for Medicare and Medicaid.[30]

---

[28] *See* Jackson, *supra* note 1, at Part I.B.2.

[29] *See id.* at Part IV.A.1.

[30] To highlight the difficulty, note that no one expects defense expenditures ever to end. One could thus consider accrual accounting for defense spending based on measures of international risk. To be sure, one distinction is that the other programs are already legislated to be permanent, whereas defense spending is appropriated each year through specific expenditure legislation. One could also draw the distinction between Social Security, which is financed fully by dedicated revenues, and other programs that are not. But this distinction is more closely related to separating budgets than the type of accounting in each budget.

Given the difficulty of drawing the line between Social Security and other programs, it is important to evaluate the effects of accrual accounting under the assumption that a substantial share of the federal budget were accounted for in this manner. That perspective highlights two principal problems.

First, accrual measures are easy to manipulate.[31] Accrual measures—like other long-term projections, including the seventy-five-year actuarial imbalance—are extremely sensitive to assumptions about the future. Seemingly small and reasonable changes in assumptions about the future manifest themselves as large changes in accruals. If these measures were the focal point for evaluating the budget, there would inevitably be pressures in some budgetary conditions to shade the projections in a manner that makes the outlook look better. That pressure could then contaminate the process by which OCACT makes its projections; OCACT's reputation as an "honest broker" is extremely valuable and the loss of that reputation would be very costly to the prospects for Social Security reform.[32] The broader point is that the more sensitive a measure is to reasonable variations in assumed parameters, the more important it is to provide a range of estimates rather than just a point estimate. And the budget process needs a point estimate to function.[33]

Second, just as accrual accounting for Social Security fails to provide a natural, reasonable target, accrual accounting in the unified budget does not provide an easily understandable target that is also a sensible one. Preserving the ratio of implicit liabilities to GDP may be sensible, but it is unlikely to be understandable as the primary goal of budgetary policy. Lacking a natural target for balance could undercut the pressure for balancing any given measure, because it is typically easier to shift the target than to enact the changes necessary to reach a given target.

---

[31] Cash-flow measures are admittedly not immune to manipulation, given the ability to move cash flows between budget years, but they are relatively difficult to manipulate compared to accrual measures.

[32] The quality of Social Security projections in the United States is much higher than in many other countries, such as Germany. In most other countries, there is no counterpart to the relative independence of the OCACT, and it is important to preserve this extremely valuable independence.

[33] This concern is similar to the one that surrounds the use of dynamic scoring in evaluating budgetary proposals. Dynamic scoring incorporates the macroeconomic feedback effects from proposals, and is very sensitive to small changes in the assumptions. It is therefore reasonable to undertake dynamic analyses in which a range of estimates can be provided, but much less reasonable to undertake dynamic scoring in which a single estimate must be provided. *See, e.g.*, *Original Jurisdiction Hearing: Hearing on the President's "Freedom to Manage" Initiative Before the Subcomm. on Legislation and Budget Process of the House Rules Comm.*, 107th Cong. 8–9 (2002) (statement of Peter R. Orszag, Senior Fellow, Brookings Inst.).

V. Conclusion

Annual accrual accounting for Social Security would be a useful addition to the information provided to the public, but it should not become the central number used in policy discussions. The issue is not whether accrual information would be helpful—it would be. Instead, the issue is whether recasting the political process to concentrate primarily on accrual measures rather than the current measures would be more likely to generate sound decision-making. Jackson argues that the current accounting system impedes reform by comforting policymakers that the problems facing Social Security are decades away. But accrual accounting would not change the basic human tendency, which also manifests itself among policymakers, to ignore or delay addressing long-term problems. Furthermore, accrual accounting measures are extremely sensitive to the assumptions employed. To the extent they were granted a much larger role in the federal budget, they would likely induce pressures that could ultimately undermine the perceived legitimacy of the actuarial and budgetary projections.

APPENDIX: SEVENTY-FIVE-YEAR HORIZON

The traditional seventy-five-year metric for achieving long-term balance in Social Security has recently been under attack from opposite directions: some would like to shorten the period to fewer than seventy-five years, and others would like to extend it well beyond seventy-five years. Ideally, the projection period should be long enough to make sure Congress is not slighting future concerns, but not so long that it is addressing problems that are very unlikely to develop as projected. It seems to us that a seventy-five-year projection period is a good compromise between these two concerns.

Plans that fail to achieve seventy-five-year actuarial balance may or may not include beneficial changes to the program, but they do not put Social Security on a firm long-term financial footing. Reforming Social Security is a difficult and complicated task. Policymakers are unlikely to be willing to undertake substantial reforms to the program every year or even every few years. It is therefore important that, when reform does occur, it put the system on a sound basis for an extended period of time. Furthermore, workers deserve to have some sense of the program's structure in the future, so that they can plan for their own retirement and other financial needs. The seventy-five-year horizon provides a reasonable approximation of the lifespan of workers just entering the labor force: A twenty-year-old worker today, for example, has less than a five percent chance of living past ninety-five.[34] Failing to achieve seventy-five-year solvency would leave too much uncertainty about the future course of the program. But extending the central measure of actuarial balance beyond seventy-five years also has downsides.

A final reason for continuing to support the seventy-five-year measure is that despite recent criticisms, this traditional measure enjoys a surprising degree of bi-partisan support among policymakers. Such bipartisan consensus is rare in Social Security reform. Even if the optimal projection period were slightly shorter than seventy-five years or slightly longer than seventy-five years, any gains from changing it may not be worth opening up the Pandora's box of precisely what that projection period should be.

The imbalance over the indefinite future does provide useful information, but a key question is what horizon policymakers should adopt for the purpose of evaluating reform. Despite some theoretical attraction of the indefinite future or "permanent" imbalance measure, in practice using that metric as the primary measure of Social Security's deficit would prove problematic. The legislative process invariably places heavy weight on certain summary statistics, such as actuarial balance. If that balance is

---

[34] *See* DIAMOND, *supra* note 12, at 33.

defined over an indefinite future, an excessive degree of reliance could be placed on the harder-to-predict, more-distant future. Focusing exclusively on a very long horizon can also make it easier to employ gimmicks: A massive tax rate increase scheduled for the years 2150 and beyond, for example, could technically address the permanent imbalance in Social Security but would be widely and correctly seen as having no credibility.[35]

To be sure, longer projection horizons are beneficial for some purposes. But calculations with a longer horizon can alert us to serious drawbacks in proposals without playing a central role in the constraints on the legislative process. Besides, there is no reason to try to fix Social Security forever—future Congresses can and should respond to changing circumstances, and the legislative process should give far more weight to concerns thirty-five years from now than 135 years from now. At the same time, we do not want to put in place a system that we expect will be out of seventy-five-year actuarial balance very shortly.

---

[35] Such a proposal might restore balance to the long-term budget of Social Security but would not pass the "Test of Long-Range Actuarial Balance," which compares the balance over the sub-periods from the present out to the seventy-five-year horizon to an allowable deviation from balance. Thus actuarial balance over seventy-five years (or solvency over seventy-five years) is neither necessary nor sufficient for this measure.

COMMENT

ACCOUNTING FOR THE FEDERAL BUDGET AND ITS REFORM

ELIZABETH GARRETT*

Social Security, the largest entitlement program in the federal government, is now accounted for on a cash-flow basis; its financial reports subtract the amount of benefits paid to retirees from the annual cash receipts from payroll taxes paid by workers. In cash-flow terms, the system is in surplus because it currently takes in billions of dollars more each year than it pays out.[1] But the cash-flow method obscures the serious challenges that Social Security faces in the future as the Baby Boom generation continues to retire and fewer workers pay the taxes necessary to sustain this pay-as-you-go system. As Professor Jackson reveals, the unfunded accrued liabilities in the Social Security trust funds at the end of fiscal year 2002 were $12.6 trillion, or 122% of the country's gross domestic product (GDP).[2]

In *Accounting for Social Security and Its Reform*, Jackson makes a persuasive case for using a form of accrual accounting to more accurately present the financial condition of the Social Security system. He argues that moving away from a cash-flow depiction and including various accrual-based presentations would change the terms of the political debate about this entitlement program.[3] It would improve accountability of lawmakers to voters for decisions they make, or avoid making, about Social Security. It would heighten the visibility of the long- and medium-term challenges faced by the retirement program. Framing the system and proposals to reform it in accrual accounting terms would alter interest-group activity, provide salience to certain matters now hidden by cash-flow presentations, and allow a serious discussion of reform proposals that cannot now pass political muster because budget rules are keyed to cash-flow measures. Jackson describes these as positive effects that flow from

---

* Professor of Law, University of Southern California. Director, USC-Caltech Center for the Study of Law and Politics. J.D., University of Virginia, 1988; B.A., University of Oklahoma, 1985. I appreciate comments from Ellen Aprill, Andrei Marmor, and Ed McCaffery; research help from Jessica Wimer of the USC Law Library; and the research assistance of Colin McNary (Chicago '04) and Tracy Daub (USC '05).

[1] *See* Howell E. Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

[2] *See* Jackson, *supra* note 1, at introduction.

[3] *See, e.g.*, *id.* at Part III.B.

"altering the optics of reform proposals."[4] Importantly, he argues that altering the optics affects more than the way politicians and citizens perceive reform proposals. Because of budget rules and other political realities, adopting different accounting methods may actually change political outcomes.

Jackson's article highlights the importance of accounting methods for fiscal policy debates and decisions. As one commentator has observed, "accurate accounting has a practical purpose: to reveal the consequences of current practices and to clarify the nature of the choices we face."[5] Determining the appropriate accounting lens through which to view fiscal policy and reform proposals is an increasingly active area of scholarship. In addition to Jackson's accrual-based proposal, scholars have proposed "generational accounting" measures designed to reflect the burden of today's policies on future generations.[6] Jagadeesh Gokhale and Kent Smetters, for example, have argued that the federal budget agencies should provide measures of fiscal imbalance and generational imbalance to provide information that is more forward-looking than current cash-flow figures.[7] Moreover, some federal programs are already depicted through accrual-based systems. Since 1990, the federal budget has used accrual principles to account for federal loan and credit-guarantee programs because, given the structure of these programs, cash-flow accounting vastly minimized the extent of the government's financial obligations.[8]

Jackson's proposal is sensible. Current budget accounting practices cannot provide an accurate picture of the fiscal condition of Social Security. Accordingly, the political debate is warped, important reforms are not properly considered, and a lack of transparency undermines accountability of legislators to their constituents. Social Security is not, however, the only part of the budget where current accounting conventions are in-

---

[4] *Id.* at Part III.C.

[5] Christopher DeMuth, *Foreword* to JAGADEESH GOKHALE & KENT SMETTERS, FISCAL AND GENERATIONAL IMBALANCES: NEW BUDGET MEASURES FOR NEW BUDGET PRIORITIES at vi (2003).

[6] *See, e.g.*, Alan J. Auerbach, Jagadeesh Gokhale & Laurence J. Kotlikoff, *Social Security and Medicare Policy from the Perspective of Generational Accounting*, *in* 6 TAX POLICY AND THE ECONOMY 129 (J.M. Poterba ed., 1992); LAURENCE J. KOTLIKOFF, GENERATIONAL ACCOUNTING: KNOWING WHO PAYS, AND WHEN, FOR WHAT WE SPEND (1992). For assessments of generational accounting, see DANIEL SHAVIRO, DO DEFICITS MATTER? 151–85 (1997); Peter Diamond, *Generational Accounts and Generational Balance: An Assessment*, 49 NAT'L TAX J. 597 (1996). One of the early Clinton budgets provided generational accounting tables as an "alternative budget presentation." *See* OFFICE OF MGMT. AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT FISCAL YEAR 1993, at 3-7 to 3-13 (1992). That decision was a political one, not required by budget laws, and the presentations have not been repeated in subsequent budgets.

[7] *See* GOKHALE & SMETTERS, *supra* note 5, at 2, 21–22.

[8] *See generally* 2003 OFFICE OF MGMT. AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT FISCAL YEAR 2004, at 189–248 [hereinafter 2004 BUDGET] (discussing current federal credit and insurance programs); James L. Chan, *The Bases of Accounting for Budgeting and Financial Reporting*, *in* THE HANDBOOK OF GOVERNMENT BUDGETING 357, at 374–79 (R. T. Meyers ed., 1999) (discussing accrual-based accounting as a budget tool).

adequate or outright misleading. The fiscal situation of the entire federal budget is primarily depicted through cash-flow measures projected for a limited period of time into the future. The budget deficit is the amount by which government outlays exceed government receipts in one fiscal year. Although many policies have substantial long-term fiscal ramifications, these consequences are not reflected in any of the deficit numbers. Instead, budget experts total all government expenditures and receipts over twelve months and compare the two numbers. Deficit projections are merely estimates of these cash-flow figures into the relatively near future, and they are the measure primarily used in arguments about the country's fiscal health.

In particular, current accounting practices obscure the true cost of legislation providing tax benefits for various groups of taxpayers. Revenue estimates provided for new tax expenditures provide cash-flow effects for only a few years into the future. Depending on the program and the current budget rules, revenue estimates may be limited to one year, five years, or at the most ten years. Changing accounting conventions so that present-value estimates of programs with substantial effects in the future are included in revenue projections provided during congressional consideration of tax bills would provide better information about the long-term financial obligations entailed in particular policies. If such projections were tied to budget enforcement rules in any new budget framework enacted by Congress, the new information would do more than change the terms of the debate. Just as Jackson argues with respect to accounting for Social Security,[9] the additional information could have real effects on legislative outcomes.

The two major tax bills passed with the strong support of the current Bush administration, the Economic Growth and Tax Relief Reconciliation Act of 2001 ("EGTRRA")[10] and the Jobs and Growth Tax Relief Reconciliation Act of 2003 ("JGTRRA"),[11] are stark examples of how inadequate methods of measuring revenue loss allow Congress to minimize the long-term fiscal impact of tax provisions that will result in substantial future revenue loss. Of course, the Bush administration and recent Congresses did not invent these gimmicks. They merely took advantage of budget ploys long used by politicians to send tax benefits to groups that support them without appearing to worsen the fiscal position of the federal government. A brief discussion of three gimmicks will give a flavor of the tricks made possible by current accounting practices.[12]

---

[9] *See, e.g.*, Jackson, *supra* note 1, at Part III.B.

[10] Economic Growth and Tax Relief Reconciliation Act of 2001, Pub. L. No. 107-16, 115 Stat. 37.

[11] Jobs and Growth Tax Relief Reconciliation Act of 2003, Pub. L. No. 108-27, 117 Stat. 764.

[12] This discussion is partly an expansion of arguments made in Elizabeth Garrett, *Budget Magic Tricks*, THE WORLD & I, July 2003, at 54.

I. BACK-LOADED REVENUE LOSS

Tax provisions can be drafted so that most of the revenue loss occurs in the future, outside the budget window of one, five or ten years. Federal budgeting rules have often required that cash-flow revenue effects be provided not only for the current fiscal year, but also for several years in the future.[13] The longer budget window was chosen because in the early years of the congressional budget process, deficit targets were met by shifting revenue loss into the next fiscal year or by speeding up revenue gain into the current fiscal year.[14] A ten-year budget window has been used in recent years in order to reduce the ability to use this strategem. At the end of fiscal year 2002, however, many of the budget rules expired and were not replaced by Congress, and the President and federal lawmakers have reverted back to shorter time horizons.[15]

Even the ten-year budget window did not eliminate the ability of lawmakers to back-load revenue loss, although it could make tax benefits less desirable for interest groups that would have to wait several years to enjoy their tax expenditures and who would therefore worry that Congress might repeal or reduce them before they were fully effective. Nonetheless, delay does not eliminate the demand for tax benefits, and sophisticated legislative drafters can meet that demand and avoid fiscal discipline by constructing tax benefits in various forms that are "equivalent tax reductions in terms of their value, but that involve quite different timing of the tax reductions and therefore have very different impacts on annual revenue estimates during a budget period."[16] The shorter the budget window, the easier the chicanery.

An example of a back-loaded tax provision is the Roth IRA,[17] a type of individual retirement account first enacted in the late 1990s and supported strongly by the current Bush administration as a candidate for expansion in its retirement account and pension reform proposals. The Roth IRA differs from a traditional IRA because contributions are not deductible when they are made, thereby avoiding current revenue loss.[18] Instead, withdrawals are tax-exempt,[19] postponing revenue loss far into the future when today's holders of Roth IRAs begin to retire and consume with

---

[13] *See* Philip G. Joyce, *Congressional Budget Reform: The Unanticipated Implications for Federal Policy-Making*, 56 PUB. ADMIN. REV. 317, 318 (1996).

[14] *See id.*

[15] For example, the Senate extended its PAYGO rule only until April 15, 2003. *See* Bill Heriff, Jr., Budget Enforcement Procedures: Senate Pay-As-You-Go (PAYGO) Rule, CRS Report for Congress (Oct. 22, 2002). The President's Fiscal Year 2004 Budget relied on a five-year budget window. *See, e.g.*, 2004 BUDGET, *supra* note 8, at 28, 311 tbl.5-1, 312 tbl.5-2.

[16] Michael J. Graetz, *Paint-By-Numbers Tax Lawmaking*, 95 COLUM. L. REV. 609, 673 (1995).

[17] *See* 26 U.S.C. § 408 (2000).

[18] *See id.* § 408A(c)(1).

[19] *See id.* § 408A(d)(1)(A).

their savings.[20] When enacted in 1997, the Roth IRA provision was estimated by the Joint Committee on Taxation to lose only $1.8 billion during the first five years and to lose more than $20.2 billion over ten years.[21] Taking a lesson from this successful use of a timing trick, the Bush administration's proposal to expand and consolidate the tax rules for tax-advantaged retirement plans in the Fiscal Year 2004 budget was constructed so that most of the revenue loss would fall well outside any budget window.[22] Leonard Burman, William Gale, and Peter Orszag estimate that, after twenty-five years, the revenue loss of the proposal would be 0.5% of GDP per year, the equivalent today of more than $50 billion a year.[23] Although this proposal has not been adopted in its entirety, it provides some perspective on the magnitude of revenue loss that can be hidden effectively by postponing it.

Accounting methods should make substantial future revenue loss salient to lawmakers and constituents during any debate on adopting such provisions. The President's Budget documents provide present-value estimates of future revenue loss for selected tax expenditures, including those that allow tax-advantaged saving for retirement,[24] but these figures are buried in obscure budget documents that few read, let alone understand. Think tanks and other interest groups have worked to publicize the information; for example, the Urban Institute-Brookings Institution Tax Policy Center developed and disseminated projections of long-term revenue loss from Bush's retirement savings proposal.[25] But budget rules need to be changed to ensure that the government budget offices also provide this kind of information and that it is considered during the committee mark-ups and floor deliberation.

New budget enforcement rules, such as points of order that can be waived only with a roll call vote and sometimes only with a supermajority vote in the Senate, could be keyed to proposals that substantially affect the present value of future revenue loss. For example, a point of order could be triggered by tax provisions projected on a present-value ba-

---

[20] *See* EDWARD J. MCCAFFERY, FAIR NOT FLAT: HOW TO MAKE THE TAX SYSTEM BETTER AND SIMPLER 50 (2002) (suggesting that back-loaded IRAs may be "deal[s] with the devil").

[21] JOINT COMMITTEE ON TAXATION, ESTIMATED BUDGET EFFECTS OF THE CONFERENCE AGREEMENT ON REVENUE PROVISIONS OF H.R. 2014, THE "TAXPAYER RELIEF ACT OF 1997" 2 (1997).

[22] President Bush's FY 2004 proposal would unify most tax-preferred accounts, including all IRAs, into two tax vehicles, retirement savings accounts and lifetime savings accounts. *See* DEP'T OF TREAS., GENERAL EXPLANATIONS OF THE ADMINISTRATION'S FISCAL YEAR 2004 REVENUE PROPOSALS 121–24 (2003) [hereinafter Treasury Explanations of FY 2004 Proposal]; *see also* CONGRESSIONAL BUDGET OFFICE, AN ANALYSIS OF THE PRESIDENT'S BUDGETARY PROPOSALS FOR FISCAL YEAR 2004, 9–10 (2003) (discussing proposals).

[23] Leonard E. Burman, William G. Gale & Peter R. Orszag, *The Administration's Savings Proposals: Preliminary Analysis*, 98 TAX NOTES 1423, 1423 (2003).

[24] *See, e.g.*, 2004 BUDGET, *supra* note 8, at 112 tbl.6-4.

[25] *See* Burman et al., *supra* note 23.

sis to cost the government revenue more than a threshold amount,[26] or a point of order could be triggered when the information about substantial revenue loss in the future did not accompany any committee report on the tax bill. Budget enforcement rules can help to ensure that revenue estimates like these affect congressional decision-making and legislative outcomes. When the budget rules are changed to track new accounting conventions, the shift in accounting has more than mere optical consequences. It will have real political effects as well, changing how Congress legislates.

In addition to future revenue losses, future revenue gains are also important. For example, traditional IRAs yield increased tax revenues when retired people pay tax on withdrawals of previously deducted amounts. The magnitude of these future tax payments is currently the subject of controversy.[27] Whatever the outcome of this particular debate, projecting a potential revenue stream is important for a realistic sense of tax policy, just as projections of long-term revenue loss are crucial for decision-making. It is also necessary to determine the fiscal effect of any future legislation that would change the ability of the government to collect the money included in projections of future revenue. For example, proposals to reduce or eliminate the tax on withdrawals from IRAs could have a substantial negative impact on the country's future fiscal health.

II. Revenue Speed-Ups

Just as revenue loss can be postponed until it is outside the budget window, revenue gains can be sped up so that they occur within the budget window and offset other revenue loss. These speed-ups usually do not represent new revenue for the government—they are provisions that ensure that the expected revenue comes in earlier. For example, Congress has increased the payments required to meet safe-harbor provisions for estimated tax payments and changed tax payment deadlines.[28] One prob-

---

[26] This proposal is a variation of the current Byrd Rule. *See infra* text accompanying note 38. The Byrd Rule looks at cash-flow effects outside the budget window of a particular bill. The proposal in the text would be based on present-value computations of long-term revenue effects.

[27] *See* Alan J. Auerbach, William G. Gale & Peter R. Orszag, *Reassessing the Fiscal Gap: The Role of Tax Deferred Saving*, 100 Tax Notes 567 (2003) (discussing preliminary study by Michael J. Boskin that projected tax-deferred savings will generate net revenue with a present value of $12 trillion dollars through 2040, criticizing this study, and finding revenue from tax-deferred savings to cause only a minor change in the fiscal gap projections but noting that Boskin's work "impl[ies] that proposals to reduce the taxation of withdrawals from retirement accounts could significantly and adversely affect an already bleak fiscal outlook"). *See also id.* at n.3 (noting that Boskin is revising this preliminary study in a way that is likely to reduce net present value of revenues); Al Kamen, *A Few Too Many Zeros*, Wash. Post, July 23, 2003, at A21 (providing a description of an e-mail from Boskin on initial paper and plan for revisions).

[28] *See* Martin A. Sullivan, *2002: A Budget Odyssey*, 76 Tax Notes 872, at 872–73 (1997).

lem with this strategy is that revenue from speed-ups is used to offset early revenue loss of tax benefits that are enacted permanently—and thus represent a long-term drain on the Treasury. Speeding up revenue is a short-term offset only.

Such timing shenanigans are not limited to tax proposals, although the durability of most tax benefit provisions[29] means this can be a particularly acute problem in the tax. Governments can take advantage of cash-flow accounting by selling government assets, thereby receiving a large lump sum payment in one year, and then leasing those same assets, thereby spreading costs over future years. Indeed, Illinois took advantage of this trick without actually selling the building. It put up for sale the James R. Thompson Center, the state office building in Chicago, a memorable structure designed by Helmut Jahn and seen in movies like *Running Scared* and *The Negotiator*. Although no one had offered to buy the office building, the state claimed a $200 million savings on paper to help it balance its budget in a time of fiscal crisis.[30] New York so frequently accelerates tax payments into a current fiscal year that budget experts have coined a term for such accelerations: spin-ups.[31]

Some tax provisions take advantage of both the first and second deceptions, obscuring the actual revenue cost by both triggering accelerated collections in early years and pushing substantial loss into out-years. The Bush administration's pension reform proposal would have increased revenue by $15 billion in the first few years as taxpayers were encouraged to switch from retirement funds allowing immediate deductions to the new accounts providing tax savings later.[32] This is just a recent example of the use of both timing tricks. The capital gains tax reduction in the 1997 Taxpayer Relief Act[33] was projected to raise money in the first five years after enactment, as lower tax rates spurred more asset sales, but then to result in a revenue loss of over $21 billion from 2003 to 2007.[34] Many states have dealt with unprecedented budget deficits by speeding up the receipt of revenue expected to be collected as a result of a settlement of a lawsuit brought against tobacco companies. They have accomplished this by selling bonds, called Tobacco Securitization Bonds in

---

[29] Recent tax bills, with their plethora of expiring tax benefit provisions, are a departure from past practice in this respect. *See infra* text accompanying notes 40–48 (discussing the expiring tax provisions ploy).

[30] David E. Rosenbaum, *States Balance Budgets with Blue Smoke and Mirrors*, N.Y. TIMES, Aug. 24, 2003, § 4, at 4.

[31] *See* Richard Briffault, *Balancing Acts: The Reality Behind State Balanced Budget Amendments*, *in* THE BALANCED BUDGET DEBATE SERIES 20 (1996).

[32] *See* Burman et al., *supra* note 23, at 1433 tbl.6 (providing revenue estimates); *id.* at 1434 (explaining why conversion feature of such plans is a budget gimmick costing the government in the long run though speeding up revenue collection in the short-run).

[33] Taxpayer Relief Act of 1997, Pub. L. No. 105-34, 111 Stat. 788 (codified as amended in scattered sections of 5 U.S.C., 19 U.S.C., 26 U.S.C., 29 U.S.C., 31 U.S.C., 42 U.S.C., and 46 app.).

[34] *See* JOINT COMMITTEE ON TAXATION, *supra* note 21, at 516.

California, to collect much of the stream of expected income now, planning to pay off debtholders with the settlement money as it comes in over the next decades.[35] Budget accounting rules should be constructed so that speeding up receipt of revenue cannot mask decisions that will cost taxpayers more in future years. Again, the cash-flow feature of much budget accounting facilitates this smoke-and-mirrors game.

III. Expiring Provisions

Budget rules have been changed over the years to reduce the ability of lawmakers to play timing games. Budget windows were extended to five and then ten years because of egregious trickery such as moving a pay date one day from one fiscal year to the next.[36] Congress, worried that legislative drafters were pushing substantial revenue loss outside even a ten-year budget window, adopted a rule, called the Byrd Rule after its author Robert Byrd (D-W. Va.), that allows senators to object to consideration of a reconciliation proposal that would increase the deficit in fiscal years beyond those covered in the measure.[37] Most tax bills are passed as omnibus budget reconciliation bills, so they are subject to the Byrd Rule. Thus, if their provisions would increase the deficit in the out-years, a member can raise a point of order on the ground that such provisions are "extraneous" (the terminology of the Byrd Rule) to the reconciliation process. To overcome the point of order and continue deliberation, sixty senators must vote to waive the objection.[38]

One way to avoid invocation of the Byrd Rule, but still pass substantial tax breaks, is to sunset the revenue-losing provisions so that they expire at the end of the budget window. Provisions that are no longer effective cannot be responsible for revenue loss in out-years under current budget accounting. Before 2001, Congress had enacted a few tax provisions on a temporary basis, called "extenders" because lawmakers invariably extended them rather than allowing them to expire.[39] One purpose for the temporary quality of these provisions was to keep their revenue costs down and thereby to make it easier to find revenue offsets in a world governed by pay-as-you-go rules.[40] The size of these expiring provisions increased by several orders, however, with the two recent major

---

[35] *See, e.g.*, Evan Halper & Thomas S. Mulligan, *Budget Plan Risky, Officials Warn*, L.A. Times, Aug. 23, 2003, at B1.

[36] *See generally* Joyce, *supra* note 13.

[37] *See* Congressional Budget Act of 1974 (codified as amended at 2 U.S.C. 644, § 313 (2000)).

[38] *See* Michael W. Evans, *The Budget Process and the "Sunset" Provision of the 2001 Tax Law*, 99 Tax Notes 405, 405, 410–12 (2003).

[39] *See, e.g.*, I.R.C. § 51 (2000); I.R.C. § 41 (2000).

[40] *See* Elizabeth Garrett, *Harnessing Politics: The Dynamics of Offset Requirements in the Tax Legislative Process*, 65 U. Chi. L. Rev. 501, 567–68 (1998) (discussing extenders and whether their structure increased congressional monitoring and review).

tax bills. Virtually all of the major tax provisions of the EGTRRA and the JGTRRA expire between 2004 and 2010, a technique used solely to avoid the Byrd Rule's requirement of a supermajority vote in a Senate where the tax cuts could barely garner majority support.

If these tax cuts are made permanent—a course the President and his congressional supporters promised to pursue before the ink was dry on the bills[41]—the Congressional Budget Office ("CBO") projects that almost $1.6 trillion will be added to the deficit over the next decade, and the government will owe an additional $289 billion of interest on the national debt.[42] In addition, lawmakers will soon need to fix the alternative minimum tax ("AMT") levied on individuals,[43] which is increasingly triggered by the deductions, credits, and other tax breaks taken by upper-middle-class Americans. CBO projects that one way to fix the AMT problem—extending the temporary legislative patch included in JGTRRA—would cost an additional $693 billion over the next decade.[44] William Gale puts the figures into better perspective: to remove the sunsets in the tax code would reduce revenue by 2.4% of GDP on a permanent basis. To fix the AMT so that only 3% of taxpayers are on the system would reduce revenues by 2.7% of GDP.[45]

The structure of many expiring tax provisions is telling. First, the vast majority of popular provisions expire in election years, thereby reducing the chance that members of Congress will be willing, even by inaction, to raise taxes on voters. Second, less popular provisions are paired with very popular ones, encouraging Congress to extend them all in a package. For example, stimulus legislation passed in March 2002 allows businesses to take advantage of "bonus" depreciation and to deduct immediately thirty percent of the costs of new investments in equipment and facilities. That provision is temporary—expiring on the same day as the increase in the Child Tax Credit, the expansion of the ten percent tax bracket, marriage-tax relief, and the AMT "fix." As researchers Robert Greenstein, Richard Kogan, and Joel Friedman write: "Having the depreciation provision expire at the same time as these other provisions sets the stage for the depreciation provision, which enjoys strong corporate support, to be included in 2004 in a package extending these other,

---

[41] *See* Treasury Explanations of FY 2004 Proposal, *supra* note 22, at 1; *GOP, Lobbyists Say Big Tax Cut Only the Start*, HOUS. CHRON., June 7, 2001 at A7 (quoting GOP lawmakers); JOEL FRIEDMAN ET AL., CENTER ON BUDGET AND POLICY PRIORITIES, THE ADMINISTRATION'S PROPOSAL TO MAKE THE TAX CUT PERMANENT 1 (2002) *available at* http://www.cbpp.org/2-4-02tax.htm.

[42] *See* CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: AN UPDATE 12–13, tbl.1-6 (Aug. 2003) [hereinafter CBO UPDATE].

[43] *See* I.R.C. § 55 (2000).

[44] *See* CBO UPDATE, *supra* note 42, at tbl.1-8 (including additional interest payments on the debt, and interaction effects if the expiring provisions are also made permanent).

[45] *Perspectives on Long-Term Budget Deficits: Hearing Before the House Comm. on the Budget*, 108th Cong. 1 (2003) (statement of William G. Gale, Brookings Inst., Tax Policy Center).

virtually-certain-to-be renewed provisions."[46] There is no legal commitment that requires Congress to extend these provisions, but there is a political commitment.[47] Congress is likely to deliver on this commitment for several reasons: because of the past experience with extenders, which were almost always extended sometimes retroactively; because of the support they received when they were passed; and because of the structure of the expirations. Budget rules should not allow lawmakers to resort to the strategem of expiring provisions that never really expire to mask the long-term cost of tax reduction bills.

Not only do such provisions obscure the reality of fiscal decisions, but they also complicate the tax code, somewhat undermining the certainty required for long-term investment planning and increasing significantly the cost of compliance. Of course, the more likely it is that lawmakers will live up to the political commitment to extend expiring provisions, the more certain the tax environment is for business planning, notwithstanding the sunsets in the Internal Revenue Code. But there are no guarantees in politics, and continuing large deficits and other fiscal conditions might convince lawmakers to renew only some of the Bush tax cuts.

Take, as an extreme example of these complicated and nonsensical provisions, the "repeal" of the estate tax. The estate tax is only repealed for one year—2010. In earlier years, the exemption level is gradually raised, and the rates are slowly reduced. In 2011, however, the estate tax returns and in the same form it had before EGTRRA. As a result, old and infirm taxpayers with large estates should be planning to time their deaths strategically if they want to be certain to escape estate tax.

This strange design of a "repeal" is primarily driven by revenue concerns and current accounting rules, but more may be going on. Professors Linda Cohen, Edward McCaffery and Fred McChesney argue that this is a trick in the interest of lawmakers and lobbyists who profit from the uncertainty of expiring provisions.[48] Although it seems likely that Congress will extend the estate tax repeal—and perhaps make it permanent if lawmakers can find the money to offset the revenue losses—it is not certain.[49] So groups on both sides of the issue contribute to sympathetic can-

---

[46] ROBERT GREENSTEIN, ET AL., CENTER ON BUDGET AND POLICY PRIORITIES, NEW TAX CUT LAW USES GIMMICKS TO MASK COSTS: ULTIMATE PRICE TAG LIKELY TO BE $800 BILLION TO $1 TRILLION 7 (2003), *available at* http://www.cbpp.org/5-22-03tax.htm.

[47] *Cf.* Jackson, *supra* note 1, at Part I.B.2 (describing legal and political status of Social Security, which is supported by a deeper political commitment than expiring provisions but similarly is not a legal entitlement that cannot be repealed in the future).

[48] *See generally* Linda Cohen, Edward J. McCaffery & Fred S. McChesney, *Shakedown at Gucci Gulch: A Tale of Death, Money and Taxes* (USC-Caltech Ctr. for the Study of Law & Politics, Working Paper No. 22) *available at* http://lawweb.usc.edu/cslp/pages/papers.html (last modified Oct. 7, 2003).

[49] Recent news stories about compromises that would retain some aspects of the estate tax contribute to the uncertainty. *See* Jonathan Weisman, *Estate Tax Opponents May Be Forced to Compromise*, WASH. POST, Oct. 22, 2003, at E1 (discussing secret efforts by

didates, hoping to succeed in the battle over what happens in 2011. The sleight of hand of expiring tax provisions that are extended for a few years regularly, but at least face some possibility of lapsing, allows legislators to extract payments from affected groups—in the estate tax example, the interest groups include the insurance industry, wealthy taxpayers, estate-tax lawyers, and large nonprofits.[50]

Lobbyists are another group prominent in lawmaking that benefit from activity associated with expiring provisions. They are guaranteed a stream of income as they work to extend the expiring provisions, and when they "manage" only to obtain a temporary extension, they can explain to their less sophisticated clients that it is only because of their hard work and skill that the provision did not expire. Lawmakers have every incentive to back up the lobbyists' story because they also benefit from the rent extraction game. Although this story is too cynical if it purports to be a full explanation of the popularity of the expiring provision gimmick, the possibility of rent extraction certainly plays a role in the legislative dynamics.

Budget accounting rules should reflect the reality that most temporary provisions are extended and never allowed to expire. The more draconian reform option is to instruct CBO and the Office of Management and Budget ("OMB") to score expiring provisions as though they are permanent, but this approach might be too inflexible because certainly there are some expiring tax provisions that are truly temporary. A benefit designed to deal with an emergency or to provide tax relief for members of the armed services after a particular military conflict, for example, might actually expire when it is supposed to. Even these provisions may linger on past their scheduled demise, however, because often the existence of a benefit creates interest groups that lobby for extension, arguing that the adverse conditions have not abated or that new emergencies have developed. For example, one of the most powerful interest groups that supported one of the extenders—the targeted jobs tax credit now called the work opportunity tax credit—was the industry that helped firms handle paperwork associated with the tax provision.[51]

Nonetheless, some expiring provisions are intended to apply for a certain period of time and then expire, and they will in fact cease to be effective. A scoring convention that considered them to be permanent would unduly hinder their enactment. Thus, an alternative reform would be to produce, during committee deliberation and before floor consideration, two revenue estimates of expiring provisions—the revenue loss if

---

Senator Kyl (R-Ariz.) to modify but retain the estate tax).

[50] *See also* FRED S. MCCHESNEY, MONEY FOR NOTHING: POLITICIANS, RENT EXTRACTION, AND POLITICAL EXTORTION (1997) (presenting theory generally and in context of tax legislation); Garrett, *supra* note 40, at 543–55 (using rent extraction to explain legislator motives in adopting offset requirements).

[51] Garrett, *supra* note 40, at 521–22.

expiration occurs as scheduled and the revenue loss if the provision is made permanent. For this information to change the political reality, the Byrd Rule or some similar budget enforcement procedure must be recalibrated to trigger protection even when a provision is slated to expire.

* * *

For many of the reasons Jackson favors changing the accounting for Social Security—reasons that include "exposing short-range cash-flow effect"[52] and "identifying back-loaded reform proposals"[53]—the government should adopt methods to improve accounting for other areas of the budget. Although this article focuses on tax legislation, in part because the gimmicks are similar to those that plague Social Security reform and in part because the recent Bush tax bills contain so many extreme examples of them, current budget accounting methods can obscure the real costs of many other fiscal decisions. Changing accounting conventions so that the present value of long-term revenue loss or gain is calculated, plays a role in decision-making, and also triggers enforcement mechanisms such as points of order, presents its own technical difficulties, such as the appropriate discount rate and the uncertainty inherent in long-range estimations. But Jackson's thoughtful proposal concerning Social Security suggests that these challenges are not insurmountable and are worth facing if new accounting procedures will improve transparency and change the terms of the fiscal debate.

Accounting methods, like budget frameworks generally, are ways to structure policy debates so that lawmakers can better make the tradeoffs necessary in a world of limited resources and so that they and voters have an accurate sense of the policy choices made. The detailed analysis of Jackson's *Accounting for Social Security* gives a flavor of the difficult work ahead to reform accounting throughout the federal budget. As the country enters a new period of sustained and substantial budget deficits, which give only a hint of the long-term fiscal imbalance of current policies, reforming budget accounting is crucial. The current burgeoning cash-flow deficits may convince policymakers to undertake this task, particularly because much of the past budget framework has expired and must be replaced. The construction of a new budget framework presents an opportunity to revamp accounting conventions as well. Of course, pandering and shell games may continue to occur. The goal of budget rules is to make such misleading subterfuge more difficulty or costly; it cannot eliminate it altogether.

---

[52] *See* Jackson, *supra* note 1, at Part III.B.1.
[53] *See id.* at Part III.B.5.

COMMENT

THE VIRTUES OF MOVING FROM CASH TO ACCRUAL ACCOUNTING FOR SOCIAL SECURITY

ROBERT C. POZEN*

Professor Jackson has written an excellent critique of the accounting methods currently applied to Social Security. In short, he articulates two major concerns.[1] First, including the annual cash-flow surplus of Social Security in calculations of the annual U.S. budget surplus or deficit misleads the public and confuses policy makers.[2] Second, emphasizing cash-flow accounting numbers for Social Security creates the illusion that its short-term finances are sound, although its long-term financial problems are serious.[3] In response to both these concerns, Jackson proposes a modified form of accrual accounting, which would show large deficits for Social Security in the short term as well as the long term. Jackson concludes that if Congress recognized the large annual deficits run by Social Security under accrual accounting, it would have a strong incentive to exclude Social Security from calculations of the annual U.S. budget surplus or deficit.[4]

This Comment begins by commenting on the two main concerns discussed by Jackson and his proposals for resolving these concerns. Next, it will discuss the implications of his proposals for evaluating various plans to reform Social Security. In this regard, I will draw upon my experience as a member of President Bush's Commission to Strengthen Social Security [hereinafter the "Bush Commission"].

I. ONLINE AND OFFLINE BUDGETS

Congress is fond of making the U.S. budget look more positive than it actually is by including the annual surpluses of Social Security based on cash-flow accounting. During the last few years of the Clinton administration, this device was used to inflate current budget surpluses by

---

* John M. Olin Visiting Professor of Law from Practice, Harvard Law School. Secretary of Economic Affairs, Commonwealth of Massachusetts. J.D., Yale Law School, 1972; B.A., Harvard University, 1968.

[1] Howell Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

[2] *See id.* at Part I.B.

[3] *See id.* at Part I.A.

[4] *See id.* at Part III.

$100 to $200 billion per year.[5] During the Bush administration, this device is being used to reduce the size of the annual budget deficit by similar amounts.[6]

The inclusion of cash-flow surpluses from Social Security in the annual numbers for the total U.S. budget is unjustified as an economic matter and inconsistent with relevant legislation, as Jackson demonstrates.[7] Additionally, by systematically understating the size of the current budget deficit, the inclusion of Social Security's cash-flow surplus is bound to mislead the public about the extent of U.S. government spending. What if the public were told that the official 2003 federal budget deficit of $400 billion is actually more than $550 billion without the addition of an ephemeral cash-flow surplus of more than $150 billion from Social Security? The size of this accounting distortion makes Enron look like a minor accounting error.[8]

Fortunately, relief is in sight. By 2007, the annual cash-flow surpluses of Social Secruity will begin to decline; by 2017, they are projected to turn negative.[9] Congress, therefore, will be inclined over the next decade to adopt Jackson's suggestion that Social Security be unconsolidated completely from the U.S. budget. Under this approach, any transfers from the U.S. government to the Social Security system would be treated in the same manner as U.S. government outlays to third parties.[10]

The placement of the Social Security system totally outside the U.S. budget would also help clarify the nature of assets held by the so-called Social Security trust fund. The status of the Social Security trust fund has been the subject of vigorous academic and political debate. In reality, the trust fund is an accounting conduit: it represents a set of claims by current and future retirees on the U.S. Treasury. By keeping the Social Security trust fund off-budget, Congress will see more clearly that Social Security claims of workers are not self-financing; rather, these claims are akin to Treasury bonds already held by public investors. In both cases, the U.S. government can pay out monies owed to third parties only if

---

[5] *See* CONGRESSIONAL BUDGET OFFICE, THE ECONOMIC AND BUDGET OUTLOOK: AN UPDATE, at 53 (Aug. 1998).

[6] *See* CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: FISCAL YEARS 2002–2011, at 19 (Jan. 2001).

[7] *See* Jackson, *supra* note 1, at Part I.B.

[8] On October 16, 2001, Enron announced that it was reducing shareholders' equity by $1.2 billion because of improper accounting. *See* REPORT OF INVESTIGATION BY THE SPECIAL INVESTIGATIVE COMMITTEE OF THE BOARD OF DIRECTORS OF ENRON CORP., at 2–3 (Feb. 1, 2002), *available at* http://news.findlaw.com/hdocs/docs/enron/sicreport/sicreport020102.pdf. Less than one month later, Enron announced a restatement of its financial statements for the years 1997 through 2000 because of other forms of improper accounting; these restatements involved a total reduction of a little more than $2 billion in shareholders' equity. *See id.*

[9] *See* 2003 BD. OF TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND THE FED. DISABILITY INS. TRUST FUNDS ANN. REP., [hereinafter 2003 TRUSTEES REPORT] at 8.

[10] *See* Jackson, *supra* note 1, at Part II.D.2.b.

Treasury sells new bonds to the public or appropriates funds for this purpose.[11]

II. CASH TO ACCRUAL ACCOUNTING

As Jackson explains, annual cash-flow accounting is inappropriate for a retirement plan such as Social Security that increases its long-term liabilities each year, even though these liabilities will not be paid out for several decades.[12] Instead, he suggests a modified form of accrual accounting, based on the way private pension plans file their annual reports under Generally Accepted Accounting Principles ("GAAP").[13] His proposed methodology would focus on the addition of retirement benefits earned by current workers reduced by the excess of taxes to be paid by such workers over the additional benefits they will accrue over the rest of their working lives.[14] If the government were to follow the accounting method proposed by Jackson, the net accrued liabilities of Social Security would have grown by roughly $370 billion in 2002, as compared with the $160 billion surplus recorded for Social Security under cash-flow accounting.[15]

The change from cash-flow to accrual accounting would have a dramatic impact on the politics of Social Security reform. Under cash-flow accounting, there is little political incentive to deal with unpopular issues such as payroll taxes and benefit formulas.[16] When members of the Bush Commission pressed for prompt action on Social Security reforms, they were often met with the response that there was no rush because the system would be running a surplus for the next fifteen to twenty years. Many Senators and Representatives said they were not willing to incur the political risks involved with advocating Social Security reforms since they would no longer be in office when the Social Security surplus turned negative. Under accrual accounting, by contrast, the Social Security deficit is growing by leaps and bounds each year. If a politician delays Social Security reform for even one more year, the Social Security system will be billions of dollars further in the red.

But Jackson stops short of banning cash-flow accounting for Social Security; he wants to introduce modified accrual accounting alongside cash-flow accounting. While it is difficult to argue against presenting multiple perspectives on a complex set of issues, it is critical to establish

---

[11] *See id.* at Part II.D.2.c.

[12] *See id.* at Part I.A.2.a.

[13] *See generally*, EMPLOYERS' ACCOUNTING FOR PENSIONS, Statement of Financial Accounting Standards No. 87 (Financial Accounting Standards Bd. 1985) [hereinafter FAS No. 87].

[14] *See* Jackson, *supra* note 1, at Part II.A.

[15] *See id.* at Part II.D.1.

[16] *See id.* at Part III.A.1.

a primary accounting approach that best captures the economic implications of Social Security. In this view, cash-flow accounting is inferior to accrual accounting not only because cash-flow accounting conflicts with GAAP for pension plans,[17] but also because cash flow accounting systematically understates the real economic cost of Social Security to taxpayers. One useful analogy is to consider how the SEC's restrictions on a company's presentation of non-GAAP measures of its financial performance—for example, the reporting of quarterly profits excluding acquisition charges—would apply to any government report on Social Security. While the SEC allows a company to announce non-GAAP measures, it must display the GAAP numbers prominently and specify which material items are omitted in the non-GAAP measures.[18] Following this SEC analogy, all government reports on Social Security would be required to use accrual accounting as the primary method of presentation. If cash accounting were used by a government report on a particular subject, then the report would also have to display prominently the results under accrual accounting and compare those to the results under cash-flow accounting.

More fundamentally, some critics argue against accrual accounting because Social Security benefits do not constitute a legally binding entitlement; they may, in theory, be reduced or eliminated by Congress.[19] In response, Jackson limits his modified accrual accounting system to Social Security retirement benefits already earned by workers, as compared to retirement benefits they may earn in the future.[20] While Jackson recognizes that Congress may retroactively change almost any aspect of Social Security, he points out that already earned benefits should be given preference over future benefits from the perspective of political realities and by analogy to private pension accounting.[21]

As Jackson recognizes, his proposal for modified accrual accounting on earned Social Security benefits is very similar to the closed-group unfunded obligation ("CGUO"), recently included in the back pages of the Social Security Trustees Annual Report ("Report").[22] That Report, however, does not specify the year-to-year increase in the CGUO; Jackson computes this increase by analyzing a set of data supplied by the Social Security Administration to experts on request. The Report also includes an estimate of the current open-end group liability for Social Security—another form of accrual accounting that includes future as well as current

---

[17] Compare cash-flow accounting with the standards set forth in FAS No. 87, *supra* note 13.

[18] *See, e.g.*, 17 C.F.R. §§ 229.10, 244.100, 244.101 (2003).

[19] *See generally*, ACCOUNTING FOR SOCIAL INSURANCE, Statement of Recommended Accounting Standards No. 17 (Federal Accounting Standards Bd. 1999) [hereinafter FAS No. 17].

[20] *See* Jackson, *supra* note 1, at Part II.A.

[21] *See id.* at Part II.C.4.

[22] *See* 2003 TRUSTEES REPORT., *supra* note 9, at 62–63.

workers.[23] Again, however, the Report does not specify year-to-year increases in open-end liabilities of Social Security.

The main objective should be for Social Security to move from cash-flow accounting to some form of accrual accounting as soon as possible. This move is critical to counteract the illusion that Social Security is in good financial shape for the next fifteen years and, as discussed below, to allow an accurate assessment of specific proposals to reform Social Security. Nevertheless, accrual accounting will be resisted by commentators who believe it is inappropriate for governmental programs where benefits or tax rates can be changed by Congress from year to year.[24] Although this general belief may have some validity, cash flow accounting on a yearly basis is unsound for retirement plans where the government is promising in the current year to make large payouts over thirty to fifty years in the future.

The CGUO should become the primary measure of the financial status of Social Security because the methodology for this form of accrual accounting is already established and because CGUO focuses on benefits already earned by current workers. CGUO is a more useful form of accrual accounting than open-end liability, because the latter includes projected Social Security benefits for people who are not currently workers but who are likely to become workers during the next seventy-five years. These projections would introduce unnecessarily a potentially large and speculative element into Social Security accrual accounting. The good news is CGUO for the current year is now included in the middle of the lengthy annual report by the Social Security trustees. The key is to move CGUO into the executive summaries of all official reports on Social Security, which should prominently display year-to-year changes in CGUO.

III. Accrual Accounting and Personal Retirement Accounts

As Jackson indicates, the advantages of accrual over cash-flow accounting for Social Secruity can readily be seen in the debate about the creation of a personal retirement account (PRA) for Social Security participants.[25] In the carve-out form of a PRA, workers would invest a relatively small portion of their Social Security payroll taxes (i.e., 3% of the 12.4% total in Social Security payroll taxes) in a portfolio of securities.[26]

---

[23] *See id.* at 61–62.

[24] *See* FAS No. 17, *supra* note 19, at Appendix A.

[25] *See* Jackson, *supra* note 1, at Part III.D.

[26] *See* Strengthening Social Security and Creating Personal Wealth for All Americans: Report of the President's Commission 82 (Dec. 21, 2001). For a summary of other PRA proposals for SS, see Liqun Liu & Andrew Rettenmaier, *Comparing Proposals for Social Security Reform*, National Center for Policy Analysis Report No. 227 (Sept. 1999), *at* http://www.ncpa.org/studies/s227/s227.html.

In return, workers would agree to a so-called offset—lower Social Security benefits than currently scheduled by an amount reasonably related to the portion of their payroll taxes invested in PRA.[27] Thus, such workers would receive two types of retirement benefits—lower scheduled benefits from Social Security plus the investment returns from their PRA.

From the perspective of many workers, especially younger ones, the concept of a carve-out PRA should have much appeal. The projected real return after inflation for Social Security contributions over the next few decades is low (i.e., 1% to 2% per year),[28] relative to the real return estimated by Social Security actuaries for a balanced portfolio invested 50% in an equity index fund and 50% in a bond index fund (4.6% annually after administrative expenses).[29] Moreover, many younger workers are deeply skeptical that they will actually receive the level of Social Security benefits scheduled for their retirement.[30] Having participated in 401(k) plans, younger workers may like the idea of controlling the investment of some portion of their payroll taxes in a PRA, rather than having all their payroll taxes funneled into an arcane system in Washington, D.C.

Additionally, carve-out PRAs can reduce Social Security's long-term deficit if the offset rate is set appropriately. While a portion of payroll taxes is allocated to PRAs during working years, workers will receive lower Social Security benefits during their retirement years. This trade-off is visible immediately under accrual accounting because the amount of earned Social Security benefits by current workers with PRAs would be significantly lower than those under the current system. An accrual accounting system would also drive home, as Jackson notes, the transitional costs of moving from the current system to a new system with lower Social Security benefits and PRAs.[31]

In contrast, the merits of PRAs are obscured by cash-flow accounting. The lower cash flows into Social Security from workers with PRAs are apparent immediately, yet the lower benefits due to such workers are not apparent for forty or fifty years. This deficiency in cash-flow accounting is accentuated by the conventional use of a seventy-five-year period to measure the long-term Social Security deficit. Consider a pro-

---

[27] *See id.* at 90.

[28] *See id.* at 88.

[29] *See id.*

[30] A 1997 survey by the Employee Benefit Research Institute and Matthew Greenwald & Associates asked whether people had greater confidence that they would receive Social Security benefits or that alien life would exist in outer space. Among Generation Xers, 63% had greater confidence in receiving Social Security benefits while 33% had greater confidence that alien life exists in outer space (4% presumably did not answer). *See The Future of Social Security for This Generation and the Next: Hearing Before the House Ways and Means Subcomm. on Soc. Sec.*, 105th Cong. (1997) (statement of Dallas L. Salisbury, President, Employee Benefit Research Institute), *at* http://waysandmeans.house.gov/legacy.asp?file=legacy/socsec/105cong/10-23-97/10-23salis.htm.

[31] *See* Jackson, *supra* note 1, at Part III.D.1.

posal for young people entering the workforce in 2037 to allocate 3% of their Social Security payroll taxes to PRAs. This allocation of 3% to PRAs would reduce the annual cash flows of Social Security every year from 2037 onwards, but the reduction in benefits accepted by these workers will not help the annual cash flows of Social Security until after 2077. Thus, under cash-flow accounting, there is no positive impact on today's estimate of the long-term Social Security deficit.

Similarly, the consolidation of the Social Security trust fund into the total U.S. budget precludes a fair evaluation of any Social Security reform proposal involving investments in the securities markets. Under current accounting practice, claims by the Social Security trust fund on the U.S. Treasury are considered intra-governmental transfers with no effect on the total U.S. budget.[32] Accordingly, if some portion of the Social Security trust fund were invested in publicly traded securities, this portion would be deducted from U.S. budget totals because that portion would no longer constitute an intra-governmental transfer. As Jackson explains, however, this accounting method distorts public debate because the amounts held by the Social Security trust fund would remain the same immediately before and after the securities investment.[33] Instead, public debate should focus on the differences, other than budget accounting, between investments in diversified securities portfolios and intra-governmental claims on the U.S. Treasury.

For example, suppose the Social Security trustees proposed to invest a small portion of trust assets directly in a broad-based index of the U.S. stock market like the Wilshire 5000.[34] The debate on this proposal should focus on the expected returns and risks of this index investment, as compared to the economic returns and political risks of holding only U.S. Treasury bonds that must ultimately be repaid by congressional appropriations. This legitimate debate is only confused by the argument against the proposed investment of Social Security assets in a stock market index on the ground that such an investment would automatically increase the consolidated budget deficit. This artificial argument would be eliminated by keeping Social Security trust assets separate from the rest of the U.S. budget.

IV. Implications for Other Reform Proposals

The move from cash-flow to accrual accounting would enhance the quality of public debate on two other types of Social Security reform proposals: addition of participants and changes in benefit formulas.

---

[32] *See id.* at Part III.D.2.

[33] *See id.*

[34] Several commentators have made this type of proposal. *See generally* Henry J. Aaron & Robert D. Reischauer, Countdown to Reform: The Great Social Security Debate 119–20 (Century Found. Press 2001) (1998).

A. New Participants for Social Security

A number of commentators have suggested that Social Security can be financially bolstered by bringing new participants into the system. One common suggestion is to bring into the Social Security system the many state and municipal workers who are currently exempt from it.[35] Another proposal is to bolster the finances of Social Security by increasing immigration of workers into the United States.[36]

Under cash-flow accounting, an addition of Social Security participants looks like a winning idea because annual revenues begin to increase immediately and higher benefits do not materialize until years later. The addition of any new cohort of workers may turn out, however, to be a net loser for Social Security over time if the total retirement benefits paid to such workers exceed their lifetime contributions plus interest. This could easily happen with low-wage workers because they will receive a relatively high level of benefits for each dollar they contribute to Social Security.

Jackson points out this deficiency of cash-flow accounting for evaluating the economic implications of expanding Social Security participation.[37] Under his approach to accrual accounting, by contrast, the financial effects of adding new categories of workers would be reflected accurately. From the start, the incremental revenues from new participants each year would be offset to some degree by the accrual of Social Security benefits earned in that year by new participants. By including increases in benefit obligations as well as increases in contributions by new participants, accrual accounting would allow policy makers to estimate better and monitor the impact of new participant proposals on the long-term deficit of Social Security.

B. Changes in Benefit Formulas

To reduce the long-term deficit of Social Security, commentators have suggested a variety of changes in the benefit formulas for retirees.[38] For example, some have advocated an increase in normal retirement age as life expectancy rises. Others have promoted the use of price indexing, rather than wage indexing, to adjust initial benefits in order to reflect the growth of the U.S. economy over the worker's career.

Because any change in benefit formulas is controversial, it may be politically important to apply the change only prospectively. For instance, one of the constraints in the executive order establishing the Bush

---

[35] *See* U.S. GEN. ACCOUNTING OFFICE, SOCIAL SECURITY: DIFFERENT APPROACHES FOR ADDRESSING PROGRAM SOLVENCY 33 (1998).

[36] *See* AARON & REISCHAUER, *supra* note 34, at 64–65.

[37] *See* Jackson, *supra* note 1, at Part III.B.1, III.C.3.

[38] *See generally*, U.S. GEN. ACCOUNTING OFFICE, *supra* note 35, at 38–49.

Commission was the maintenance of the current Social Security benefit structure for all workers "at or near retirement." This phrase was interpreted by the Commission members to mean that any change in the benefit structure would not apply to any worker fifty-five years old or older. Such an interpretation implies that, for workers younger than fifty-five, a change in some aspect of the Social Security formulas might apply to benefits already earned in the past, as well as retirement benefits to be earned in the future.

Jackson's proposal for modified accrual accounting supplies a particularly useful way to implement any reform proposal involving the grandfathering of benefits for various categories of workers.[39] Because his proposal captures the long-term cost of Social Security benefits already accrued by current workers, it would allow for the application of new formulas only to retirement benefits that accrue after the date the reforms are instituted. For instance, wage indexing could be applied only to initial Social Security benefits accrued in the past, while price indexing could be applied to benefits earned in the future. Such an approach to Social Security reform, as Jackson emphasizes, would have the political virtue of honoring existing commitments, while offering a rapid improvement in the long-term Social Security deficit by reducing the rate of benefit accrual in the years immediately following the adoption of the reform.[40]

Conclusions

In this comprehensive article, Jackson sharply criticizes the cash-flow accounting currently used in the Social Security system and puts forward a thoughtful proposal for a modified form of accrual accounting that should be employed in Social Security accounting. Although the exact details of his proposal can be debated, his argument in favor of accrual accounting over cash-flow accounting is compelling.

Accrual accounting for Social Security is, however, an arcane subject with little popular appeal, and politicians have a vested interest in the overly optimistic picture of Social Security created by cash-flow accounting. If Jackson wants his proposal adopted for Social Security accounting, he should publish a series of one-page editorials in newspapers and magazines.[41] In this article, he explains in detail why cash-flow accounting for Social Security is inappropriate, but useful, to the Washington establishment. Possibly through brief editorials, he can generate

---

[39] *See* Jackson, *supra* note 1, at Part III.F.

[40] *See id.* However, Congress may want the flexibility of changing already accrued Social Security benefits, at least for younger workers.

[41] During the process of editing this comment, Professor Jackson did publish such an editorial. *See* Howell E. Jackson, *It's Even Worse Than You Think*, N.Y. Times, Oct. 9, 2003, at A35.

enough public support to make accrual accounting the primary measure of the financial status of Social Security.

COMMENT

ACCRUAL ACCOUNTING AND THE FISCAL GAP

DANIEL N. SHAVIRO*

I. INTRODUCTION

As Professor Jackson's article[1] helps to show, two important points about the federal budget have won increasingly wide recognition in recent years. The first is that traditional cash-flow measures cannot provide meaningful budgetary information when the government has taken on huge long-term commitments, such as those under Social Security and Medicare, rather than simply spending discretionary annual appropriations on soldiers and roads. The second is that the current set of policies is unsustainable because the financing behind these policies falls short of the benefits due. The basic problem is that, under current policy, Social Security and Medicare spending, but not their financing, are on track to grow rapidly relative to the economy for many decades to come. This fact reflects pervasive demographic and technological trends toward longer lifespans and more expensive, albeit better, health care.[2] As a result the government will ineluctably be forced, at some point not too far in the future, to enact substantial tax increases or benefit cuts, or both. Current fiscal policy, unfortunately, has since 2001 been headed in precisely the opposite direction, featuring huge tax cuts, a potentially open-ended new Medicare entitlement for prescription drugs, and costly foreign engagements as to which no extra financing is even suggested. Future generations will have to pay the bill for all this.

Against this background of inadequate information encouraging unsustainable policies that poor political decisions have made even worse, Jackson's analysis of accounting for Social Security is helpful and illuminating. His basic point about the need for accrual accounting, rather than annualized cash-flow accounting, is so obviously and unanswerably correct that the continued existence, or even prevalence, of skeptics is deeply discouraging. Even those who believe that they are protecting So-

---

* Wayne Perry Professor of Taxation, New York University School of Law. J.D., Yale Law School, 1981; A.B., Princeton University, 1978.

[1] Howell E. Jackson, *Accounting For Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

[2] *See* DANIEL N. SHAVIRO, WHO SHOULD PAY FOR MEDICARE? (forthcoming 2004).

cial Security by insisting on cash-flow accounting are deluding themselves, unless they are indifferent to the program's prospects beyond the very short term. About the only logically coherent reason for favoring continued cash-flow accounting is that it helps to encourage favoring current generations relative to future ones through the enactment of further tax cuts and spending increases. The theory would have to be that current voters are not selfish enough to favor themselves sufficiently relative to future voters unless misleading accounting measures are used to encourage them to ignore the burdens they are leaving for others to meet. Yet it is hard to see why current voters should be discouraged from giving due, or at least some, consideration to the interests of future generations.

Additionally, Jackson should be commended for making a unique contribution in one respect. Others have written about the economic accrual of unfunded Social Security obligations and of the overall fiscal gap and its likely impact on future generations.[3] Jackson, however, adds an accounting background that has not been included previously in this literature. This background permits him not only to compare Social Security reporting options to the techniques used under Generally Accepted Accounting Principles (GAAP) to address similar types of problems, but also to consider what Social Security accounts ought to look like if they are to be effective in communicating information to users, be they political actors or prospective beneficiaries.[4]

Social Security is an especially good area for making certain of the distinctions Jackson explores, such as that between currently accrued and other future benefits. Under the Social Security benefit formula, one can specify the value of the benefits that a given individual has accrued as of a given point of time. By contrast, one becomes eligible for the full array of Medicare benefits after only forty quarters, effectively ten years, of work by oneself or one's spouse.[5] Further work and payment of payroll taxes yields no additional Medicare benefits. It might seem artificial to treat people who have worked forty quarters as entitled to the full array of Medicare benefits, and those who have worked thirty-nine or fewer quarters as entitled to nothing. Moreover, Medicare benefits, being in-kind, have only an estimated value at any given time that depends on health care trends among other considerations. In the rest of the fiscal system, even insofar as future outlays such as military spending can be reliably predicted, there is even less sense in which any portion of the expected cost could be described as having accrued to anyone's benefit.

---

[3] *See, e.g.*, JAGADEESH GOKHALE & KENT SMETTERS, FISCAL AND GENERATIONAL IMBALANCES: NEW BUDGET MEASURES FOR NEW BUDGET PRIORITIES (2003); LAURENCE J. KOTLIKOFF, GENERATIONAL ACCOUNTING (1992); DANIEL N. SHAVIRO, DO DEFICITS MATTER? (1997); Alan J. Auerbach, *Where We Are, How We Got Here, and Where We're Going*, NBER MACROECON. ANN., at 141–75 (1994).

[4] *See generally* Jackson, *supra* note 1.

[5] *See* 42 U.S.C. § 414 (2000).

Because Social Security benefits can be divided into those that have accrued and those that have not, Social Security allows a more differentiated accounting treatment than the rest of the budget. Normatively, however, this need not imply that already accrued benefits ought to be treated differently than other expected future outlays. After all, even if the assumption is made that precise and definite expectations should be honored more fully than looser expectations, it is not clear to what extent people are aware of the exact state of their accrued Social Security benefits.[6]

Notwithstanding the interest of some of the details, the biggest takeaway point for readers of Jackson's article should be his overall bottom line. Under his estimates, Social Security's fiscal gap—or its unfunded obligations, in his terminology—has a present value of $10.5 trillion,[7] an amount that is almost exactly equal to the annual gross domestic product (GDP) of the United States economy.[8] Moreover, as new Social Security obligations accrue and prior ones appreciate, the Social Security fiscal gap is rapidly increasing. In 2002, for example, when the Social Security Trustees reported an operating surplus of $165.4 billion, Jackson suggests that, under GAAP principles, they would have had to report a deficit of $467.8 billion.[9]

In computing these amounts, Jackson treats the Social Security system as having accumulated reserves of $1.4 trillion, which yielded $80.4 billion of interest income in 2002.[10] Had these amounts been excluded from his calculations, the Social Security fiscal gap would have been $11.9 trillion, and the 2002 Social Security deficit would have been $547.9 billion.[11] One could certainly argue, although it would raise complicated issues not worth pursuing in full here, that this is a more reasonable presentation.[12] After all, Social Security's accumulated reserves are merely an inter-government I.O.U., reflecting historical bookkeeping records concerning the subset of total government revenues and outlays that are officially attributed to the system. Moreover, the interest income on these supposed Social Security assets was paid by the government to itself, via government bonds that are attributed to the Social Security Trust Fund.[13] Self-paid interest has no effect on the government's overall fiscal position or the long-term affordability of any given program.

---

[6] DANIEL N. SHAVIRO, MAKING SENSE OF SOCIAL SECURITY REFORM 12–13 (2000).

[7] Jackson, *supra* note 1, at Part II.C.1.

[8] The United States GDP in 2002 was approximately $10.4 trillion. CENT. INTELLIGENCE AGENCY, WORLD FACTBOOK 2003 (2003), *available at* http://www.cia.gov/cia/publications/factbook/geos/us.html#Econ.

[9] Jackson, *supra* note 1, at Part II.B.2.

[10] *See id.*

[11] *See id.*

[12] *See* SHAVIRO, *supra* note 6, at 92–93.

[13] CONGRESSIONAL BUDGET OFFICE, THE BUDGET AND ECONOMIC OUTLOOK: AN UPDATE 33 (Aug. 1998) (noting that trust fund surpluses are "invested in interest bearing government securities, and that interest is part of the funds' income").

In any event, however, the numbers are so large either way that the treatment of self-paid interest almost does not matter. Moreover, Social Security is but one piece of the overall fiscal picture, and far from being the worst piece. Jagadeesh Gokhale and Kent Smetters estimate the Medicare fiscal gap at $36.6 trillion,[14] a conservative estimate given their use of relatively short life expectancy and a high discount rate.[15]

What really matters the most, because the government's various commitments are likely to stand or fall together, is the overall fiscal gap. As it happens, this figure is not far from being the sum of the Social Security and Medicare fiscal gaps, because the rest of government taxes and projected spending are approximately in balance.[16] Gokhale and Smetters projected that the overall U.S. government fiscal gap, as of 2003, would stand at $45.5 trillion.[17] Their estimate of the accrual equivalent for the annual cash-flow federal budget deficit stood at around $1.3 trillion for 2003 and $1.5 trillion for 2004.[18] These deficits are more than double the projected cash-flow budget deficits for those years, and the cash-flow deficits were already high enough to cause considerable heartburn in the Washington policy community.[19] Once again, moreover, other reasonable estimates might be considerably higher than theirs. Another recent estimate by economists suggests a fiscal gap of about $74 trillion, if one applies a 3% discount rate and assumes that real GDP will grow at an annual rate of 1.5%.[20]

The fiscal gap is purely a measure of under-specification of the Federal government's actual long-term policy path. That is, it describes the extent to which announced or reasonably projected future outlays have a present value in excess of that for assets on hand plus projected future revenues, under current policy. Given the fundamental budget constraint that all economic actors face, however, over the long run the government cannot spend more than it collects.[21] Thus, the main thing, or at least the first thing, that can be learned from the fiscal gap is that the projected policy will not in fact be the actual policy. At some point, taxes will have to be a lot higher than current policy suggests, or spending (especially on Social Security and Medicare) will have to be a lot lower, or both.

---

[14] GOKHALE & SMETTERS, *supra* note 3, at 39.

[15] *See, e.g.*, Jackson, *supra* note 1, at Part II.C.3 (employing longer life expectancy predictions and lower discount rates).

[16] *See* GOKHALE & SMETTERS, *supra* note 3, at 39.

[17] *See id.* at 36.

[18] *See id.* The estimates attributed to them for the 2003 and 2004 accrual deficits were made by taking the projected fiscal imbalance for each year and subtracting the fiscal imbalance shown for the prior year.

[19] *See generally* William G. Gale & Peter R. Orszag, *The Budget Outlook: Analysis and Implications*, 101 TAX NOTES 145 (2003).

[20] Daniel N. Shaviro, *The Growing US Fiscal Gap*, 3 WORLD ECON. J. 1, 2–3 (2002).

[21] Even if the government could pay for its programs indefinitely by printing money, such a policy would function as the equivalent of a tax on people who lost purchasing power due to the resulting inflation.

The main implications of this disconnect between projected and feasible policy include the following. First, older generations are receiving an enormous lifetime wealth transfer from younger and future generations. By one recent estimate that predates the 2001 through 2003 budgetary enactments, future generations will face lifetime net tax rates (taxes paid minus transfers received, divided by lifetime income) more than double those being borne by members of currently living generations.[22] Given that future generations are not likely to receive commensurately more government services in exchange for their much higher net taxes, this disparity amounts to a direct transfer from future to current generations. Second, current workers who will still be alive when the "generational storm"[23] hits may experience unexpected hardship in their retirement years if they are saving too little apart from what Social Security and Medicare will give them.[24] Third, we face the significant possibility of an Argentina-style meltdown in the U.S. government's position as a borrower in world capital markets, potentially yielding chronic inflation, unemployment, and bank and currency crises.[25] Whether this happens, and the seriousness of the crises if they occur, depends in large part on how well the American political system responds as the fiscal implosion nears. In this regard, however, the last few years, and in particular the capture of the Republican Party by rabid anti-tax ideologues who lack any matching fervor for cutting government spending, do not bode well.

Jackson's focus is on how the government should account for Social Security, rather than on what its policy should be, so he spends relatively little time on the normative implications of its fiscal gap or that for the government as a whole.[26] He correctly notes that the case for accrual accounting, as a matter of good information reporting, need not imply that Social Security should be fully funded, a position that turns on separate issues of public finance.[27] Those who are interested in the pre-funding or fully funded issue, however, should keep in mind the distinction between questions of (1) full specification of a sustainable long-term policy course, and (2) generational distribution.[28]

---

[22] *See* LAURENCE J. KOTLIKOFF, THE COMING GENERATIONAL STORM (2001).

[23] LAURENCE J. KOTLIKOFF & SCOTT BURNS, THE COMING GENERATIONAL STORM (forthcoming 2004).

[24] Consider, for example, that the fiscal gap was estimated as of 2003 to equal 45.6% of projected Social Security and Medicare outlays. *See* GOKHALE & SMETTERS, *supra* note 3, at 36. One way of eliminating the fiscal gap, therefore, would be to reduce those outlays by 45.6%. While this is unlikely to be the chosen method, it helps make the point that significant reductions in these outlays are possible.

[25] *See* discussion in KOTLIKOFF & BURNS, *supra* note 23, and in Daniel N. Shaviro, *Reckless Disregard: The Bush Administration's Policy of Cutting Taxes in the Face of an Enormous Fiscal Gap*, 45 B.C. L. REV. (forthcoming 2004).

[26] *See generally* Jackson, *supra* note 1.

[27] *See id.* at Part III.F.

[28] *See, e.g.*, SHAVIRO, *supra* note 25 (manuscript at 12–18, on file with author).

To illustrate the difference between the two, suppose that Congress this year enacted massive Social Security tax increases and benefit cuts, to take effect in 2050. Suppose these changes were large enough to eliminate the Social Security fiscal gap, as measured under officially announced policy.[29] These changes would set a sustainable policy course, but they would not alter address the tendency of current fiscal policy to transfer wealth from younger to older generations.

This author has argued elsewhere that the impact of the current fiscal policy on future generations is unjustifiable.[30] While ongoing technological advances make it likely that they will be wealthier than Americans today (leaving aside the possibility of environmental, terrorism-related, or other catastrophes),[31] such advances also suggest that wealth will be more valuable to them, in terms of the actual uses to which they could put it, than it is to people today. "Suppose that you could give a million dollars to either of two individuals who suffer from advanced colon cancer: one living in 2005 who cannot be helped, or one living in 2045 who can actually, at great expense, be cured. It is hard to argue against giving the money to the latter individual, even if he is better-off in absolute terms."[32] Given the course of health care technology and Medicare's more than eighty-percent share of the overall fiscal gap,[33] this is no idle hypothetical. America faces the possibility of hurting future generations a lot more than helping ourselves if the government does not start lowering the fiscal burdens that current policy would impose on future generations.

Whether one agrees with this argument of inter-generational distribution or not, there is no justification for simply hiding from current voters the tradeoffs that America faces in Social Security and other budget policy choices. The adoption of Jackson's Social Security accounting proposals, and of a similar accrual-based approach to the entire federal budget, would be an important first step toward achieving greater sanity and realism in our public policy debate. The replacement of cash-flow accounting by accrual accounting is long overdue, and can no longer be reasonably or honestly argued against.

---

[29] Whether we should regard the announced 2050 changes as credible is another matter.

[30] *See* Shaviro, *supra* note 20, at 7.

[31] *See id.* at 6.

[32] *See* Shaviro, *supra* note 20, at 7

[33] *See* GOKHALE & SMETTERS, *supra* note 3, at 39.

COMMENT

THE INADEQUACIES OF ACCRUAL ACCOUNTING FOR SOCIAL SECURITY

KENT SMETTERS*

I agree with Professor Jackson's assessment of the problems arising from the traditional budgetary treatment of the Social Security system.[1] The Trust Fund's exhaustion date, which was recently projected as 2042 under the Social Security Actuary's "Intermediate Assumptions,"[2] has received most of the attention in the past.[3] After the Interim Report of The Commission to Strengthen Social Security, more attention is now also given to the date future Social Security benefits are expected to exceed future Social Security contributions, which was most recently projected as 2018.[4] This so-called "cross-over date" could have more relevance than the Trust Fund exhaustion date if past Trust Fund surpluses have been spent, rather than used to reduce the debt held by the public, and will likely continue being spent in the future.[5] These measures, however, do not provide adequate information concerning the financial problems facing Social Security. Policymakers can manipulate both of these measures, and neither will indicate whether we have actually increased national saving sufficiently to address baby-boomer demographic problems.

The current focus on cash-flow budgeting, deficits, and debt in the federal government's accounting system not only understates the problems facing the Social Security system, but this focus *biases* the choices of policymakers toward minimizing short-term deficits at the cost of long-term imbalances. A more relevant accounting framework would include both short- and long-term shortfalls.

---

* Assistant Professor of Insurance and Risk Management, The Wharton School, University of Pennsylvania. Research Assistant, Public Economics, National Bureau of Economic Research (NBER). Ph.D., Harvard University, 1995; M.A., Harvard University, 1992; B.S., Ohio State University, 1990.

[1] *See* Howell E. Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

[2] *See* 2003 BD. OF TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP., H.R. DOC. NO. 108-49, at 2, *available at* http://www.ssa.gov/OACT/TR/TR03/tr03.pdf [hereinafter 2003 TRUSTEES REPORT].

[3] *See, e.g.*, Christian E. Weller, Editorial, *Securing Social Security*, WASH. POST, Aug. 20, 2002, at A12; Michael Barone, *A Knack for Reframing*, US NEWS & WORLD REP., Sept. 8, 2003, at 23.

[4] *See* Kent Smetters, *Is the Social Security Trust Fund Worth Anything?*, 94 AM. ECON. REV. (forthcoming 2004) (manuscript at 11, *available at* http://irm.wharton.upenn.edu/WP-security-smetters.pdf) (last modified June 2003).

[5] *See id.*

Jackson instinctively approaches the problem as a lawyer well versed in the approach used under private pension systems: accrual accounting. He asks why the government should not be held to the same standard that it imposes on the private sector.[6] This intuition certainly has some rhetorical appeal. Indeed, in testimony before Congress, I once made the off-the-cuff remark that current federal budgetary practices would be illegal in the private sector.[7]

Accrual accounting is reasonable for a private defined-benefit pension system in which firm shareholders cannot assume that the firm will exist forever. In this environment, pension rules must be constructed to determine how much money a company must set aside in order to meet its pension obligations in case the firm goes bankrupt. Retirement benefits in private pension systems are inherently "back loaded," because benefits are based on the last few years of earnings prior to retirement, and some fairly ad hoc assumptions must be used to calculate the accrued liability; vesting requirements, in turn, necessitate additional assumptions. Except for vesting requirements, these particular technical problems are less of an issue for Social Security, in which retirement benefits are calculated using a worker's best thirty-five years of indexed earnings.[8]

More significantly, accrual accounting is less attractive for government accounting than for private accounting where it is more natural to assume that the government will last well into the future. Accrual accounting tells us nothing about Social Security's long-run sustainability. Whether Social Security will need parametric changes in order to remain sustainable depends on the infinite-horizon open-group obligation, a concept discussed by Jackson.[9] This obligation was recently estimated by the Social Security Trustees as $10.5 trillion in present value—a significant shortfall.[10] In other words, benefits must be cut or taxes increased by $10.5 trillion in present value for Social Security contributions to cover scheduled benefits.

While the infinite-horizon open-group obligation is fairly consistent with the system's accrued liability as calculated by Jackson,[11] this result is coincidental. In fact, if Congress were to decrease benefits or increase taxes enough to eliminate the long-run imbalance, the infinite-horizon

---

[6] *See* Jackson, *supra* note 1, at Part I.A.2.a.

[7] *See Balanced Budget Amendment: Hearing on H.R. Res. 22 Before the Subcomm. on the Constitution of the Comm. on the Judiciary*, 108th Cong. 26 (2003) (statement of Kent Smetters, Assistant Professor, The Wharton School, University of Pennsylvania).

[8] *See* Soc. Sec. Admin., Examples of Benefit Calculations for Workers Attaining Age 62 in 2004 (last modified Oct. 15, 2003), *available at* http://www.ssa.gov/OACT/ProgData/retirebenefit1.html (last visited Nov. 7, 2003). Social Security has its own technical problems, however; unlike the linear retirement benefit formulas in the private sector, Social Security's benefit formula is progressive, requiring some arbitrary choices when calculating accrued liability.

[9] *See* Jackson, *supra* note 1, at Part II.C.3.

[10] *See* 2003 TRUSTEES REPORT, *supra* note 2, at 62.

[11] *See* Jackson, *supra* note 1, at Part II.B.1.

open-group liability would become zero. In this case, the system would be in long-run balance, even though it continued to transfer resources from one generation to another. The accrued liability, however, would remain positive, unless Social Security were fully privatized or fully prefunded so that each generation paid for its own benefits. So these two concepts really are unrelated.

At first glance, it might appear that the accrued liability measure would indicate the amount of "implicit debt" that people alive today and in the past are passing to future generations. Accrued liability does not accurately measure "implicit debt," however, because accrued liability is inherently backward-looking, as it is calculated based on previous contributions. The advantage of being backward-looking is that it is quite easy to calculate. This ease of calculation makes it potentially suitable for certain privatization experiments when the government cannot estimate the individual productivity of each worker, as in the optimal income tax literature.[12] Alternative approaches, however, can be used to finance transitions to a privatized social security system that are less expensive than the costs suggested by accrual accounting.[13]

The cost of the simplicity of the accrued liability calculation is that it overestimates the true liability that the Social Security system is passing to future generations because the accrued measure fails to net out the future taxes that will be paid by younger and richer workers in excess of their future benefits. The more inclusive closed-group obligation measure gives the correct measure of the obligations that are being passed forward in time. This measure also corresponds more closely to the negative impact that a pay-as-you-go Social Security system has on household saving decisions and national saving.[14]

Jackson agrees that the standard accrued liability measure would overestimate the obligations that the Social Security system is passing to future generations.[15] As a result, he presents a modified accrued liability measure that appears to be very similar, if not identical, to the more standard closed-group liability measure.[16] While the usefulness of his modified measure is clear, one wonders why he then even uses the language of "accrual accounting." Why not focus the paper on the merits of the closed-group obligation measure? To be sure, the positive difference in size between the accrued liability and the closed-group liability measure reminds us that many generations alive today will receive less money in

---

[12] *See generally* J. A. Mirrlees, *An Exploration in the Theory of Optimum Income Taxation*, 38 REV. ECON. STUD. 175 (1971).

[13] *See generally* Kent Smetters & Jan Walliser, *Opting Out of Social Security*, 88 J. PUB. ECON. (forthcoming 2004), *available at* http://irm.wharton.upenn.edu/WP1-Opt-Smetters.pdf (last modified Aug. 7, 2002).

[14] *See id.* at 5.

[15] *See* Jackson, *supra* note 1, at Part II.C.5.

[16] *See id.*

present value from Social Security than they are projected to pay into the system. But the difference is spread across many living cohorts, so it is not particularly indicative of the amount of money that, for example, an individual born into the system will pay in excess of future benefits.

As noted above, the accrual accounting approach might have some rhetorical appeal. This spin does not come free. The key problem with accrual accounting is that it gives a sense that Social Security benefits accrued to date are somehow "bonded," that is, that they represent legal liabilities in the same sense as government debt.

In *Fleming v. Nestor*,[17] however, the Supreme Court recognized that a Social Security benefit represents the policy of Congress and not an express or implied contract between beneficiaries and the Social Security Administration. The Social Security Administration, therefore, picks its terminology very carefully. For example, in the Trustees Reports, future projected benefits are referred to as "scheduled benefits" instead of "promised benefits."[18] Moreover, the open- and closed-group statistics are reported as measures of "obligations" instead of "liabilities."[19] Jackson seems to brush aside the importance of these distinctions, almost taking for granted that Congress would never reduce accrued benefits. In doing so, Jackson restricts the degrees of freedom that Congress has available when making policy. Indeed, the President's recent Commission to Strengthen Social Security proposed three plans, all of which would have reduced the accrued benefits of some people in exchange for personal accounts.[20]

In contrast, the infinite-horizon open-group and closed-group obligation measures make no pretense of bonding. Instead, they simply describe the implications of the continuation of current policy. The infinite-horizon open-group obligation indicates whether current policy can continue without changes. The closed-group obligation indicates the value of obligations that are being passed to future generations if current policy were to continue. Together, these two measures make a powerful combination, and they can be integrated easily with the rest of the federal budget.[21] Both measures can also be applied to the U.S. Medicare system, which faces a long-run imbalance that is about five times larger than the Social Security system.[22] The nation's Medicaid program also faces big problems.[23]

---

[17] 363 U.S. 603, 610–11 (1960).

[18] *See, e.g.*, 2003 TRUSTEES REPORT, *supra* note 2, at 8.

[19] *See, e.g.*, *id.* at 44–61, 62, tbl.IV.B7, 63, tbl.IV.B8, 64–73.

[20] *See* President's Comm. to Strengthen Social Security, Strengthening Social Security and Creating Personal Wealth for All Americans, at 119–42 (Dec. 2001) *available at* http://www.csss.gov/reports/Final_report.pdf.

[21] *See* JAGADEESH GOKHALE & KENT SMETTERS, FISCAL AND GENERATIONAL IMBALANCES: NEW BUDGET MEASURES FOR NEW BUDGET PRIORITIES 27 (2003).

[22] *See id.*

[23] *See generally* Jan Ellen Rein, *Misinformation and Self-Deception in Recent Long-*

In fact, using the same economic and policy assumptions in the President's 2004 Budget, the U.S. federal government faces a $45 trillion fiscal imbalance across all expenditure categories, including Social Security, Medicare, Medicaid, defense, transportation, and education.[24] In other words, the current level of debt held by the public plus the present value of all future projected expenditures exceeds the present value of all future projected taxes and other receipts by $45 trillion.[25] Jackson is correct that it is important that the government make these shortfalls more transparent in the federal budget. His paper presents a very persuasive set of arguments toward this end, but accrual accounting fails to reflect adequately the current state of the Social Security system.

Term Care Policy Trends, 12 J.L. & POL. 195 (1996); William Alvarado Rivera, *A Future for Medicare Managed Care: The Lessons of California's San Mateo County*, 7 STAN. L. & POL'Y REV. 105 (1995-96).

[24] *See* GOKHALE & SMETTERS, *supra* note 21, at 27.

[25] *See id.*

REPLY

HOWELL E. JACKSON

I am grateful to all of the commentators for their thoughtful and probing responses to my Article. Their analyses clarify both the advantages and challenges of moving toward a system of modified accrual accounting for Social Security and our budgetary processes more generally. In the same spirit, I would like to reply to a few of the points raised.

First, Professor Clark asserts that reforms in the 2003 Trustees Reports obviate the need for the reforms I propose.[1] As the Clark comment stresses and as I noted in my Article, the 2003 Trustees Report included for the first time alternative presentations of trust fund solvency, including estimates of the trust funds' open-group and closed-group liability. Unlike Professor Clark, I regard this change as modest and of little importance to the public debate over Social Security reform.

Let me first address the modesty of the 2003 changes. As my Article noted, the change appeared sixty pages into the 2003 Trustees Report and was limited to two and a half pages of analysis. The information presented in the analysis falls short of the FASAB disclosure requirements for Social Insurance programs—that is, the disclosures do not satisfy generally accepted accounting standards for the federal government. In addition, the disclosures do not contain any estimates of accrued liabilities, which my proposal would emphasize and which Professor Clark admits are "an important and useful indicator of the financial status of Social Security."[2] Moreover, if one reviews the introductory materials to the 2003 Trustees Report, the inconsequential nature of the reforms becomes apparent. While the materials do report for the first time the net present value estimate of the open-group seventy-five-year horizon liability, the information is imbedded into a format that is identical to past reports and conveyed in a way unlikely to attract the attention of anyone but Social Security aficionados.[3] Neither the 2003 Trustees Report itself

---

[1] *See* Robert L. Clark, *Liabilities, Debts, Revenues, and Expenditures: Accounting for the Actuarial Balance of Social Security*, 41 HARV. J. ON LEGIS. 161 (2004).

[2] *See id.* In the Spring of 2003, I wrote Professor Clark in his capacity as chair of the Technical Panel on Assumptions and Methods, outlining the additional information that the trustees' report would need to include to allow outside analysts to derive more complete accrual accounting statements for Social Security. *See* Memorandum from Howell Jackson to Robert Clark (May 13, 2003), *available at* http://www.law.harvard.edu/faculty/hjackson/projects.

[3] Reproduced below are the two key sections of the introductory materials of the 2003 Trustees Report:

> The combined OASDI Trust Funds are projected to become insolvent in 2042 under the long-range intermediate assumptions. For the trust funds to remain solvent throughout the 75-year projection period, the combined payroll tax rate

nor the press release accompanying its release in March of 2003 made any claim that the content of the report had been altered to any noteworthy degree.[4] Even more telling, I have found no general press accounts after the release of the 2003 Trustees Reports that have made any mention of the fact the report included important new materials, much less that the change was in some way significant.[5] Indeed, press coverage of the 2003 Trustees Report tracked the traditional measures of performance that the trustees' reports have included for many years.

Second, several of the commentators disagree with me over the desirability of distinguishing between accrued benefits and benefits to-be-accrued in the future. I personally find it useful to quantify the cost of retirement benefits that the system has incurred to date and will not restate my reasons here. I would, however, like to address some of the political aspects of quantifying these liabilities. As Professor Clark notes and as I acknowledged, we have reduced accrued benefits in the past, most notably in the 1983 reforms.[6] To a large degree these reductions undid retroactive enhancements in benefits from the preceding decade. Without revisiting the wisdom of the 1983 reforms, these retroactive reductions in benefits have complicated reform efforts in recent years. Seniors are now suspicious that "Social Security Reform" is just a code word for retroactive reductions of benefits. Professor Clark, for one, recognizes this difficulty when he notes "virtually all reform proposals be-

---

> could be increased immediately by 1.92 percentage points, benefits could be reduced immediately by 13 percent, a transfer of $3.5 trillion in general revenue (in net present value) could be made, or some combination of approaches could be adopted. Significantly larger changes, would be required to achieve solvency beyond 75 years . . . .
>
> Another way to illustrate the financial shortfall of the OASDI system is to examine the cumulative value of taxes less costs, in present value. Figure II.D4 shows the present value of cumulative OASDI taxes less costs over the next 75 years. The balance of the combined trust funds peaks at $2.3 trillion in 2017 (in present value) and then turns downward. Through the end of 2077, the combined funds have a present-value unfunded obligation of $3.5 trillion . . . .
>
> The open group unfunded obligation over the 75-year projection period has increased from $3.3 trillion to $3.5 trillion.

2003 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 3, 10, 12 [hereinafter 2003 TRUSTEES REPORT]. Some information that Professor Clark cites as new to the 2003 report actually appeared in earlier reports. For example, prior reports had routinely reported how much benefits would need to be cut after the trust fund exhaustion date in order to equal projected revenues. *See, e.g.*, 2002 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 16–18 [hereinafter 2002 TRUSTEES REPORT].

[4] *See* Press Release, Social Security Administration, Social Security Not Sustainable for the Long Term (Mar. 17, 2003), *available at* http://www.ssa.gov/pressoffice/pr/trustee03-pr.htm.

[5] *See* Howell Jackson, *Accounting for Social Security and Its Reform*, 41 HARV. J. ON LEGIS. 59 (2004).

[6] *See id.* at Part II.C.2; Clark, *supra* note 1.

gin by stating that current retirees will not be affected," but then he and others resist my suggestion that the financial statements of Social Security should recognize these obligations of liabilities. One way to address the legitimate concerns of seniors and to facilitate sensible reforms is to recognize their claims as financial liabilities of the system and then move on to other issues.

To a large degree, I suspect, it is the recognition of accrued liabilities to mid-career workers that the commentators really oppose. As several comments noted,[7] the recognition of accrued liabilities for mid-career workers may make it more difficult for politicians to reduce these benefits. Accrual accounting would not, however, constitute an absolute barrier to retroactive adjustment in accrued benefit of mid-career workers, and such reduction could make sense particularly if these workers were given access to other forms of retirement savings such as individual accounts. To be sure, there is some tension between my proposing the recognition of these liabilities and then leaving open the possibility that they might be adjusted. The goal of liability recognition, however, is to convey the accretion of *probable* economic sacrifices over time.[8] Mid-career workers are earning statutory entitlements under the Social Security Act and, with the passage of time, these entitlements will mature into politically binding obligations once these workers reach retirement age. Any accounting system that does not recognize the gradual accrual of these obligations over time is substantially misleading, at least in my view.

One lesson that emerges from the foregoing debate is the impossibility of articulating a fully unbiased estimate of the present value of future Social Security benefits.[9] If one focuses on absolute legal entitlements, then future benefits have no present value, because Congress could eliminate the program completely.[10] Instead, one could focus on accrued benefits, as I propose, on the ground that this system best comports with statutory formulae and best approximates political reality. A third option, following President Bush's Commission,[11] analyzes payable benefits—that level of current benefits that the projected revenues would cover over the next seventy-five years or some other period of time. Fourth, as traditionalists prefer,[12] one could focus on scheduled benefits,

---

[7] *See* Kent Smetters, *Inadequacies of Accrual Accounting in Social Security*, 41 HARV. J. ON LEGIS. 215 (2004); Peter A. Diamond & Peter R. Orszag, *Accrual Accounting for Social Security*, 41 HARV. J. ON LEGIS. 173 (2004).

[8] Contrary to Professor Clark's suggestion, *see* Clark, *supra* note 1, recognition of liabilities under accrual accounting does not require certainty of payment.

[9] *But see id.*

[10] Professor Smetters seems to advance this view, *see* Smetters *supra* note 7, although his preferred measures of fiscal imbalance rely on present value calculations of scheduled benefits and revenues over the infinite horizon.

[11] *See generally* THE PRESIDENT'S COMM'N TO STRENGTHEN SOC. SECURITY, REPORT OF THE PRESIDENT'S COMMISSION: STRENGTHENING SOCIAL SECURITY AND CREATING PERSONAL WEALTH FOR ALL AMERICANS (2001).

[12] *See, e.g.*, Clark, *supra* note 1; Diamond & Orszag, *supra* note 7.

because that is the level of benefits that the statutory structure currently mandates. All are defensible approaches and each has an element of bias.

Finally, the overarching question at issue in this debate asks which financial format the trustees should use to report the financial performance of the trust funds. The traditional format features the system's net change in assets as a short-term measure of performance along with the seventy-five-year actuarial deficit and trust fund exhaustion dates as measures of long-term solvency. Modified accrual accounting highlights the system's annual profit or loss and the level of implicit trust debt as a percentage of GDP. The accompanying table presents these measures under both approaches for the past six years.[13] Which format most effectively communicates the financial condition of the trust funds?

Under the modified accrual accounting approach, the message for the general public is clear. The Social Security trust funds are incurring annual substantial losses—on the order of hundreds of billions of dollars a year—and the level of implicit trust fund debt is increasing as a percentage of GDP. The take home lesson to the general public is that Social Security is losing ground fast and its implicit debts are mounting rapidly. The case for reform is clear.

The traditional measures convey a very different message. Net assets have increased markedly in each of the last six years, rising $164.4 billion in 2002. The seventy-five-year actuarial deficit, measured as a percentage of total payroll, has declined (that is, improved) in four of the last six years, and the deficit at the end of 2002 was actually a good deal lower than it was at the of 1997. Over the same period, the projected trust fund exhaustion date has actually been pushed back in each of the last six years. At year-end 2002, the exhaustion date was a full decade later than it was five years earlier. One reason the trustees have had such difficulty in communicating the seriousness of the crisis in Social Security finances is that the financial statement they produce each year highlights financial measures that imply the systems finances are, at the very least, holding their own. The crisis does not seem imminent.

With financial statements, there must be a bottom line. The bottom line for Social Security's current financial statements should be substantially and unambiguously negative. The traditional presentation does not satisfy this criterion. A modified system of accrual accounting would.

---

[13] For the source of the data in the table, see 2003 TRUSTEES REPORT, *supra* note 3, at 2, 4; 2002 Trustees Report, *supra* note 3, at 6, 16; 2001 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 2, 4; 2000 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 4, 5, 24; 1999 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 2–4, 23; 1998 TRS. OF THE FED. OLD-AGE AND SURVIVORS INS. AND DISABILITY INS. TRUST FUNDS ANN. REP. 2-4, 23.

	Traditional Measures			Modified Accrual Accounting	
Year-End	Net Change in Assets ($ billion)	Actuarial Deficit (% of payroll)	Year of Trust Fund Exhaustion	Annual Loss ($ billion)	Implicit Debt (% of GDP)
1997	+ $ 88.6	– 2.19%	2032	– $515	97.7%
1998	+ $107.0	– 2.07%	2034	– $287	95.8%
1999	+ $133.7	– 1.89%	2037	– $550	96.8%
2000	+ $153.3	– 1.86%	2038	– $761	98.9%
2001	+ $163.1	– 1.87%	2041	– $535	101.1%
2002	+ $165.4	– 1.92%	2042	– $370	101.7%

ARTICLE

SEVERABILITY, INSEVERABILITY, AND THE RULE OF LAW

MICHAEL D. SHUMSKY*

The pending decision on the constitutionality of the Bipartisan Campaign Finance Reform Act (BCRA) prompts the question of whether the Supreme Court should allow BCRA's increases in hard-money spending limits to stand if BCRA's ban on soft-money is declared unconstitutional—that is, whether the Court should "sever" the soft-money ban from any remaining constitutional components of this sweeping legislation. In this Article, Michael D. Shumsky argues that, when confronted with unambiguous directives either to sever or to invalidate a statute in its entirety, federal courts are bound by constitutional norms to give full effect to these statutory provisions, including the severability clause Congress enacted as part of BCRA. In the absence of a clear congressional directive regarding severability, Shumsky contends that principles of judicial restraint point toward severing partially unconstitutional statutes.

TABLE OF CONTENTS

I. Introduction 228
II. Severability: A Doctrinal History 232
A. Early Cases Addressing Severability in the Absence of Legislative Direction 232
B. Early Caselaw Addressing Severability Clauses 234
C. The Alaska Airlines *Formulation* 240
D. Recent Caselaw Addressing Inseverability Clauses 243
III. Alaska Airlines*'s Unjustified Approach to Severability Clauses* 245
A. Article I, Section 7 as a Binding Rule of Recognition 245
B. Severability Clauses and the Judicial Power 249
C. Severability Clauses and Non-Delegation 252

* Law Clerk to the Hon. Diarmuid F. O'Scannlain, U.S. Court of Appeals for the Ninth Circuit, 2003-2004 term. A.B., Harvard College, 2000; J.D., Yale Law School, 2003. I owe special thanks to William N. Eskridge, Jr. for his guidance and support, and to Eric Shumsky for helpful criticisms and editorial assistance. I benefited greatly from conversations on this topic with Akhil Amar, Neal Katyal, and Jerry Mashaw, and from Judge Guido Calabresi's excellent course "The Role of Courts in the Age of Statutes." I worked on unrelated aspects of the legal challenge to McCain-Feingold while employed at Kirkland & Ellis during the summer of 2002. All views expressed herein are my own and do not necessarily reflect those of any party to *McConnell v. FEC* or its companion cases. A draft version of this Article was awarded Yale Law School's 2003 Thomas I. Emerson Prize for the best student paper or project on a subject related to legislation.

D. *Severability and Statutory Functionality* *255*
E. *Severability and Changed Circumstances* *259*
IV. *Special Problems with Non-Enforcement of Inseverability Clauses* *267*
A. *The Inseverability Clause as Structural Enforcement Mechanism* *267*
B. *The Countermajoritarian Consequences of Ignoring Inseverability Clauses* *268*
C. *Could Congress Repeal the Remaining Statutory Structure?* *269*
V. *Severability in the Absence of Explicit Legislative Guidance* *271*
A. *Toward a Clear Statement Rule Favoring Severability* *272*
B. *Cautions Against a Clear Statement Rule Favoring Severability* *276*
VI. *Conclusion* *279*

I. Introduction

The legislative bargain struck at the heart of the McCain-Feingold campaign-finance reform legislation was simple: In exchange for agreeing to a ban on soft-money contributions, a majority of Members of Congress successfully insisted upon increasing then-existing hard-money contribution limits.[1] At the time it was made, this deal had the potential to lead to a surprisingly difficult and important question of statutory interpretation: If the Supreme Court invalidated McCain-Feingold's soft money ban on First Amendment grounds, what should it have done about the legislation's accompanying increase in allowable hard-money contributions?

As it happened, the Supreme Court upheld the ban, so it never answered this question.[2] The special panel of district judges that first considred the law, however, divided on the issue, and those judges' approaches suggest some of the ways severability might be addressed.[3] One approach

[1] *See* Bipartisan Campaign Finance Reform Act ("BCRA"), Pub. L. No. 107-155, §§ 101–103, 307, 116 Stat. 81, 82–88, 102–03 (2002). For a discussion of the tradeoff between hard-money and soft-money contribution limits, see Richard Briffault, *The Future of Reform: Campaign Finance After the Bipartisan Campaign Reform Act of 2002*, 34 Ariz. St. L.J. 1179, 1214–15 (2002).

[2] *See* McConnell v. FEC, Nos. 02-1674, 02-1675, 02-1676, 02-1702, 02-1727, 02-1733, 02-1734, 02-1740, 02-1747, 02-1753, 02-1755, 02-1756, 2203 WL 22900467, at *29–*38 (U.S. December 10, 2003). Although the Court did strike down two minor provisions of the law, it simply assumed without discussion that they could be severed. *See id.* at *63–*65 (invalidating a provision of BCRA requiring political parties to choose between coordinated and independent expenditures after nominating a candidate); *72–*73 (invalidating a provision prohibiting contributions by minors).

[3] *See* McConnell v. FEC, 251 F. Supp. 2d 176 (D.D.C. 2003). With each judge issuing his or her own opinion (and two judges jointly submitting a fourth "per curiam" opinion and order), the deeply fractured panel upheld parts of the law's ban on raising and spend-

could have been to ask whether Congress would have independently increased hard money limits if it had not also banned soft money. The answer is probably not, and therefore one may have suggested that the Court should excise the contribution limit increase from McCain-Feingold at the same time that it struck down the law's soft money ban. This line of reasoning seems to be what most influenced Judge Henderson at the lower-court level.[4]

Another approach might have been to question whether the remaining statute—including its increase in hard money contribution limits—could function without a soft money ban, or whether that ban was so integral to the statutory scheme that none of its other provisions could operate independently. In this case, the remaining scheme surely could operate independently from the soft money ban, and therefore, one could have argued that those aspects of the statute not dependent on the soft money ban, including the contribution-limit increase, ought to have been preserved.

A third approach—and perhaps the most obvious one—would have been to ask whether Congress itself addressed the possibility of partial unconstitutionality in crafting McCain-Feingold. Cognizant that courts routinely must decide whether to sever or strike down entirely a partially unconstitutional law, Congress frequently enacts a severability or inseverability clause designed to govern the reviewing court's severability determination should the court declare part of a statute constitutionally invalid. And that is exactly what Congress did in constructing McCain-Feingold. Section 401 of the legislation provides:

> If any provision of this Act or amendment made by this Act, or the application of a provision or amendment to any person or circumstance, is held to be unconstitutional, the remainder of this Act and amendments made by this Act, and the application of the provisions and amendment to any person or circumstance, shall not be affected by the holding.[5]

In fact, severability was a critical issue during the debate over McCain-Feingold. After two Republican senators attempted to insert an insever-

---

ing soft money but struck down numerous others, and it retained the law's hard money increases without reaching their constitutionality.

[4] *See McConnell*, 251 F. Supp. 2d at 270 n.5. (Henderson, J.) ("I am convinced that no plaintiff has standing to challenge the increased contribution limits set out in BCRA sections 304, 307, 316 and 319. Therefore, I would not decide the constitutionality of those provisions even though upon examination of the record and *despite BCRA's severability provision, I doubt that the Congress, upon elimination of the numerous provisions I believe are invalid, would have been satisfied with the contribution limit increases.*") (citations and quotation omitted) (emphasis added); *see also id.* at 266 ("Further, the opinions are similarly flawed in their dissection of the statute's dense and interlocking provisions, upholding a portion here and striking down a fragment there until they have drafted legislation the Congress would never have enacted . . . .").

[5] BCRA § 401.

ability clause into the legislation,[6] the bill's sponsors and the Democratic leadership strenuously and successfully campaigned for passage of the severability clause that became Section 401.[7] At the time, the Senate's debate over severability was considered so important to the legislation—which was certain to face a constitutional challenge—that the *New York Times* published excerpts of that debate the following day.[8] Congress thus purposefully directed reviewing courts to preserve all parts of the statute that are constitutional; if the Supreme Court had held that McCain-Feingold's soft money ban violated the First Amendment, its duty would have been to preserve the law's increase in hard money limits in accord with Congress's clearly expressed intent.

Not so fast, the Court's severability doctrine surprisingly teaches. Despite the unambiguous command of severability and inseverability clauses, the Court has repeatedly held that they create only a rebuttable presumption that guides—but does not control—a reviewing court's severability determination. Instead of deferring to Congress's clearly expressed preference either for or against severability, and thereby allowing the legislature confidently to control the form of the statutory schemes it creates, the Court has chosen instead to focus on extrinsic evidence of legislative intent and on the potential functionality of the post-severance statutory scheme in evaluating whether to sever an unconstitutional provision from the surrounding statute.[9] This appears to be how Judges Leon and Kollar-Kotelly approached the issue of severability in their McCain-Feingold opinions.[10]

That this decidedly atextual approach remains dominant within the judiciary is particularly startling given the Court's increasing reliance on

---

[6] *See* 147 CONG. REC. S3084, S3088–90 (daily ed. Mar. 29, 2001) (statement of Sen. Frist).

[7] *See, e.g.*, 147 CONG. REC. S3084, S3095 (daily ed. Mar. 29, 2001) (statement of Sen. Feingold); *see also* Press Release, Sen. Tom Daschle, News Briefing Regarding BCRA (Mar. 28, 2001) (on file with Fed. Document Clearing House).

[8] *See Excerpts From Senate Debate on Donations: Skirmishing and Predictions*, N.Y. TIMES, Mar. 30, 2001, at A16.

[9] *See, e.g.*, Alaska Airlines v. Brock, 480 U.S. 678, 686 (1987) (holding that the presence of a severability clause merely creates a rebuttable presumption of severability); INS v. Chadha, 462 U.S. 919, 931–34 (1983) (acknowledging the presence of a severability clause but finding it necessary to examine the act's legislative history before severing its unconstitutional legislative veto provision from the remainder of the act); United States v. Jackson, 390 U.S. 570, 586 n.27 (1968) ("[T]he ultimate determination of severability will rarely turn on the presence or absence of such a clause."); Biszko v. RIHT Fin. Corp., 758 F.2d 769, 773 (1st Cir. 1985) ("[A] non-severability clause cannot ultimately bind a court . . . .").

[10] *See* McConnell v. FEC, 251 F. Supp. 2d 176, 776 (D.D.C. 2003) (Leon, J.) ("[A severability] clause only creates a presumption of severability. It does not relieve this Court of its obligation to determine if the [remaining statutory provisions] can stand alone, and if Congress would have enacted [them] knowing [other sections] would be held unconstitutional."); *id.* at 435 (Kollar-Kotelly, J.) ("The provisions I have found unconstitutional are all provisions of BCRA that are not central to its core mission and are entirely severable without doing injustice to the remainder of the law.").

a more robust textualism in deciding statutory cases.[11] Likewise, the lack of a contemporary scholarly focus on severability is puzzling during a time in which statutory interpretation has emerged not only as a distinct field of academic inquiry, but as one with burgeoning prominence within the legal academy.[12] To this day, the leading article on severability doctrine and theory is Robert Stern's 1937 *Separability and Separability Clauses in the Supreme Court*,[13] though John Copeland Nagle's now decade-old *Severability*[14] follows close behind. Outside of these two articles, however, severability remains largely ignored as a subject of sustained theoretical inquiry[15] and has faded almost completely from the academy's radar—if it is fair to say that it was ever really on the academy's radar.[16]

This is a fundamentally troubling oversight. As this Article will demonstrate, severability doctrine is intimately connected to a number of critical issues at the heart of the Constitution's separation of powers. These include the debates over competing paradigms of statutory interpretation and interpretive practices (e.g., textualism, dynamic statutory interpretation, the use of legislative history); the appropriate scope of judicial review; non-delegation; and key elements of the Article III jurisdictional requirements. The aim of this Article is to explore these thematic connections by highlighting crucial deficiencies in the Supreme Court's approach to severability, and to develop an alternative theory of severabil-

---

[11] *See generally* William N. Eskridge, Jr., *The New Textualism*, 37 UCLA L. REV. 621 (1990).

[12] *See* John Copeland Nagle, *Newt Gingrich: Dynamic Statutory Interpreter*, 143 U. PA. L. REV. 2209, 2210–11 (1995) (reviewing WILLIAM N. ESKRIDGE, JR., DYNAMIC STATUTORY INTERPRETATION (1994)) ("A long list of leading scholars have turned their attention to the theory and practice of statutory interpretation during the 1980s and 1990s, and three of the Justices now sitting on the Supreme Court have written academic works in the field.").

[13] Robert L. Stern, *Separability and Separability Clauses in the Supreme Court*, 51 HARV. L. REV. 76 (1937).

[14] John Copeland Nagle, *Severability*, 72 N.C. L. REV. 203 (1993).

[15] An exception is Adrian Vermeule, *Saving Constructions*, 85 GEO. L.J. 1945 (1997), which focuses more on the relationship between the Court's general presumption of severability and its use of the canon of constitutional avoidance than on the separation of powers issues at stake in the Court's treatment of severability clauses. The leading casebook on statutory interpretation devotes only about a single page to severability, *see* WILLIAM N. ESKRIDGE, JR. & PHILIP P. FRICKEY, LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 506, 933–34 (2d ed. 1995), as does the leading casebook on federal jurisdiction. *See* RICHARD H. FALLON ET AL., HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 197–99 (4th ed. 1996).

[16] *INS v. Chadha*, 462 U.S. 919 (1983), which invalidated the single-house veto clause contained in the Immigration and Nationality Act, 8 U.S.C. § 1254(c)(2) (1982), did quickly prompt a number of student notes addressing severability, but it occasioned no sustained scholarly discussion of severability theory. *See, e.g.*, Note, Steven W. Pelak, *The Severability of Legislative Veto Provisions: An Examination of the Congressional Budget and Impoundment Control Act of 1974*, 17 U. MICH. J.L. REFORM 743 (1984); Note, *The Aftermath of* Chadha: *The Impact of the Severability Doctrine on the Management of Intergovernmental Relations*, 71 VA. L. REV. 1211 (1985); Note, *Severability of Legislative Veto Provisions: A Policy Analysis*, 97 HARV. L. REV. 1182 (1984).

ity—one more firmly rooted in our constitutional tradition and that more faithfully reflects the principle of separated powers.

The Article proceeds as follows: Part II outlines the awkward judicial development of the current severability and inseverability doctrines, tracing the early federal preference for holding statutes severable through the development of severability clauses (and corresponding doctrinal upheaval of the early twentieth century) and into the modern-day formulation. Part III critiques the courts' disregard for severability clauses, arguing that because those clauses have been through the Article I, Section 7 process, they are authoritative laws that bind judges confronting a partially unconstitutional statute. In doing so, the Article addresses a number of challenges to this textualist approach to severability, including constitutional claims that severability clauses violate the separation of powers and prudential arguments for evaluating severability in light of changed circumstances. Part IV focuses on the even greater dangers inherent in modeling an approach to inseverability clauses on the Supreme Court's current approach to severability. Doing so would offend the separation of powers not only because inseverability clauses have been through the Article I, Section 7 process, but also because disregarding them risks leaving in force a statutory scheme that neither the enacting Congress nor the President would have supported, with no way for those lawmaking branches to reverse the results. Part V moves beyond the textualist case for addressing severability and inseverability and argues that, in the absence of a clear congressional statement against severability, courts should hold statutes severable.

II. Severability: A Doctrinal History

A. *Early Cases Addressing Severability in the Absence of Legislative Direction*

No early federal statutes contained severability clauses, and in its first cases exercising judicial review, the Supreme Court seems to have simply assumed that a constitutionally offensive provision could be severed from an otherwise enforceable statute. In *Marbury v. Madison*,[17] for instance, Chief Justice Marshall referred only to a portion of Section 13 of the Judiciary Act of 1789, and the remainder of the Act stayed in force after the Court's decision. Making explicit this early assumption that partially unconstitutional statutes were to be severed, the Court declared in 1829 that "[i]f any part of the act be unconstitutional, the provisions of that part may be disregarded while full effect will be given to such as are not repugnant to the constitution of the United States . . . ."[18]

---

[17] 5 U.S. (1 Cranch) 137, 173 (1803).
[18] Bank of Hamilton v. Lessee of Dudley, 27 U.S. 492, 526 (1829).

It did not take long for an eventually vexing exception to emerge. As Professor Stern first noted,[19] the 1854 Massachusetts case *Warren v. Mayor & Aldermen of Charlestown*[20] is particularly important to the development of the modern severability doctrine. After holding that part of a state statute annexing Charlestown to Boston violated the federal Constitution by (among other things) effectively denying federal representation to the citizens of Charlestown, Chief Justice Shaw of the Supreme Judicial Court of Massachusetts reasoned for the Court that the presumption of severability

> must be taken with this limitation, that the [statute's] parts, so held respectively constitutional and unconstitutional, must be wholly independent of each other. [For] if they are so mutually connected with and dependent on each other, as conditions, considerations or compensations for each other, as to warrant a belief that the legislature intended them as a whole, and that, if all could not be carried into effect, the legislature would not pass the residue independently, and some parts are unconstitutional, all the provisions which are thus dependent, conditional or connected, must fall with them.[21]

Shaw then concluded that because the "various provisions of the act, . . . all providing for the consequences of . . . annexation, . . . are connected and dependent[,] . . . look to one object and its incidents, and are so connected with each other," the legislature could not have intended the dysfunctional but constitutionally valid statutory remnants to stay in force.[22] Shaw therefore struck down the entire statute.[23]

Approximately a quarter of a century later, the U.S. Supreme Court had adopted this reasoning, citing *Warren* for the proposition that "[t]he point to be determined in all such cases is whether the unconstitutional provisions are so connected with the general scope of the law as to make it impossible, if they are stricken out, to give effect to what appears to

---

[19] Stern, *supra* note 13, at 79–80; *see also* Nagle, *supra* note 14, at 211–13.

[20] 68 Mass. (2 Gray) 84 (1854).

[21] *Id.* at 99.

[22] *Id.* at 100. Elsewhere, citing *Warren*, Chief Justice Shaw would write:

> It is undoubtedly a correct rule of construing a statute in reference to its constitutionality, to consider only such part void as is plainly repugnant to the Constitution; and therefore, where there are different provisions in the same statute, so distinct and independent, that the one may not have been the motive or inducement to the other, one may be held valid and the other void . . . . [But where] one may have been the motive, inducement or consideration on which the other was founded, . . . they must stand or fall together.

Jones v. Robbins, 74 Mass. 329, 338–39 (1857) (citations omitted).

[23] *Warren*, 68 Mass. at 107.

have been the intent of the legislature."[24] Thus, by 1880, the federal inquiry into statutory severability required a court to consider extrinsic—that is, non-text based—sources in order to gauge legislative intent—specifically, whether the constitutional provisions of a partially invalid statute could function independently of their unconstitutional counterparts such that Congress could have intended them to stand alone.

B. Early Caselaw Addressing Severability Clauses

Just as the Court appeared to be finalizing its approach to severability, it had to address a novel development: severability clauses. These clauses first appeared during the late nineteenth century and had become fairly common by 1910,[25] probably in response to the courts' search for legislative intent in extrinsic sources and their increasing tendency to declare partially unconstitutional statutes inseverable.[26] Most of these clauses, like the one at issue in the *Ohio Tax Cases*,[27] stated simply,

> The sections of this act, and every part of such sections, are hereby declared to be independent sections and parts of sections, and the holding of any section or part thereof to be void or ineffective shall not affect any other section or part thereof.[28]

When challenges first arose to the validity of these clauses, both state and federal courts enforced them without hesitation.[29]

In the 1922 case *Hill v. Wallace*,[30] however, the Supreme Court limited the enforceability of a severability clause for the first time. Although the Court seemed to recognize that a severability clause "undoubtedly . . . furnishes assurance to courts that they may properly sustain separate sections or provisions of a partly invalid act without hesitation or doubt as to whether they would have been adopted, even if the legislature had

---

[24] Allen v. Louisiana, 103 U.S. 80, 84 (1880).

[25] Nagle, *supra* note 13, at 222.

[26] Alfred Hayes, Jr., *Partial Unconstitutionality with Special Reference to the Corporation Tax*, 11 COLUM. L. REV. 120, 124–25 n.8 (1911) (collecting late 19th and early 20th century cases declaring statutes partially unconstitutional and indicating the courts' decisions with regard to severability); Stern, *supra* note 13, at 107–08 n.138 (same).

[27] 232 U.S. 576 (1914).

[28] 102 Ohio Laws 224 § 160, *quoted in Ohio Tax Cases*, 232 U.S. at 594.

[29] *See, e.g.*, *Ohio Tax Cases*, 232 U.S. at 594. ("Finally, it is contended that the act is unconstitutional because of the severity of the penalties imposed for withholding the tax. But these actions do not involve any present attempt to enforce the penalties; and the act contains a [severability clause]. The penalty clauses, if themselves unconstitutional, are severable, and there is therefore no present occasion to pass upon their validity."). *See also* State *ex rel.* Clarke v. Carter, 56 So. 974, 977 (Ala. 1911); *In re* Questions of the Governor, 123 P. 660, 662 (Colo. 1912); State *ex rel.* Davis-Smith Co. v. Clausen, 117 P. 1101, 1114 (Wash. 1911).

[30] 259 U.S. 44, 68–70 (1922) (invalidating various regulations and penalties contained in the Future Trading Act, Pub. L. No. 67-66, 42 Stat. 187 (1921)).

been advised of the invalidity of part,"[31] it nonetheless concluded: "Section 4, with its penalty to secure compliance with the regulations of [B]oards of [T]rade, is so interwoven with those [unconstitutional] regulations that they cannot be separated. None of them can stand."[32]

At first glance, it might appear that the *Hill* Court merely transplanted *Warren*'s then-familiar functionality test into the context of severability clauses. On closer examination, however, it becomes clear that *Hill* worked a critical shift in the *Warren* formulation. In *Warren*, statutes lacking a severability clause were deemed presumptively severable unless the remaining statutory structure could not function, because the residual statute's inability to operate constituted evidence that the legislature would not have wanted the statute to remain only partially in force. In *Hill*, by contrast, there was no need to search for legislative intent in extrinsic sources—after all, the legislature had enacted a severability clause explicitly conveying its intent, and the Court had recognized that the severability clause conclusively demonstrated that intent. With their decision in *Hill*, then, the Justices seemed to go from looking for legislative intent to defying it. The Court failed to provide a thorough explanation of why it retained a stand-alone inquiry into the functionality of the remaining statutory scheme. Part II.B will suggest that politics may have played a role.[33]

Indeed, *Hill* signaled the beginning of an era characterized by a troubling judicial eagerness to look beyond severability clauses. In *Dorchy v. Kansas*,[34] decided shortly after *Hill*, the Court emphasized that a severability clause "provides a rule of construction which may sometimes aid in determining [legislative] intent. But it is an aid merely; not an inexorable command."[35] And then in *Williams v. Standard Oil Co.*,[36] the Justices combined the *Warren* and *Hill* tests to undercut the force of severability clauses further. After declaring unconstitutional parts of a state statute regulating gasoline prices and containing a severability clause, the *Williams* Court began its discussion of severability by citing the above-quoted language from *Dorchy*.[37] Then, pointing to a single 1903 opinion of the New Jersey Court of Errors and Appeals, the Court declared that "[i]n the absence of such a legislative declaration, the presumption is that the legislature intends an act to be effective as an entirety"[38] and held that a severability clause merely reverses this presumption:

---

[31] *Id.* at 71.
[32] *Id.* at 70.
[33] *See infra* text accompanying notes 47–56.
[34] 264 U.S. 286 (1924).
[35] *Id.* at 290.
[36] 278 U.S. 235, 241–42 (1928).
[37] *Id.* at 241.
[38] *Id.* (quoting Riccio v. Mayor of Hoboken, 55 A. 1109, 1113 (N.J. 1903)).

> That is to say, we begin, in the light of the [severability] declaration, with the presumption that the legislature intended the act to be divisible; and this presumption must be overcome by considerations which make evident the inseparability of its provisions *or* the clear probability that[,] the invalid part being eliminated[,] the legislature would not have been satisfied with what remains.[39]

Williams had far-reaching consequences. Not only did it jettison the Court's longstanding presumption of severability; it limited the effect of statutory severability clauses to a greater degree than before. Under *Williams*, courts could disregard severability clauses not only if the remaining statutory structure could not function (as under *Hill*), but also if there was strong evidence that Congress would have been dissatisfied with the remaining statute. This latter proposition appears to have been based on *Warren*'s general notion that legislative intent controlled the severability determination—but under *Warren*, of course, such a search for legislative intent was necessary only because legislatures had not yet developed a clear, textual means to indicate their preference—severability clauses. Once the *Hill* Court held that the presence of a severability clause conclusively demonstrated the legislature's preference for severability, it is questionable why the *Williams* Court found necessary a *Warren*-like search for legislative intent beyond *Hill*'s stand-alone inquiry into post-severance statutory operability. The Court offered no rationale for retaining the inquiry, and it failed to explain where litigants would be able to find evidence of legislative intent sufficient to contravene an explicit severability clause.

Over the next decade, the Court put the final touches on its early severability jurisprudence—at least partially reversing its course twice more. In *Champlin Refining Co. v. Corporation Commission*,[40] the Court held:

> The unconstitutionality of a part of an Act does not necessarily defeat or affect the validity of its remaining provisions. *Unless it is evident that the legislature would not have enacted those provisions which are within its power*, independently of that which is not, the invalid part may be dropped *if what is left is fully operative as a law*.[41]

Contrary to *Williams*, partially invalid statutes would once again be presumed severable, but this presumption would hold only (1) in the absence of evidence that the legislature would not have enacted the remaining

---

[39] *Id.* at 242 (emphasis added).
[40] Champlin Refining Co. v. Corp. Comm'n, 286 U.S. 210 (1932).
[41] *Id.* at 234 (emphasis added).

provisions on its own and (2) if the remaining statutory provisions could function independently. Of special note here is the fact that the *Champlin* Court imported the bifurcated *Warren* test it had created in *Hill* and *Williams* back into the original *Warren* context: Under *Champlin*, *either* contrary legislative intent *or* the inability of the remaining statutory scheme to function could defeat the general presumption of severability accorded statutes lacking a severability clause. But once again, where were litigants to find evidence of legislative intent apart from the ability of the remaining statutory scheme to function, which now constituted an independent test divorced from any inquiry into legislative intent?

In *Carter v. Carter Coal Co.*,[42] the Court again flipped the reigning presumption on its head by restoring *Williams*'s general presumption of inseverability.[43] Critically, the Court also began to show litigants how to seek out independent evidence of legislative intent contrary to an explicit severability clause (aside from the ability of the remaining statute to function). In the Court's words:

> [T]he presumption [created in favor of severability can] be overcome by considerations which establish . . . "the clear probability that the legislature would not have been satisfied . . . unless it had included the invalid part." Whether the provisions of a statute are so interwoven that one being held invalid the others must fall, presents a question of statutory construction and of legislative intent, to the determination of which the [severability clause] becomes an aid. "But it is an aid merely; not an inexorable command." The presumption in favor of se[ve]rability does not authorize the court to give the statute "an effect altogether different from that sought by the measure viewed as a whole."
>
> The [severability clause] in no way alters the rule that in order to hold one part of a statute unconstitutional and uphold another part as separable, they must not be mutually dependent upon one another. Perhaps a fair approach to a solution of the problem is to suppose that while the bill was pending in Congress a motion to strike out the labor provisions had prevailed, and to inquire whether, in that event, the statute should be so construed as to justify the conclusion that Congress, not-withstanding,

---

[42] 298 U.S. 238 (1936).

[43] *Id.* at 312 ("In the absence of [a severability clause], the presumption is that the legislature intends an act to be effective as an entirety—that is to say, the rule is against the mutilation of a statute; and if any provision be unconstitutional, the presumption is that the remaining provisions fall with it.").

probably would not have passed the price-fixing provisions of the code.[44]

Leave aside that these two paragraphs are internally inconsistent, first linking the operability of the remaining statutory scheme to legislative intent, as in *Warren*, but then reiterating the stand-alone functionality test from *Hill*. What matters most is the Court's vision of "imaginative reconstruction,"[45] which practically invited litigants to comb through a statute's legislative history looking for "considerations" that could rebut the plain language of a severability clause.

The Court's severability jurisprudence was now almost fully formed. As a general rule, statutes were presumed to be inseverable absent legislative intent to the contrary. But even when the legislature enacted an explicit severability clause, its unambiguous, intrinsic statement of legislative preference would give rise only to a presumption in favor of severability, rebuttable by evidence of contrary extrinsic legislative intent (derived from imaginative reconstruction based upon the statute's legislative history) or by showing that the remaining statutory scheme could not function. Summarizing the Court's jurisprudence in 1937, Professor Stern wryly observed: "Separability clauses are thus now significant only because of their absence. Like articles of clothing, if they are present little attention is paid to them, but if they are absent they may be missed."[46]

As suggested above, these decisions largely failed to grapple with any of the critical questions raised by their approaches to severability—so much so that one begins to wonder whether something was going on behind the scenes between 1914 (when the Court decided the *Ohio Tax Cases*) and 1936 (when it handed down *Carter Coal*). While it is possible only to speculate about any unacknowledged motivations that could have prompted the Court's marked reluctance to sever partially unconstitutional statutes—even when the legislature explicitly declared its intent—it bears mention that the Court's New Deal–era severability cases fit nicely into the "Lochnerism" that characterized its substantive jurisprudence during this period. As Professor Stern contemporaneously observed, many of the Court's leading severability opinions seemed to track the various Justices'

---

[44] *Id.* at 312–13 (citations omitted).

[45] I borrow this term from Judge Posner, though the idea belonged originally to Professor (and later Dean) Roscoe Pound. *Compare* Richard A. Posner, *Statutory Interpretation—in the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800, 817 (1983) ("[T]he task for the judge called upon to interpret a statute is best described as one of imaginative reconstruction. The judge should try to think his way as best he can into the minds of the enacting legislators and imagine how they would have wanted the statute applied to the case at bar.") (citation omitted), *with* Roscoe Pound, *Spurious Interpretation*, 7 COLUM. L. REV. 379, 381 (1907) ("[The Judge must endeavor to discover] what the law-maker meant by assuming his position, in the surroundings in which he acted, and endeavoring to gather from the mischiefs he had to meet and the remedy by which he sought to meet them, his intention with respect to the particular point in controversy.").

[46] Stern, *supra* note 13, at 122.

personal views of the underlying substance of the statutes at issue: When a Justice voted to uphold part of a statute, he also voted for severability; when he voted to strike part of a statute down, he voted against severance.[47]

In *Lemke v. Farmers Grain Co.*,[48] for instance, the Court's six-member majority (Taft, McKenna, Day, Van Devanter, Pitney, and McReynolds) invalidated certain provisions of a North Dakota grain regulation statute and declared those provisions inseverable from price-fixing regulations contained in another section of the same statute. The dissenting Justices (Holmes, Brandeis, and Clarke) argued that the challenged provisions were constitutional, and, in any event, severable from the statute's price-fixing provisions.[49] *Williams* showed a similar divide: Justice Sutherland delivered the opinion of the Court;[50] Holmes dissented without an opinion,[51] and Brandeis and Stone concurred in the result without joining the Court's opinion or issuing one of their own.[52] Likewise, in *Railroad Retirement Board v. Alton Railroad*,[53] decided the year before *Carter Coal*, the Court's five-Justice majority (Van Devanter, McReynolds, Sutherland, Butler, and Roberts) held the Act at issue unconstitutional and inseverable; the dissenters (Hughes, Brandeis, Stone, and Cardozo) would have held it constitutional and severable.[54] And in *Carter Coal*, the same five-Justice majority declared the Bituminous Coal Conservation Act of 1935 partially unconstitutional and inseverable while the same four dissenters again contended that the challenged provisions were constitutional—or at least severable from the remaining statutory scheme.[55] Finally, it is critical to note that the early New Deal Court's newly minted severability tests themselves pointed increasingly toward declaring statutes inseverable. *Hill* introduced a stand-alone functionality test for statutes containing a severability clause, giving litigants an opportunity to defeat completely only partially invalid statutes—even when the legislature had declared its preference for severance. *Williams* not only enunciated a general presumption of inseverability, but encouraged litigants to seek evidence of legislative intent contrary to a severability clause. And

---

[47] *Id.* at 113–14.
[48] 258 U.S. 50, 59–61 (1922).
[49] *Id.* at 64–65.
[50] Williams v. Standard Oil Co., 278 U.S. 235, 237 (1928).
[51] *Id.* at 245.
[52] *Id.*
[53] 295 U.S. 330, 360–63 (1935).
[54] *Id.* at 387–89, 391–92.
[55] *See* Carter v. Carter Coal Co., 298 U.S. 238, 277, 317, 324, 341 (1936). It bears noting here that the fractured lower-court opinions in the McCain-Feingold case uncomfortably replicate this pattern: Judge Kollar-Kotelly, who voted to uphold all but three provisions of BCRA, found the Act severable; Judge Leon, who voted to uphold some provisions and applications of the Act but also to strike down a roughly equivalent number, found the Act severable; and Judge Henderson, who had trouble finding even a single provision of BCRA she could support, declared emphatically that she would invalidate the statute in its entirety. *See supra* notes 4 and 10.

Carter Coal—the era's last great severability case—solidified rules that could allow a politically conservative Court to strike down in its entirety state and federal regulatory legislation only partially unconstitutional in substance. In the final analysis, the Court's pre-1937 severability jurisprudence seems to have either (at best) developed so clumsily that no conscious design can be attributed to the Court, or (at worst) resulted from deliberate manipulation by the Court's conservative core to enable it to strike down otherwise constitutional regulatory legislation that conflicted with its libertarian political preferences.

C. The Alaska Airlines *Formulation*

Despite the Supreme Court's machinations and its peculiar rules governing severability, things progressed rather smoothly after *Carter Coal.* Since 1936, the Court has not struck down in its entirety a statute only partially invalid and containing a severability clause,[56] and while it had some opportunities in the wake of *Carter Coal* to sort out its jurisprudence, severability questions seem to have mostly faded into the background—perhaps as a consequence of the Court's willingness to accept the constitutionality of most New Deal–style regulatory legislation.[57] Over the course of the next fifty years, the Court decided only two notable severability cases: *United States v. Jackson*,[58] in which it noted that "the ultimate determination of severability will rarely turn on the presence or absence of . . . a [severability] clause,"[59] and *Regan v. Time*,[60] where a four-member plurality invited the full Court to resurrect the general presumption of severability articulated earlier in its history:

> In exercising its power to review the constitutionality of a legislative Act, a federal court should act cautiously. A ruling of unconstitutionality frustrates the intent of the elected representatives of the people. Therefore, a court should refrain from in-

---

[56] Relying on legislative history, however, lower courts have done so. *See, e.g.*, W. States Med. Ctr. v. Shalala, 238 F.3d 1090, 1097 (9th Cir. 2001); United States v. Spokane Tribe of Indians, 139 F.3d 1297, 1299, 1302 (9th Cir. 1998); Jane L. v. Bangerter, 61 F.3d 1493, 1498 (10th Cir. 1995), *rev'd sub nom.* Leavitt v. Jane L., 518 U.S. 137 (1996); Spokane Arcades, Inc. v. Brockett, 631 F.2d 135, 139 (9th Cir. 1980), *aff'd mem.*, 454 U.S. 1022 (1981). The *Spokane Tribe* court did not totally invalidate the Indian Gaming Regulatory Act, but it did hold, against the statute's severability clause, that certain applications of the statute were invalid. *Spokane Tribe*, 139 F.3d at 1301. For a discussion of the issues raised by *Spokane Tribe*, see *infra* Part III.E.

[57] For an illuminating discussion of the Court's historic transformation from Lochnernism to non-interventionism, see generally Richard D. Friedman, *Switching Time and Other Thought Experiments: The Hughes Court and Constitutional Transformation*, 142 U. PA. L. REV. 1891 (1994).

[58] 390 U.S. 570 (1968).

[59] *Id.* at 586.

[60] 468 U.S. 641 (1984).

> validating more of the statute than is necessary. As this Court has observed, "whenever an act of Congress contains unobjectionable provisions separable from those found to be unconstitutional, it is the duty of this court to so declare, and to maintain the act in so far as it is valid." Thus, this Court has upheld the constitutionality of some provisions of a statute even though other provisions of the same statute were unconstitutional.[61]

Three years after *Regan*, the Court issued its leading contemporary opinion on severability. *Alaska Airlines v. Brock*[62] arose out of an airline industry challenge to the constitutionality of section 43 of the Airline Deregulation Act of 1978.[63] As enacted, section 43 contained both an "Employee Protection Program" ("EPP"), designed to shield long-term employees from the potentially adverse economic effects of deregulation, and a legislative veto provision that allowed either house of Congress to override Department of Labor regulations implementing the program.[64] Airline industry lawyers argued that the legislative veto provision was unconstitutional under *INS v. Chadha*[65] and that the entire section should be invalidated because the EPP and legislative veto provisions were inextricably intertwined.[66] The district court granted summary judgment to the airlines and struck down section 43 in its entirety.[67] Union members appealed the finding of nonseverability and prevailed before the United States Court of Appeals for the District of Columbia Circuit, which restored the EPP without the legislative veto provision.[68]

On further appeal, the Supreme Court seized the opportunity—created by doubt over whether a severability clause applied to the Act[69]—to rearticulate its severability standards both for statutes that contain a severability clause and for those that do not:

> [W]hen Congress has explicitly provided for severance by including a severability clause in the statute, . . . the inclusion of

---

[61] *Id.* at 652–53 (plurality opinion) (quoting El Paso & Northeastern Ry. Co. v. Gutierrez, 215 U.S. 87, 96 (1909)).

[62] 480 U.S. 678, 680 (1987).

[63] Airline Deregulation Act of 1978, Pub. L. No. 95-504, 92 Stat. 1705 (1978) (Section 43 codified at 49 U.S.C. § 1552).

[64] *Alaska Airlines*, 480 U.S. at 680–81.

[65] 462 U.S. 919 (1983).

[66] *Alaska Airlines*, 480 U.S. at 682–83.

[67] Alaska Airlines v. Donovan, 594 F. Supp. 92, 96 (D.D.C. 1984).

[68] Alaska Airlines v. Donovan, 766 F.2d 1550, 1565 (D.C. Cir. 1985).

[69] While the Airline Deregulation Act itself did not contain such a clause, it amended the Federal Aviation Act of 1958, Pub. L. No. 85-726, 72 Stat. 731 (1958), which did. *See Alaska Airlines*, 480 U.S. at 687 n.8. Rather than resolve the question of whether severability clauses attach to amending legislation, the Court simply enunciated the broad standards herein discussed and held the statute severable under a general presumption of severability.

> such a clause creates a presumption that Congress did not intend the validity of the statute in question to depend on the validity of the constitutionally offensive provision. In such a case, *unless there is strong evidence that Congress intended otherwise*, the objectionable provision can be excised from the remainder of the statute.[70]

When Congress has not included a severability clause, however, "Congress' silence is just that—silence—and does not raise a presumption against severability."[71] Instead, quoting *Champlin*, the Court held:

> Unless it is evident that the Legislature would not have enacted those provisions which are within its power, independently of that which is not, the invalid part may be dropped if what is left is fully operative as a law .... Congress could not have intended a constitutionally flawed provision to be severed from the remainder of the statute if the balance of the legislation is incapable of functioning independently. [When the unconstitutional provision is completely separate from the functional aspects of a statute, as with a legislative veto, t]he more relevant inquiry in evaluating severability is whether the statute will function in a *manner* consistent with the intent of Congress. The final test ... is the traditional one: the unconstitutional provision must be severed unless the statute created in its absence is legislation that Congress would not have enacted.[72]

After extensively reviewing the statute's legislative history, the Court concluded that Congress would have "enact[ed] the Airline Deregulation Act, including the EPP's first-hire program, [even] if the legislative veto had not been included,"[73] and therefore affirmed the Court of Appeals's judgment against the airlines.[74]

---

[70] *Alaska Airlines*, 480 U.S. at 686 (citing Champlin Refining Co. v. Corp. Comm'n, 286 U.S. 210, 235 (1932)) (emphasis added).

[71] *Id.* at 686 (citations omitted).

[72] *Id.* at 684–85 (quotations omitted). Although the *Alaska Airlines* regime appears to recognize that severing an unconstitutional provision from a statute may leave in force a statutory scheme that Congress would not have adopted, the Court's articulated test is terribly misleading. As the Court noted in footnote 7 of its opinion, the Court of Appeals's "statement that an invalid portion of a statute may be severed unless ... it is proved 'that Congress would have preferred no airline employee protection provision at all to the existing provision sans the veto provision' [is] completely consistent with the established severability standard." *Id.* at 685 n.7 (citations omitted). Thus, what appears to matter to the Court is whether Congress "would have enacted some [legislation on the subject at issue], rather than whether Congress would have enacted the same protections currently found in the Act [without the unconstitutional provision(s)]." *Id.*

[73] *Id.* at 697.

[74] *Id.*

As was typical of its prior jurisprudence in the area, the Court spent little time justifying the severability tests it enunciated. Failing even to mention *Carter Coal*, *Williams*, or its recent division in *Regan*, the Court conclusively rejected a general presumption of inseverability. After *Alaska Airlines*, statutes are presumed to be severable unless, following *Champlin*'s bifurcated *Warren* test, (1) Congress "intended otherwise"—that is, in the Court's words, Congress would have preferred no legislation at all to the enacted legislation without its unconstitutional provision(s)—or (2) the remaining statutory structure cannot function independent of its unconstitutional parts. As the Court had held in its various prior severability formulations, Congress's inclusion of a severability clause does not settle the issue. Instead, such a clause only preserves the general presumption of severability, which remains rebuttable by evidence of contrary legislative intent.

Neatly hidden in this articulation of the severability-clause test is the old *Hill* functionality requirement. Although the Court did not mention *Hill* in its discussion of severability clauses, it noted in the course of establishing its general presumption of severability that "Congress could not have intended a constitutionally flawed provision to be severed from the remainder of the statute if the balance of the legislation is incapable of functioning independently."[75] Taken as a whole, then, *Alaska Airlines* largely formalized what may well have been the peculiar results of the New Deal Court's struggle over the constitutionality of the growing regulatory state.

D. Recent Caselaw Addressing Inseverability Clauses

Interestingly enough, *in*severability clauses have not featured as prominently in the development of the current doctrine. This can probably be attributed to the related facts that Congress only occasionally enacts such clauses[76] and that the Supreme Court has never addressed a federal inseverability clause.[77] Lower federal courts, however, have some-

---

[75] *Id.* at 684.

[76] *See, e.g.*, Mobile Telecommunications Sourcing Act, Pub. L. No. 106-252, § 125, 114 Stat. 626, 632 (2000) (codified at 4 U.S.C.A. § 125) ("If a court of competent jurisdiction enters a final judgment on the merits that is based on Federal law[,] is no longer subject to appeal[,] and substantially limits or impairs the essential elements of sections 116 through 126 of this title, then sections 116 through 126 of this title are invalid and have no legal effect as of the date of entry of such judgment."); Catawba Indian Tribe of South Carolina Land Claims Settlement Act of 1993, Pub. L. No. 103-116, § 15, 107 Stat. 1118, 1136 (1993) (codified at 25 U.S.C. § 941(m)) ("If any provision of section 4(a), 5, or 6 of this Act is rendered invalid by the final action of a court, then all of this Act is invalid.").

[77] In *Zobel v. Williams*, 457 U.S. 55, 64–65 (1982), the Supreme Court did note the presence of an inseverability clause in a partially unconstitutional state statute. Though its treatment of the clause appears to have been more deferential than its traditional approach to severability clauses, the Court ultimately left the issue to the state courts to determine as

times considered inseverability clauses, and for the most part their approach mirrors *Alaska Airlines*.[78] Representative is *Biszko v. RIHT Financial Corp.*[79] At issue was a challenge to a Rhode Island statute restricting the ability of non–New England companies to acquire shares in Rhode Island financial institutions. Plaintiffs—stockholders of a Rhode Island bank holding company—alleged that the restrictions were unconstitutional and had artificially constrained the value of their stock by limiting the pool of potential purchasers.[80]

Dismissing the plaintiffs' complaint, the district court first held that plaintiffs had not suffered a legally cognizable injury-in-fact because Rhode Island had no constitutional obligation to allow interstate banking at all.[81] The state's decision to allow some interstate banking—even though it simultaneously placed restrictions on the activities of non-New England banks—had therefore raised the value of the plaintiffs' stock to a level above what it would have been had the state altogether refused to authorize interstate banking.[82] The court then reasoned that plaintiffs had also failed to prove that their asserted injuries were redressable.[83]

Anticipating this second objection, plaintiffs had argued that their injuries were remediable because the court could strike down the regional banking restrictions but leave the remainder of the statute in force—allowing institutions outside of New England to acquire stock in Rhode Island banks and increasing the value of plaintiffs' shares by expanding the pool of potential purchasers.[84] This, the court responded, it simply could not do: The Rhode Island statute contained an inseverability clause declaring that if any parts of the law were declared unconstitutional, the whole law (except a single provision applicable only to credit unions) should be struck down.[85] Because valid federal law requires states to

---

a matter of state law. *Id.* at 65 ("Here, we need not speculate as to the intent of the Alaska Legislature; the legislation expressly provides that invalidation of any portion of the statute renders the whole invalid . . . . However, it is of course for the Alaska courts to pass on the severability clause of the statute."). To the extent that the Court indicated that inseverability clauses merit greater deference than severability clauses, it never explained why; and, as we shall see, lower courts have not followed through on *Zobel*'s suggestion.

[78] *See, e.g.*, Biszko v. RIHT Fin. Corp., 758 F.2d 769, 773 (1st Cir. 1985) (holding that "a non-severability clause cannot ultimately bind a court"). Not withstanding this line of cases, at least one state supreme court has invalidated an entire statute notwithstanding its inseverability clause. *See* Stiens v. Fire & Police Pension Ass'n, 684 P.2d 180, 185–86 (Colo. 1984). For a brief synopsis of *Stiens*, see Israel E. Friedman, Comment, *Inseverability Clauses in Statutes*, 64 U. CHI. L. REV. 903, 907–08 (1997).

[79] 102 F.R.D. 538 (D.R.I. 1984), *aff'd* 758 F.2d 769 (1st Cir. 1985). The discussion here focuses on the district court opinion because it more fully analyzes the validity of the statute's inseverability clause and because the First Circuit's analysis of the issue essentially adopts the lower court's reasoning and opinion. *See Biszko*, 758 F.2d at 774.

[80] *Biszko*, 102 F.R.D. at 541–42.

[81] *Id.* at 542–43.

[82] *Id.* at 543.

[83] *Id.* at 543–44.

[84] *Id.* at 543.

[85] *Id.*

authorize interstate bank acquisition before any out-of-state institution can acquire ownership of in-state banks, the value of the plaintiffs' stock would plummet if the court obeyed the inseverability clause and struck down the entire statute: No out-of-state institutions—not even those in New England—could purchase the plaintiffs' shares.[86]

Yet in suggesting that it would enforce the state legislature's inseverability clause, the court did not indicate that it was doing so because it was obliged to obey the legislature's unambiguous command. Instead, it repeated the *Dorchy* maxim that "[a] severability or, in this case non-severability, clause is a guideline for statutory interpretation but not a mandate to the court" and concluded that a partially unconstitutional statute "can be extended only if the portion remaining would be operative law and follow legislative intent."[87] Looking beyond the inseverability clause to the history and structure of the Rhode Island banking bill, the court then determined that the state legislature would not have adopted an interstate banking statute without regional restrictions.[88]

Here, at least in preliminary form, was the *Alaska Airlines* approach to severability clauses applied to an inseverability clause. Like the Supreme Court in its severability cases, neither the district court nor the First Circuit fully explained why it felt compelled to look beyond an explicit legislative command for extrinsic evidence of legislative intent, beyond arguably applicable precedent borrowed from the Court's severability clause jurisprudence.

III. *ALASKA AIRLINES*' UNJUSTIFIED APPROACH TO SEVERABILITY CLAUSES

This Part begins to make the case for a textualist approach to severability and inseverability clauses. It starts with the notion that severability clauses are binding legislation due judicial recognition under Article I, Section 7 of the Constitution. From there, the Article addresses a number of concerns: that severability clauses usurp the Article III judicial power, that they violate principles of non-delegation, that they may leave in place a dysfunctional statutory scheme if enforced, and that they ought to be interpreted in light of changed circumstances. None of these objections justifies disregarding the clear command of severability clauses.

A. Article I, Section 7 as a Binding Rule of Recognition

The Court's severability jurisprudence—and particularly its notion that a severability clause creates only a presumption of severability re-

---

[86] *Id.* at 543–44.
[87] *Id.* at 543.
[88] *Id.* at 544.

buttable by extrinsic evidence of legislative intent—raises troubling separation of powers problems. After all, severability clauses have been through the Article I, Section 7 process[89] and therefore are valid legislation that binds judicial decisionmaking. It is difficult to evaluate the Court's treatment of this concern, however, as the Justices have never fully justified their willingness to overlook severability clauses.

Over the course of the past century, scholars and other commentators have attempted to fill in the gaps created by the Court's cavalier treatment of severability directives. Most often, they justify judicial disregard for severability clauses on grounds that such clauses are merely "boilerplate" provisions inserted into statutes without substantive deliberation by the legislature.[90] In many cases, this is true: Even legislators have occasionally acknowledged that little debate is necessary before adding a severability clause to pending legislation.[91] But as the debate over McCain-Feingold's severability clause suggests, this is not always the case.[92] The fact that Congress sometimes chooses *not* to enact a severability clause[93] (or to include an *inseverability* clause in legislation likely to face a constitutional challenge)[94] suggests that, despite the frequency with which it opts for severability, Congress's decisions to do so are purposeful.[95] It is well-settled that "[a]bsent clear congressional intent to the contrary, [courts should] assume the legislature did not intend to pass

---

[89] *See* U.S. CONST. art. I, § 7, cl. 2 (bicameralism and presentment).

[90] 2 SUTHERLAND STATUTES AND STATUTORY CONSTRUCTION § 44.08 (5th ed. 1992) [hereinafter SUTHERLAND] ("[T]o say that a saving clause is 'indisputable evidence' of legislative intent to pass part of a statute irrespective of void provisions is to put too great an emphasis on the mechanical inclusion of such provisions."); Max Radin, *A Short Way With Statutes*, 56 HARV. L. REV. 388, 419 (1942) ("Are we really to imagine that the legislature had, as it says it had, weighed each paragraph literally and come to the conclusion that it would have enacted that paragraph if all the rest of the statute were invalid?"); Stern, *supra* note 13, at 122 ("When legislatures declared that '[t]he invalidity of any part of this statute shall not affect the remainder,' they did not mean it.").

[91] *See, e.g.*, 140 CONG. REC. 53 (1994) (statement of Rep. Slaughter) ("This is a standard 'boilerplate' severability clause; similar language has been included in a wide variety of laws including the Emergency Unemployment Compensation Amendments of 1993, the Americans with Disabilities Act, the Civil Rights Restoration Act, the Fair Labor Standards Act, the Education for Economic Security Act, and the Comprehensive Drug Abuse Prevention and Control Act."); 134 CONG. REC. 75 (1988) (statement of Rep. Frank) ("This is just boilerplate severability.").

[92] *See supra* notes 1–8 and accompanying text.

[93] *See, e.g.*, Ethics Reform Act of 1989, Pub. L. No. 101-194, 103 Stat. 1716 (codified in scattered sections of 5 U.S.C.) (partially invalidated in *United States v. National Treasury Employees Union*, 513 U.S. 454 (1995)); Low-Level Radioactive Waste Policy Amendments Act of 1985, Pub. L. 99-240, 99 Stat. 1842 (codified at 42 U.S.C. §§ 2021–2022) (partially invalidated in New York v. United States, 505 U.S. 144 (1992)); Sentencing Reform Act of 1984, Pub. L. No. 98-473, 98 Stat. 1987 (codified as amended at 18 U.S.C. §§ 3551–3586 and 28 U.S.C. §§ 991–998) (challenged repeatedly until upheld in Mistretta v. United States, 488 U.S. 361 (1989)).

[94] *See supra* note 76.

[95] *Cf.* Nagle, *supra* note 14, at 243–45 (arguing that the diversity of severability clauses undermines the boilerplate justification for disregarding them).

vain or meaningless legislation,"[96] and the fact that Congress almost always prefers statutes to be severable—and usually chooses to reveal that preference explicitly to the courts—does nothing to show that Congress does not really mean what it says.

But even if Congress devoted little thought to including a severability provision in legislation, and it did so all the time, such clauses would still be valid legislation requiring judicial enforcement. Article I, Section 7, clause 2 of the Constitution sets forth the exclusive *procedural* requirements for legislation to bind.[97] Under this framework, a bill—and each of its subsidiary provisions—becomes law once it is passed by majority votes in both the House and Senate and is signed by the President (or, if the President refuses to sign it, after a two-thirds majority of each chamber votes to pass the bill over the President's veto).[98] Once a bill goes through this arduous legislative process, it becomes a legally binding statutory enactment subject only to *substantive* constitutional constraints; "[j]udicial nullification of statutes, admittedly valid and applicable, has, happily, no place in our system."[99]

Bicameralism and presentment thus do more than undercut the usual justification for disregarding severability clauses: They condemn the Court's practice of turning to legislative history to determine whether "Congress really intended" unconstitutional statutory provisions to be severable when the bill at issue contains an unambiguous severability clause.[100] In

---

[96] Coyne & Delany Co. v. Blue Cross & Blue Shield of Va., 102 F.3d 712, 715 (4th Cir. 1996).

[97] Although the discussion here focuses on the U.S. Constitution's structure for lawmaking, every state constitution except Nebraska's contains bicameralism and presentment requirements. *See* John Dinan, *Framing a "People's Government": State Constitution-Making in the Progressive Era*, 30 RUTGERS L.J. 933, 963–64 (1999). As a result, the concerns articulated here apply with equal force when state courts are called upon to address state law severability questions. Federal courts are, of course, bound by the constraints of federalism to defer to state law severability doctrine—even if that doctrine may appear to violate a state's own system of separated powers.

[98] U.S. CONST. art. I, § 7, cl. 2.

[99] Sorrells v. United States, 287 U.S. 435, 450 (1932); *see also* Laker Airways v. Sabena, 731 F.2d 909, 949 (D.C. Cir. 1984). In addition, the judicial creation of standards for legislative decisionmaking—no matter how desirable that may seem—appears to violate Article I, Section 5, which provides that "[e]ach House may determine the rules of its proceedings . . . ." U.S. CONST. art. I, § 5, cl.2. It also seems to put the courts squarely in the middle of purely "political questions" that they are compelled to avoid. *See, e.g.*, Baker v. Carr, 369 U.S. 186, 217 (1962) ("Prominent on the surface of any case held to involve a political question is found a textually demonstrable constitutional commitment of the issue to a coordinate political department; or a lack of judicially discoverable and manageable standards for resolving it . . . .").

[100] The search for "legislative intent" has long been criticized as futile. More than seventy years ago, Max Radin put it this way: "The chances that of several hundred men each will have exactly the same determinate situations in mind as possible reductions of a given [statutory issue] are infinitesimally small." Max Radin, *Statutory Interpretation*, 43 HARV. L. REV. 863, 870 (1930). The chances of this group having a single, shared mindset are especially slim because legislators' individual (and collective) policymaking decisions depend so strongly on non-policy factors, such as the order in which they consider various alternatives and the logrolling opportunities available to each of them in a given statutory

Judge Easterbrook's words, it "demean[s] the constitutionally prescribed method of legislating to suppose that its elaborate apparatus for deliberation on, amending, and approving a text is just a way to create some evidence about the law, while the real source of legal rules is in the mental processes of legislators."[101] Only the severability clause itself is law under our Constitution's framework for lawmaking, and the judiciary is constitutionally bound to honor it without regard for what the surrounding statute's legislative history might be read to suggest.[102]

---

debate at a particular moment in time. DANIEL A. FARBER & PHILIP P. FRICKEY, LAW AND PUBLIC CHOICE 38–39 (1991). *See also generally* KENNETH ARROW, SOCIAL CHOICE AND INDIVIDUAL VALUES (2d ed. 1963).

Legislative history is, in any event, often a bad place to look for legislative intent. Professors Frickey and Eskridge—nonetheless staunch defenders of using legislative history—observe:

> Lobbyists and lawyers maneuver endlessly to persuade staff members (who write committee reports) or their legislative bosses to throw in helpful language in the reports when insertion of similar language would be inappropriate or infeasible for the statute itself. "Smuggling in" helpful language through legislative history is a time-tested practice.

ESKRIDGE & FRICKEY, *supra* note 15, at 744. The fact that, among other things, these reports are: rarely read by committee members; never voted upon by the committees that issue them (much less by the full House or Senate); and not amendable aides the "smuggling" process. *See* Hirschey v. FERC, 777 F.2d 1, 7 n.1 (D.C. Cir. 1985) (Scalia, J., concurring) (quoting an exchange on the Senate floor between Senators Armstrong and Dole). Consequently,

> legislative history can be cited to support almost any proposition, and frequently is. The propensity of judges to look past the statutory language is well known to legislators. It creates strong incentives for manipulating legislative history to achieve through the courts results not achievable during the enactment process.

Wallace v. Christensen, 802 F.2d 1539, 1559 (9th Cir. 1986) (Kozinski, J., concurring). This critique is doubly damning when applied to hearings and floor debates, as opposed to committee reports. *See* Reed Dickerson, *Statutory Interpretation: Dipping into Legislative History*, 11 HOFSTRA L. REV. 1125, 1131–32 (1983). For contrasting views on the use of legislative history, see Stephen Breyer, *On the Uses of Legislative History in Interpreting Statutes*, 65 S. CAL. L. REV. 845 (1992) (arguing that the case for abandoning the use of legislative history has not yet been made); William N. Eskridge, Jr., *Legislative History Values*, 66 CHI.-KENT L. REV. 365 (1990) (identifying three values that legislative history might serve: "authority value," "purpose value," and "truth value" and concluding that there are both significant problems and substantial benefits to using legislative history); Patricia M. Wald, *The Sizzling Sleeper: The Use of Legislative History in Construing Statutes in the 1988–89 Term of the United States Supreme Court*, 39 AM. U. L. REV. 277 (1990) (arguing that courts generally have used legislative history in a sensible fashion).

[101] *In re* Sinclair, 870 F.2d 1340, 1344 (7th Cir. 1989).

[102] *See* R.R. Comm'n of Wis. v. Chi., Burlington, and Quincy R.R. Co., 257 U.S. 563, 589 (1922) ("[Legislative history is] only admissible to solve doubt and not to create it."); *see also* HUD v. Rucker, 535 U.S. 125, 132 (2002) ("[R]eference to legislative history is inappropriate when the text of the statute is unambiguous."); W. Va. Univ. Hosps., Inc. v. Casey, 499 U.S. 83, 98–99 (1991) ("The best evidence of [legislative] purpose is the statutory text adopted by both Houses of Congress and submitted to the President. Where that contains a phrase that is unambiguous—that has a clearly accepted meaning in both legislative and judicial practice—we do not permit it to be expanded or contracted by the

B. Severability Clauses and the Judicial Power

The preceding discussion presumes, of course, that Congress has the constitutional authority to direct courts to sever unconstitutional statutory provisions. Although no court has held that severability clauses are unconstitutional *per se*,[103] some commentators have argued that severability clauses unconstitutionally usurp the Article III "judicial Power"[104] by directing courts to exercise judicial review under legislatively mandated standards.[105] While the argument is tempting—after all, judicial review *is* a quintessentially judicial power—it is ultimately unconvincing.

To begin with, the Supreme Court has recognized that Congress enjoys plenary control over the creation of statutory rights, and that in exercising that control, Congress has broad authority to condition the exercise of judicial power. In *Northern Pipeline Constr. Co. v. Marathon Pipe Line Co.*, where the Court addressed the constitutionality of Congress's decision to vest Article III judicial power in "adjunct" bankruptcy courts staffed by judges lacking Article III protections, a four-Justice plurality explained:

---

statements of individual legislators or committees during the course of the enactment process.").

[103] This may merely reflect the fact that courts have never really considered severability clauses to be absolutely binding.

[104] U.S. CONST. art. III, § 1.

[105] *See, e.g.*, SUTHERLAND, *supra* note 90, § 44.08 ("[I]t should be kept in mind that the authority of a court to eliminate invalid elements of an act and yet sustain the valid elements is not derived from the legislature, but rather flows from powers inherent in the judiciary."); Note, *Constitutional Law—Partial Unconstitutionality of Statutes—Effect of Saving Clause on General Rules of Construction*, 25 MICH. L. REV. 523, 527 (1927) (concluding that courts will not allow the presence of a severability clause to take away their power to independently determine questions of severability). One federal court has found a severability clause unconstitutional in its particular circumstances on similar grounds. In *Mathews v. Schweiker*, No. 79-G-5251-NE, 1982 U.S. Dist. LEXIS 1812 (N.D. Ala. Aug. 24, 1982), plaintiffs challenged certain provisions of the Social Security Act requiring that they receive more than half of their financial support from their wives to be eligible for certain benefits. After concluding that the statute violated plaintiffs' equal protection rights by senselessly denying them benefits they would have been eligible to receive if they were women, the district court addressed the constitutionality of the statute's peculiar severability clause. *Id.* at *10. Because its effect would have been to deny everyone spousal benefits, the court held that it represented

> an unconstitutional usurpation of judicial power by the legislative branch of the government [insofar as it] attempted to mandate the outcome of any challenge to the validity of the [law] by making such a challenge fruitless. Even if a plaintiff achieved success in having the gender-based classification stricken, he would derive no personal benefit from the decision . . . .

Id. at *11–*12. On appeal directly from the district court, the Supreme Court later reversed the decision, but found it unnecessary to address the constitutionality of the statute's severability clause. *See* Heckler v. Mathews, 465 U.S. 728, 737–40 (1984).

> The constitutional system of checks and balances is designed to guard against encroachment or aggrandizement by Congress at the expense of the other branches of government. *But when Congress creates a statutory right, it clearly has the discretion, in defining that right, to create presumptions, or assign burdens of proof, or prescribe remedies*; it may also provide that persons seeking to vindicate that right must do so before particularized tribunals created to perform the specialized adjudicative tasks related to that right. *Such provisions do, in a sense, affect the exercise of judicial power, but they are also incidental to Congress' power to define the right that it has created.*[106]

Although severability clauses do not necessarily arise in the context of congressional "rights creation," the principle enunciated by the *Northern Pipeline* plurality (and its implications for severability) holds true across statutory types. The fact that Congress passes a statute that "*affect[s]* the exercise of judicial power"[107] does little to render such legislation unconstitutional on grounds that it *usurps* the judicial power. Two examples help illustrate this conclusion. First, the Supreme Court has held that legislative "fallback" provisions, largely analogous to severability clauses, bind courts. In *Bowsher v. Synar*,[108] where the Court was forced to determine how to remedy constitutional defects in the Balanced Budget and Emergency Deficit Control Act of 1985,[109] the plurality explained:

> The language of the [Act] itself settles the issue. In § 274(f), Congress has explicitly provided "fallback" provisions in the Act that take effect "[i]n the event . . . *any* of the reporting procedures described in section 251 are invalidated." The[se] fallback provisions are fully operative as a law [and were intended] to be given effect in this situation.[110]

The fallback provisions at issue in *Bowsher* closely resemble severability clauses. When Congress includes a severability clause, the en-

---

[106] 458 U.S. 50, 83 (1982) (plurality opinion) (emphasis added) (citation omitted) (footnote omitted). Of special note, the dissenting opinion in *Northern Pipeline* favorably cites an analogous claim. *See id.* at 104 ("In fact, the plurality [acknowledges that the exceptions to Article III's reach are broad in] announc[ing] that 'when Congress creates a substantive federal right, it possesses substantial discretion to prescribe the manner in which that right may be adjudicated.'") (White, J., dissenting) (quoting plurality opinion at 80).

[107] *Id.* at 83 (emphasis added).

[108] Bowsher v. Synar, 478 U.S. 714, 717–18 (1986).

[109] Balanced Budget and Emergency Deficit Control Act of 1985, Pub. L. No. 99-177, 99 Stat. 1037 (codified as amended at 2 U.S.C. § 901 (2000)).

[110] Bowsher, 478 U.S. at 735 (citations omitted).

acted statutory scheme operates until the courts strike down any of its provisions, at which point the clause acts (as a kind of statutory shorthand) to repeal the partially invalid scheme and replace it with one that contains all of the original scheme's provisions except for those declared unconstitutional.[111] No aspect of the *judicial power* precludes Congress from providing this kind of statutory remedy for unconstitutionality—at least not according to the Court.[112]

Second, courts have always (at least implicitly) recognized Congress's authority to bind judicial interpretation by legislating definitions for the terms it uses in the statutory schemes it creates.[113] As Nick Rosenkranz observes:

> [W]hen Congress inserts a definitional section, courts resort not to their usual grabbags of interpretive tools, but to the statutory definition alone. Congress in effect replaces a complicated and fuzzy algorithm with a simple cut-and-paste function. "Where one sees *X*, one shall read *Y*." No guesswork is necessary (and no litigation likely) to determine whether [*X* is *Y*]. It is; Congress said so. Cut and paste.[114]

Interpreting statutory terms is arguably just as judicial a power as judicial review. Of course, it is possible that courts could interpret a statutorily defined term within a much broader range if Congress had chosen not to define it. But the fact that "courts would have an additional power absent a particular statutory provision hardly suffices to show that the provision violates Article III."[115] Likewise, the fact that courts might otherwise be able to strike down in its entirety a statute that is only unconstitutional in part, and that does not contain a severability clause, fails to show how the enactment of a binding severability clause impinges upon the federal courts' inherent and exclusive Article III powers.[116]

---

[111] *See* Evan H. Caminker, Note, *A Norm-Based Remedial Model for Underinclusive Statutes*, 95 YALE L.J. 1185, 1208 n.87 (1984).

[112] This conceptualization, however, helps to highlight possible non-delegation problems inherent in severability clauses. This potential pitfall is addressed *infra* at Part III.C.

[113] *See, e.g.*, United States v. Hodge, 321 F.3d 429, 436 (3d Cir. 2003) ("Absent an absurd departure from conventional English, Congress of course is free to define terms in statutes differently than any particular dictionary does."); United States v. Harkey, 709 F. Supp. 977, 984 (E.D. Wash. 1989) ("Clearly, if Congress desires to include burglaries with elements other than those of common law burglary, it is free to so legislate.").

[114] Nicholas Quinn Rosenkranz, *Federal Rules of Statutory Interpretation*, 115 HARV. L. REV. 2085, 2104 (2002) (footnotes omitted). Of course, as Rosenkranz notes, "there is no guarantee that definition *Y* will be less ambiguous than defined term *X*, and a court may require all its interpretive ingenuity to give content to *Y*. The point, though, is that no interpretive work will be necessary *as to term X* except a simple cut and paste." *Id.* at 2104 n.71 (emphasis added).

[115] *Id.* at 2105.

[116] Rosenkranz goes on to note that definitional enactments need not be statute-specific to obtain constitutional validity. *See id.* at 2109–26. Consider, as he does, the Defense of

More broadly understood, this conclusion flows from the fundamental structure of American law. As Professors Eskridge and Ferejohn explain:

> [The] Constitution committed the national government to lawmaking by elected representatives deliberating for the public good. Article I's vesting legislative authority in Congress and Article III's vesting the Supreme Court and inferior federal courts with jurisdiction to interpret federal statutes (and only implicit jurisdiction to hear federal common law claims) suggest the principle that the primary source of law at the federal level would be statutes—a striking contrast to England and the states, where Blackstonian common law precedents remained the main source of law.[117]

Given these background norms, it is beyond question that constitutional requirements trump any contrary statutory enactments and that valid statutory enactments, in turn, trump judicially crafted common law.[118] To the extent that the Court's current tests for severability are not constitutional rules—and there has never been any suggestion that they are anything but prudential—Congress may effectively reverse them by statute without impinging upon the federal judiciary's inherent Article III powers.

C. Severability Clauses and Non-Delegation

A second constitutional critique of severability clauses takes the opposite tack, suggesting that these clauses grant too *much* power to courts. Suggested by Max Radin as early as 1942,[119] this critique argues that severability clauses represent an unconstitutional delegation of lawmaking

---

Marriage Act: "In determining the meaning of any Act of Congress, . . . the word 'marriage' means only a legal union between one man and one woman as husband and wife, and the word 'spouse' refers only to a person of the opposite sex who is a husband or a wife." 1 U.S.C. § 7 (2000). Whatever its arguable constitutional defects, the fact that its definition of marriage binds judicial interpretation of all federal statutes using the term "marriage" is not one of them. *See, e.g.*, Andrew Koppelman, *Dumb and DOMA: Why the Defense of Marriage Act Is Unconstitutional*, 83 IOWA L. REV. 1, 14, 24 (1997) (arguing that DOMA violates the Full Faith and Credit Clause and Equal Protection principles); Kristian D. Whitten, *Section Three of the Defense of Marriage Act: Is Marriage Reserved to the States?*, 26 HASTINGS CONST. L.Q. 419, 421 (1999) (arguing that DOMA infringes upon the Tenth Amendment). Similarly, a general severability clause passed by Congress—stating, for example, that "[u]nless otherwise indicated by a clear statutory statement, all statutes passed after the effective date of this Act shall be severable"—violates Article III no more than the statute-specific severability clauses contained in most contemporary federal legislation (which is, of course, to say not at all).

[117] William N. Eskridge, Jr. & John Ferejohn, *Super-Statutes*, 50 DUKE L.J. 1215, 1221 (2001).

[118] *See also* Rosenkranz, *supra* note 114, at 2107.

[119] Radin, *supra* note 90, at 419 ("Are we really to imagine that the legislature had, as it says it had, weighed each paragraph literally and come to the conclusion that it would have enacted that paragraph if all the rest of the statute were invalid?").

authority to the courts because Congress could not have considered every possible statutory permutation that might result from a court's post-enactment decision to sever a subset of unconstitutional statutory language.[120]

The objection proves too much. Congress often leaves critical statutory issues unresolved,[121] thereby inviting the judiciary to develop federal common law to supplement the statutory scheme and effectively delegating significant policymaking authority to the courts. This practice is well-accepted. Consider the Sherman Act,[122] which represents a classically broad grant of policymaking authority to the judiciary.[123] Although it outlaws "[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations,"[124] the Sherman Act fails to define even one specific practice that constitutes a "conspiracy . . . in restraint of trade."[125] Its legislative history (assuming that it is an appropriate source of interpretive advice in the face of such genuine and deliberate statutory ambiguity) offers virtually no guidance.[126] As a result, the Court has declined to adopt a static approach to interpreting the Act,[127] and it long ago rejected non-delegation challenges.[128] Other landmark statutes contain similarly sweeping delegations of policymaking authority to the courts without raising constitutional doubts.[129]

---

[120] *Cf.* Ohio Oil Co. v. Wright, 53 N.E.2d 966, 974 (Ohio 1944) ("Literally interpreted [the severability clause] would result in a new statute going into effect as a result of the judgment of this court deciding that either some classes were improperly included, or other classes improperly excluded. The new law would be created by this court and not by the General Assembly, because it enacted a different one. This would amount to a delegation of legislative powers to the courts, which is contrary to article III of the constitution, as well as numerous decisions of this court.").

[121] *See, e.g.*, Frank H. Easterbrook, *Statutes' Domains*, 50 U. CHI. L. REV. 533, 540 (1983) ("Almost all statutes are compromises, and the cornerstone of many a compromise is the decision . . . to leave certain issues unresolved.").

[122] Act of July 2, 1890, ch. 647, 26 Stat. 209 (codified at 15 U.S.C. §§ 1–2 (2000)).

[123] Cass R. Sunstein, *Interpreting Statutes in the Regulatory State*, 103 HARV. L. REV. 405, 421 (1989).

[124] 15 U.S.C. § 1 (2000).

[125] *See id.* §§ 1–2 (2000).

[126] *See generally* 1 THE LEGISLATIVE HISTORY OF THE FEDERAL ANTITRUST LAWS AND RELATED STATUTES (Earl W. Kitner ed., 1978).

[127] *See, e.g.*, Bus. Elecs. Corp. v. Sharp Elecs. Corp., 485 U.S. 717, 732 (1988) ("The Sherman Act adopted the term 'restraint of trade' along with its dynamic potential. It invokes the common law itself, and not merely the static content that the common law had assigned to the term in 1890.").

[128] *See* Standard Oil Co. v. United States, 221 U.S. 1, 59–64 (1911).

[129] In enacting the Employee Retirement Income Security Act of 1974 ("ERISA"), for instance, Congress delegated policymaking authority to the judiciary "to develop a 'federal common law of rights and obligations under ERISA-regulated plans.'" Firestone Tire & Rubber Co. v. Bruch, 489 U.S. 101, 110 (1989) (citations omitted). Similarly, the Court has interpreted section 301(a) of the Labor Management Relations Act of 1947, 29 U.S.C.S. § 185, as an authorization for the courts to develop a federal common law of labor-management relations (subject to jurisdictional limits). *See* Textile Workers v. Lincoln Mills of Ala., 353 U.S. 448 (1957).

In contrast to these broad delegations, severability clauses are restrained. When a court enforces a severability clause, it may not add provisions or limitations to legislation but can only enjoin enforcement of constitutionally offensive provisions—as directed by the statute itself.[130] If this limited grant of authority offends non-delegation principles, then the practice of crafting federal common law—and any statutes that confer such broad authority on the courts—are likewise unconstitutional. Even the most serious "new textualists"[131] would likely reject this result.[132]

In any event, the objection is exaggerated. Although courts could conceivably declare any provision of a statute unconstitutional, there are, in fact, likely to be only a few provisions of any single statute that raise constitutional questions, and even fewer that are likely to be declared unconstitutional.[133] The possible number of post-severance statutory permutations accordingly is limited. If three statutory provisions (*A*, *B*, and *C*) raise serious constitutional problems, there are seven possible post-severance permutations: all provisions except *A*; all provisions except *B*; all provisions except *C*; all provisions except *A* and *B*; all provisions except *B* and *C*; all provisions except *A* and *C*; and all provisions except *A*, *B*, and *C*. Given that Congress evaluates hundreds of statutory permutations during the lawmaking process, considering seven, twenty, or even fifty possible post-severance permutations is certainly within its institutional capacity.[134]

---

[130] *See* Hill v. Wallace, 259 U.S. 44, 70 (1922) ("[The severability clause] did not intend the court to dissect an unconstitutional measure and reframe a valid one out of it by inserting limitations it does not contain. This is legislative work beyond the power and function of the court.").

[131] For an introduction to, and critique of, the "new textualist" philosophy, see Eskridge, *supra* note 13, at 41–47, 230–34, 272–73. For a new textualist's view on statutory interpretation, see, for example, Easterbrook, *supra* note 122.

[132] As John Manning has noted:

> [T]extualist judges allow that statutory indeterminacy, and the resulting need for norm-specification, may at times involve judges in the exercise of substantial policymaking discretion. For example, Justice Scalia acknowledges that "policy evaluation is . . . part of the traditional judicial tool-kit." Although he often finds that a statute's meaning "is apparent from its text and from its relationship with other laws," he recognizes that judicial, like agency, elaboration of an ambiguous statutory text may entail some exercise of delegated policymaking authority.

John F. Manning, *Textualism as a Nondelegation Doctrine*, 97 COLUM. L. REV. 673, 701 (1997) (quoting Antonin Scalia, *Judicial Deference to Administrative Interpretations of Law*, 1989 DUKE L.J. 511, 515).

[133] Between 1995 and 2001, the Supreme Court partially invalidated twenty-seven federal laws. Neal Devins, *Congress as Culprit: How Lawmakers Spurred on the Court's Anti-Congress Crusade*, 51 DUKE L.J. 435, 464 n.1 (2001). Although this is a large number of invalidations when gauged against historical standards, the total nonetheless represents only a miniscule fraction of the total number of laws Congress enacted during this period.

[134] *See also* Stern, *supra* note 13, at 126, 128 ("Although all problems of separability cannot be recognized in advance, the most important of them can be. Particularly in the case of new statutes regulating new subjects can a preliminary examination by a person familiar with constitutional authorities reveal the principle points of danger . . . . [T]he

D. Severability and Statutory Functionality

The oldest rationale for disregarding severability clauses was first articulated in *Warren*, reiterated by the Supreme Court in *Carter Coal*, and is implicit in the *Alaska Airlines* test: a statutory scheme that cannot function independently from its unconstitutional subsections should be struck down in its entirety, notwithstanding Congress's inclusion of a severability clause. As explained earlier, this "functionality standard" seems out of place in the severability clause context because it was designed as an extrinsic mechanism for gauging legislative intent in an era before legislatures developed an intrinsic statutory mechanism to express their preference for severability explicitly (a development, in fact, that was apparently spurred by the courts' very use of the standard). Once legislatures began including severability clauses in constitutionally questionable legislation, however, their intent with regard to severability was clear—even if the post-severance statutory scheme envisioned by a reviewing court appeared odd or incapable of precisely serving its intended purpose. What justification for the functionality standard remains?

Before going further, it is important to distinguish between two different contexts in which courts have purported to apply the standard. The first is the scenario at issue in *Hill v. Wallace*.[135] Recall that there, the Supreme Court invalidated regulations contained in the 1921 Future Trading Act[136] and then limited the scope of the statute's broad severability clause by explaining that "Section 4 with its penalty to secure compliance with the regulations of Boards of Trade is so interwoven with those [unconstitutional] regulations that they can not be separated. None of them can stand."[137]

The *Hill* Court's decision to invalidate the statutory penalties was right—but not for any reason related to severability. The real problem with enforcing penalties designed to secure compliance with unconstitutional regulations is that doing so is unconstitutional, not that the underlying regulations and their enforcement mechanisms are too "interwoven" to be separated without leaving behind a dysfunctional statute. Once it is determined that Congress is constitutionally proscribed from *prohibiting* or *requiring* certain conduct, it follows that Congress is constitutionally precluded from *punishing* someone for engaging in (or failing to engage in) that conduct. In cases like *Hill*, the issue of whether to invalidate penalties has nothing to do with severability, and everything to do with "substantive" constitutional law.

---

basic problems of separability raised by most important statutes can be discovered and cared for when the statute is being drafted.").

[135] 259 U.S. 44 (1922).

[136] Pub. L. No. 67-66, 42 Stat. 187 (1921).

[137] *Hill*, 259 U.S. at 70.

The second context in which courts have applied the functionality standard to limit severability clauses is arguably a more appropriate one, and is illustrated by *Warren*. There, in wholly invalidating a state statute annexing Charlestown into Boston, the Supreme Judicial Court of Massachusetts explained that the "various provisions of the act, therefore, all providing for the consequences of [the impermissible] annexation, . . . are connected and dependent[,] . . . look to one object and its incidents, and are so connected with each other" that the dysfunctional but constitutionally valid statutory remnants could not be permitted to remain in force.[138]

Despite the fact that the *Warren* court's stated rationale for the standard—that a dysfunctional post-severance scheme cannot comport with legislative intent—is misplaced in the severability clause context, the standard itself continues to hold purchase in the modern era. I suspect that this is largely due to the fact that the standard seems to recognize something *constitutionally* troubling about a residual statutory scheme that cannot function. As Professor Mashaw has observed, "some form of reasonableness check on legislative activity is embedded in our general commitment to the protection of individual liberty and the idea of a government of limited powers."[139] And although the Due Process Clauses[140] speak in broad procedural terms, they seem to embody the quasi-substantive principle that both federal and state legislation must at least employ means reasonably calculated to achieve a legitimate legislative purpose.[141] When a post-severance statutory scheme truly "cannot function," it cannot serve any purpose. Indeed, there is a risk that its inability to function will lead to arbitrary state action.[142]

---

[138] *Warren*, 68 Mass. (2 Gray) at 100.

[139] JERRY L. MASHAW, GREED, CHAOS, AND GOVERNANCE: USING PUBLIC CHOICE TO IMPROVE PUBLIC LAW 52 (1997).

[140] U.S. CONST. amend. V; U.S. CONST. amend. XIV, § 1.

[141] *See id.* During an earlier era, the Due Process Clauses were read to require much more than this from legislation. *See, e.g.*, Lochner v. New York, 198 U.S. 45 (1905). Since the New Deal, however, the Court has swung to the other extreme—retaining a rationality requirement in name, but rendering it almost toothless unless some sort of fundamental right is implicated by state action. *See, e.g.*, Williamson v. Lee Optical, 348 U.S. 483, 488 (1955) ("The day is gone when this Court uses the Due Process Clause of the Fourteenth Amendment to strike down state laws, regulatory of business and industrial conditions, because they may be unwise, improvident, or out of harmony with a particular school of thought."); *see also* United States v. Carolene Prods., 304 U.S. 144, 153 n.4 (1938) (announcing a strong presumption of constitutionality and reserving heightened scrutiny only for situations in which legislation "appears on its face to be within a specific prohibition of the Constitution[,] . . . restricts those political processes which can ordinarily be expected to bring about repeal of undesirable legislation[, or is] . . . directed at . . . discrete and insular minorities").

[142] Professor Nagle has also attempted to re-rationalize this strand of the Court's severability jurisprudence. In place of a Due Process formulation, he suggests that courts should strike down the remaining statutory scheme only when it leads to an "absurd result." Nagle, *supra* note 14, at 235–36. There are two difficulties with his suggestion, each illustrated by Justice Kennedy's admonition that the absurd results canon is to be applied only

But this re-rationalization of the stand-alone functionality standard—the idea that a post-severance statutory scheme might raise constitutional concerns *independent* of those giving rise to the underlying lawsuit—suggests a doctrinal problem that merits consideration without regard to how the standard is formulated. Although a plaintiff challenging a regulatory scheme unquestionably has *standing* to bring both her challenge and to be heard regarding the form that a remedy for the scheme's constitutional infirmities should take (in this context, whether the statute at issue should be severed or invalidated entirely),[143] the issue of whether a post-severance statutory scheme can function in practice (upon which the severability determination allegedly pivots under this justification for disregarding a severability clause) is almost surely not *ripe* for review at the time that a litigant asserts her claim.

Consider *Abbott Laboratories v. Gardner*,[144] the Supreme Court's leading case on ripeness (both inside and outside of the administrative law context).[145] Addressing a pre-enforcement challenge seeking injunctive and declaratory relief against regulatory prescription drug labeling requirements, the Court explained:

---

when a "result . . . would be, in a genuine sense, absurd, i.e., where it is quite impossible that Congress could have intended the result, and where the alleged absurdity is so clear as to be obvious to most anyone," Pub. Citizen v. United States, 491 U.S. 440, 470–71 (1989) (Kennedy, J., concurring) (citation omitted). First, the absurd results formulation fails to remedy the key conceptual difficulty with the stand-alone functionality test; resort to the absurd results canon depends on the same underlying justification—legislative intent—even in the face of severability clauses designed by legislatures to convey their intent unambiguously. Second, courts apply the canon far too frequently—often when results are not obviously absurd. *See, e.g., id.* at 454, 465 (invoking the absurd results canon to construe a statute, but concluding only that "on the whole we are fairly confident" in the decision reached). In recent years, Nagle appears to have changed his position on the absurd results canon. *See* John C. Nagle, *Textualism's Exceptions*, ISSUES IN LEGAL SCHOLARSHIP: DYNAMIC STATUTORY INTERPRETATION #15 (2002) (*available at* http://www.bepress.com/ils/iss3/art15) (last visited Mar. 12, 2003) (arguing that the absurd results canon is unprincipled and unnecessary, decreases the likelihood of remedial legislative action, and encourages frivolous claims that squander the legal system's limited resources).

[143] Indeed, asserting that a particular remedy for an alleged constitutional violation exists is a critical component of the Court's standing jurisprudence. *See* Lujan v. Defenders of Wildlife, 504 U.S. 555, 560–61 (1992) ("[T]he irreducible constitutional minimum of standing contains three elements. First, the plaintiff must have suffered an injury in fact—an invasion of a legally protected interest which is (a) concrete and particularized, and (b) actual or imminent, not conjectural or hypothetical. Second, there must be a causal connection between the injury and the conduct complained of—the injury has to be fairly trace[able] to the challenged action of the defendant, and not the result [of] the independent action of some third party not before the court. *Third, it must be likely, as opposed to merely speculative, that the injury will be redressed by a favorable decision*.") (citations and quotations omitted) (emphasis added).

[144] 387 U.S. 136 (1967).

[145] *See* FALLON ET AL., *supra* note 15, at 249 ("Abbott Laboratories is invariably cited as the leading case on the ripeness of challenges to federal administrative regulations, and its two-part test is often applied in cases involving constitutional attacks on state and federal statutes."); *see also* Gene R. Nichol, Jr., *Ripeness and the Constitution*, 54 U. CHI. L. REV. 153 (1987) (criticizing the constitutionalization of ripeness doctrine and illustrating its broad application outside of the administrative law context).

> [Ripeness doctrine's] basic rationale is to prevent the courts, through avoidance of premature adjudication, from entangling themselves in abstract disagreements over administrative policies, and also to protect the agencies from judicial interference until an administrative decision has been formalized and its effects felt in a concrete way by the challenging parties. The problem is best seen in a twofold aspect, requiring us to evaluate both the fitness of the issues for judicial decision and the hardship to the parties of withholding court consideration.[146]

Applying the two-part test, Justice Harlan concluded that the issues raised by the manufacturers' challenge to the labeling requirements were fit for judicial review (because the primary issue tendered for adjudication was interpretive—not fact-dependent—and because the regulation was "final" for purposes of the APA),[147] and that present resolution of the issue was required (because the provision would have "a direct effect on the day-to-day business of all prescription drug companies"[148] and because "to require [plaintiffs] to challenge these regulations only as a defense to an action brought by the government might harm them severely and unnecessarily.")[149]

At the point in time when a litigant claims that a post-severance statutory scheme *may* function so poorly *in practice* that it *could* raise Due Process concerns *if* applied, it is difficult to see how that issue is ripe for adjudication under *Abbott Labs*. Regarding the fitness of the issues presented for adjudication, the question of whether a post-severance statutory scheme can function rationally in practice is assuredly fact-dependent. At least, it is not a purely legal determination akin to interpreting the meaning of statutory words. In addition, given that Congress generally reviews judicial decisions invalidating federal laws, it is not

---

[146] *Abbott Labs.*, 387 U.S. at 148–49. A more recent elaboration of the principles at stake teaches:

> Ripeness doctrine . . . mixes various mutually reinforcing constitutional and prudential considerations. One such consideration is the need to prevent the courts, through avoidance of premature adjudication, from entangling themselves in abstract disagreements. Another is to avoid unnecessary constitutional decisions. A third is the recognition that, by waiting until a case is fully developed before deciding it, courts benefit from a focus on particular facts . . . . These rationales spring, in part, from the recognition that the scope of judicial power is bounded by the Constitution. It is a principle of first importance that the federal courts are courts of limited jurisdiction. Article III of the Constitution limits jurisdiction to "cases" and "controversies," and prudential doctrines may counsel additional restraint.

Doe v. Bush, 322 F.3d 133, 138 (1st Cir. 2003) (citations and quotations omitted).

[147] *Abbott Labs.*, 387 U.S. at 149.

[148] *Id.* at 152.

[149] *Id.* at 153.

beyond contemplation that the legislature will take action on the post-severance statute before the scheme can work to harm anyone (assuming *arguendo* that a statute that cannot function can still be used to harm someone). And in any event, a reviewing court would benefit from seeing the operation of a post-severance statutory scheme in practice before it holds the scheme utterly irrational under the Due Process Clause.

If these claims fail to convince a court that the constitutional issues they must adjudicate in applying the functionality standard are not yet ripe for review, the question still remains: How will delaying a decision on the constitutional acceptability of the remaining statutory scheme harm any party to the lawsuit? It is unlikely that a truly dysfunctional post-severance statutory scheme would have far-reaching effects on private conduct that demand immediate adjudication on its constitutional merits. To use the language of *Toilet Goods Ass'n v. Gardner*,[150] a companion case to *Abbott Labs*, it is hard to see how the presence of a non-operational post-severance statutory scheme would "be felt immediately by those subject to it in conducting their day-to-day affairs." And there are not likely to be any "irremediable adverse consequences . . . from requiring a later challenge"[151] to the remaining statutory scheme in the (unlikely) event that some sort of (obviously) arbitrary action is taken against a putative plaintiff.

To sum up: The functionality standard has outlived its original justification. In many cases, it is applied unnecessarily; in others, its application raises important jurisdictional concerns. In short, wholesale invalidation of a partially unconstitutional statute is best left for another day—not handled at the end of an opinion the bulk of which is devoted to determining whether a statutory scheme is, to begin with, partially unconstitutional.

E. Severability and Changed Circumstances

A final justification for disregarding severability clauses is less sweeping and more compelling—circumstances may have changed between the enactment of a statute containing a severability clause and a judicial decision to invalidate a subsection of the statutory scheme.[152] Consider *United States v. Spokane Tribe of Indians*.[153] For some time, the Spokane Tribe had operated a bingo hall and card tables on its Washington State reser-

---

[150] 387 U.S. 158, 164 (1967).

[151] *Id.*

[152] For an illustrative discussion of the various situations in which courts might be justified in "accommodating" statutory directives to new factual circumstances, see WILLIAM N. ESKRIDGE, JR., DYNAMIC STATUTORY INTERPRETATION § 2 (1994). A synopsis of this discussion can be found in ESKRIDGE & FRICKEY, *supra* note 15, at 604–06.

[153] 139 F.3d 1297 (9th Cir. 1998).

vation.[154] Wishing to expand its gaming operations, the Tribe attempted to negotiate a required compact with the state pursuant to procedures set forth in the Indian Gaming Regulatory Act ("IGRA"),[155] which, importantly, contained a severability clause. After two years, negotiations broke down and the Tribe filed a federal lawsuit alleging that state officials had refused to bargain in good faith.[156] While the lawsuit was pending, the Tribe began to expand its gaming operations.[157]

In the meantime, the Supreme Court held in *Seminole Tribe of Fla. v. Florida* that "[t]he Eleventh Amendment prevents congressional authorization of suits by private parties against unconsenting States."[158] The State of Washington quickly invoked sovereign immunity, thereby terminating the Spokane Tribe's lawsuit,[159] and federal authorities filed suit against the Tribe in U.S. district court alleging that the Tribe's expanded gaming activities were illegal in the absence of the tribal-state gaming compact required by IGRA.[160] The court granted a preliminary injunction preventing the Tribe from operating most of its gaming activities, and the Tribe filed an interlocutory appeal.[161]

At the heart of its defense, the Tribe claimed that Congress would not have passed IGRA—with its requirement that Indian Tribes and State governments reach an accord before the Tribes could operate certain games of chance on their reservations—if it could not also authorize Indian tribes to sue states that refused to negotiate with them in good faith.[162] On this view, Congress had passed IGRA in response to the Court's decision in *California v. Cabazon Band of Mission Indians*, which held that the states lack authority to regulate gaming on Indian reservations.[163] IGRA authorized state governments to regulate Indian gaming for the first time, but in crafting the law, Congress deliberately chose not to leave the states "hold[ing] all the cards."[164] Instead, it struck a careful balance between state and tribal interests by enabling the Tribes to force states to bargain with them in good faith under threat of a federal lawsuit and subject to approval by the National Indian Gaming Commission.[165] Congress thus could not have meant what it said when it enacted IGRA's severability clause; it simply had not imagined that the Court would interpret the Eleventh Amendment to bar suits by Indian Tribes against

---

[154] *Id.* at 1298.
[155] Indian Gaming Regulatory Act ("IGRA"), 25 U.S.C. §§ 2701–2721 (1988).
[156] *Spokane Tribe*, 139 F.3d at 1298.
[157] *Id.*
[158] 517 U.S. 44, 72 (1996).
[159] *Spokane Tribe*, 139 F.3d at 1298.
[160] *Id.*
[161] *Id.*
[162] *Id.* at 1299.
[163] 480 U.S. 202 (1987).
[164] *Spokane Tribe*, 139 F.3d at 1301.
[165] *Id.*

state governments, and it would not have left the states with such disparate bargaining power over the Tribes if it had known in advance how the Court would decide *Seminole Tribe*.[166] *Spokane Tribe* thus presents a compelling example of a case in which changed circumstances might prompt a court to look beyond an explicit severability clause.[167]

How should the courts deal with such situations? Doubtless many, including the Ninth Circuit in *Spokane Tribe*,[168] would argue that crafting exceptions to (or otherwise overlooking) a severability clause is justified in these circumstances.[169] Professor Eskridge, for instance, has suggested that, in applying statutes, judges ought to consider "not only what the statute means abstractly . . . but also what it ought to mean in terms of the needs and goals of our present day society."[170] On this view, "[e]ven when one can figure out the legislature's specific intent as to an issue when it enacted the statute, there may be considerable doubt that the legislature 'would have' specifically intended that the issue be resolved in that way if it could have predicted future circumstances."[171] When legal rules and social norms have shifted, therefore, judges may read statutes in ways that "contradict as well as transcend the legislature's original expectations."[172]

In defense of this kind of "critical pragmatism,"[173] Eskridge deconstructs the textualist vision of legislative supremacy.[174] On that view, courts

---

[166] *See id.* at 1300. (quoting IGRA sponsor Sen. Daniel Inouye, who, after it became clear that the Eleventh Amendment might be interpreted to bar IGRA suits, stated: "[I]f we had known that this proposal of tribal state compacts that came from the States and was strongly supported by the States, would later be rendered virtually meaningless by the action of those states which have sought to avoid entering into compacts by asserting the Tenth and Eleventh Amendments to defeat federal court jurisdiction, we would not have gone down this path.").

[167] *See also INS v. Chadha*, 462 U.S. 919 (1983). This case suggests how changed circumstances—there, the Court's invalidation of legislative veto provisions—might prompt courts to look beyond an explicit severability clause.

[168] *Spokane Tribe*, 139 F.3d at 1302 ("[W]e [could be] left . . . with a tribe that believes it has followed IGRA faithfully and has no legal recourse against a state that allegedly hasn't bargained in good faith. Congress did not intentionally create this situation and would not have countenanced it had it known then what we know now. Under the circumstances, IGRA's provisions governing class III gaming may not be enforced against the Tribe.").

[169] Given longstanding interpretive norms favoring Native Americans, this may be especially true on the facts of *Spokane Tribe*. *See* Montana v. Blackfeet Tribe of Indians, 471 U.S. 759, 766 (1985) ("[S]tatutes are to be construed liberally in favor of Indians, with ambiguous provisions interpreted to their benefit."). *See also Choate v. Trapp*, 224 U.S. 665, 675 (1912) ("[I]n the Government's dealings with the Indians . . . construction, instead of being strict, is liberal; doubtful expressions, instead of being resolved in favor of the United States, are to be resolved in favor of a weak and defenseless people, who are wards of the nation, and dependent wholly upon its protection and good faith. This rule of construction has been recognized, without exception, for more than a hundred years . . . .").

[170] ESKRIDGE, *supra* note 152, at 50.

[171] *Id.* at 121.

[172] *Id.* at 120–21.

[173] *Id.* at 200.

[174] *See generally* William N. Eskridge, Jr., *Spinning Legislative Supremacy*, 78 GEO. L.J. 319 (1989).

are bound to act as "faithful agents" of the legislature by adhering to the legislature's explicit commands. Otherwise they impinge upon powers reserved to the legislature (and ultimately the people) in our system of separated powers.[175] Eskridge turns this analysis on its head. If courts are supposed to act as faithful agents of the legislature, he asks, shouldn't they interpret statutes to do what the legislature would have wanted the legislation to do? And, having imagined the circumstances which brought the case at hand before the court, might the legislature have voted *against* including a severability clause? Article III judges, he concludes, are "both agents carrying out directives laid down by the legislature and partners in the enterprise of law elaboration, for they (like the legislature) are ultimately agents of 'We the People.'"[176] Professor Eskridge has impressively traced this view of the judicial role to Founding-era practice, though his interpretation of the evidence is disputed.[177]

There is much to admire—but more to disagree with—in the theory of dynamic statutory interpretation. Rather than fully rehash the debate, I employ public choice theory to expose two central tensions between dynamic statutory interpretation and our constitutional structure,[178] and to provide insights into the proper resolution of severability determinations in the face of changed circumstances.

On the public choice view, politicians are rational actors who provide legislation to those interest groups and other constituencies most willing to deliver them votes, contributions, and other "payments" designed to prolong their political careers.[179] Though this vision could be criticized as cynical, it should not be surprising: It was the conscious design of the Framers. As James Madison explained in *Federalist No. 57*, Congress was

---

[175] *See, e.g.*, ANTONIN SCALIA, A MATTER OF INTERPRETATION 22 (1997) ("It is simply not compatible with democratic theory that laws mean whatever they ought to mean, and that unelected judges decide what that is."). For a detailed elaboration of the "faithful agent" theory, see John F. Manning, *Textualism and the Equity of the Statute*, 101 COLUM. L. REV. 1, 15–21 (2001).

[176] William N. Eskridge, Jr., *All About Words: Early Understandings of the "Judicial Power" in Statutory Interpretation, 1776–1806*, 101 COLUM. L. REV. 990, 992 (2001).

[177] *Compare id., with* John F. Manning, *Deriving Rules of Statutory Interpretation from the Constitution*, 101 COLUM. L. REV. 1648 (2001) [hereinafter Manning, *Deriving Rules*], *and* Manning, *supra* note 175.

[178] For another public choice-inspired analysis of dynamic statutory interpretation, see Manning, *supra* note 175, at 70–78.

[179] *See generally* DAVID R. MAYHEW, CONGRESS: THE ELECTORAL CONNECTION (1974); *see also* William M. Landes & Richard A. Posner, *The Independent Judiciary in an Interest-Group Perspective*, 18 J.L. & ECON. 875, 877 (1975) "[L]egislation is supplied to groups . . . that outbid rival seekers of favorable legislation [with payment taking] the form of campaign contributions, votes, implicit promises of future favors, and sometimes outright bribes."); Terry Sullivan, *Presidential Leadership in Congress: Securing Commitments*, *in* CONGRESS: STRUCTURE AND POLICY 286 (Mathew D. McCubbins & Terry Sullivan eds., 1987) ("[Members of Congress] are guided by a desire to maximize their expected electoral return . . . .").

> so constituted as to support in the members an habitual recollection of their dependence on the people. Before the sentiments impressed on their minds by the mode of their elevation can be effaced by the exercise of power, they will be compelled to anticipate the moment when their power is to cease, when their exercise of it is to be reviewed, and when they must descend to the level from which they were raised; there forever to remain unless a faithful discharge of their trust shall have established their title to a renewal of it.[180]

Merely subjecting legislators to popular elections was considered insufficient to safeguard liberty from legislative encroachment. In addition, prospective legislation would have to obtain majority support in both the House and Senate and be signed by the President before becoming law,[181] with intra- and inter-branch divisions secured by the expanse of the Republic and the branches' varied electoral procedures.[182] For legislation to survive this procedural gamut, sponsors must assemble a supporting coalition that reaches across party, ideological, and other interest group lines[183]—a hurdle raised even higher by the proliferation of internal leg-

---

[180] THE FEDERALIST No. 57, at 352 (James Madison) (Clinton Rossiter ed., 1961). Madison was addressing the House of Representatives, but his words are generally applicable to the Senate (in the wake of the Seventeenth Amendment) and to the President as well (despite the Electoral College).

[181] *See* U.S. CONST. art. I, § 7 (bicameralism and presentment). If the President vetoes legislation, his rejection can, of course, be overridden by a two-thirds vote of both legislative chambers. U.S. CONST. art. I, § 7, cl. 2.

[182] In *The Federalist No. 10*, for instance, Madison argued:

> Extend the [republic] and you take in a greater variety of parties and interests; you make it less probable that a majority of the whole will have a common motive to invade the rights of other citizens; or if such a common motive exists, it will be more difficult for all who feel it to discover their own strength and to act in unison with each other.

THE FEDERALIST No. 10, at 83 (James Madison) (Clinton Rossiter ed., 1961). *See also* U.S. Const. art. I, §§ 2–3 (Representatives elected for two-year terms from relatively small legislative districts); U.S. Const. art. I, § 3 & amend. XVII (Senators elected for six years by statewide popular vote); U.S. Const. art. II, §§ 1–3 (President elected to serve a four-year term by the Electoral College).

[183] *See* M. Douglass Bellis, *Drafting in the U.S. Congress*, 22 STATUTE L. REV. 38, 40 (2001) (In the American legislative process, "all the various interest groups in the country . . . must forge a compromise. Rarely does one party have enough of a majority to ignore everyone else and override a potential Presidential veto, so the compromise must have a very large consensus, usually not limited to one political party, even the majority party, to succeed. Even where the majority is strong, the existence of debate, however futile it often seems, gives the minority a chance to persuade or at least embarrass publicly the majority. In the typical case, the majority feels compelled to make some sort of compromise with the minority. Ours is not the winner takes all until the next election system that prevails in some parliamentary democracies.").

islative checkpoints designed to allow small numbers of Congressmen to impede a proposed statute's progress toward passage and enactment.[184]

Article I, Section 7 thus impedes factional domination of Congress and limits the strength of legislation. It also makes federal legislation exceedingly difficult to enact and to amend. Legislating under these requirements necessarily involves high transaction costs, and laws must therefore endure over the long-run. Otherwise, "the enacting Congress and the [interests seeking legislation] may [incur] substantial costs that would not prove worthwhile."[185] And so Article I, Section 7 serves to stabilize the legislative process by ensuring that laws do not change willy-nilly in response to modest shifts in the prevailing political consensus. This, too, was intended by the Framers.[186]

This portrait raises two objections to dynamic statutory interpretation.

> First: Given the Constitution's concern with abuse of the law-making power—evinced most notably by its subjection of members of Congress and the President to electoral control—why would it insulate courts from the political pressures to which it subjects the other branches of government if it also expects courts to exercise the same powers?
>
> Second: Why does the Constitution make legislation so difficult to enact if it expects the courts to amend those laws unilaterally in order to better address the needs of present-day society?

Understanding exactly how dynamic theorists respond to these objections is critical to resolving whether courts can legitimately disregard severability clauses in the face of changed circumstances.

In response to the first objection, Eskridge observes that "our system tolerates electoral schemes that permit candidates and parties receiving a minority of votes to win elections."[187] He then makes two claims: (1) that the power to interpret statutory enactments contrary to congressional expectations is vested in courts as part of the separation of powers, which he argues is designed to prevent a single branch of the federal government from "creating law and controlling policy";[188] and (2) that Congress can ultimately overrule a statutory decision with which it disagrees, thus

---

[184] Article I, section 5 gives each House of Congress the power to "determine the Rules of its Proceedings." U.S. CONST. art. I, § 5, cl. 2. This includes plenary control over both committee structure (and membership) and more immediately procedural matters like how to manage day to day business (including voting-related matters, like the Senate's filibuster).

[185] Landes & Posner, *supra* note 179, at 877.

[186] *See generally* THE FEDERALIST No. 63 (James Madison), Nos. 70, 78 (Alexander Hamilton).

[187] ESKRIDGE, *supra* note 152, at 156.

[188] *Id.* at 113.

minimizing any arguable judicial encroachment into the legislative domain.[189]

These arguments fail to persuade. That the Electoral College is not fully majoritarian does little to explain why the Framers felt it necessary to subject legislators, *but not* judges, to electoral control (however imperfect) if they also expected legislators *and* judges to exercise effectively equal control over enacted federal statutory policies. Eskridge's appeal to the separation of powers succumbs to the same objection it represents. A judiciary vested with power to interpret statutes in ways that contradict legislative choices wields the authority to both "create law and control policy,"[190] thus violating the same conception of the separation of powers relied upon to justify granting courts this power.

Ultimately, the analytical trump must be Congress's ability to override judicial decisions. But given that the theory under consideration posits that courts can deliberately contradict legislative enactments, endless cycling could result. If Congress enacts a statute, the Supreme Court rejects Congress's choice, and Congress then overrules the Court's decision, could the Court then overrule Congress's override? Doing so seems plainly out of place in our constitutional order (though Eskridge has come surprisingly close to suggesting that such repeated overrides are possible.)[191] In any event, absent a limiting principle that—and here's the rub—the Constitution fails to supply because its structure does not contemplate the courts' exercise of such authority, this game could go on forever. The assertion that policy choices must ultimately be left to Congress undermines any claim that courts have inherent authority to contradict unambiguous, constitutionally valid statutory directives in the first place.

Addressing the second objection, Eskridge acknowledges that legislation is purposefully difficult to enact and amend[192] and that "the assumption in both 1789 and today is that statutes will have an indefinite life—well beyond that of the enacting Congress."[193] He responds by asserting first that "An important reason for having an independent judiciary is to reassure Congress that the statutes it enacts will remain efficacious over time and not run wild or expire because of the inattention to subsequent Congresses,"[194] and second, with crucial implications for severabil-

---

[189] *Id.* at 151.

[190] *Id.* at 113.

[191] *See* William N. Eskridge, Jr. & Philip P. Frickey, *Foreword: Law as Equilibrium*, 108 HARV. L. REV. 26, 78 (1994) ("In statutory interpretation cases, *absent a strong substantive justification*, the Court *should be unwilling* to disturb a stable equilibrium, especially one that was recently reinstated by Congress in response to a temporary displacement of that equilibrium.") (emphasis added).

[192] ESKRIDGE, supra note 152 at 20 ("The constitutional goal of [the Article I, Section 7] procedural requirements is to protect against hasty, faction-driven changes in the status quo.").

[193] *Id.* at 130–31.

[194] *Id.* at 131.

ity, that it is "unrealistic to expect Congress to monitor every nook and cranny of statutory policy from year to year, for that is the reason why Congress delegates policy-making authority to agencies and courts."[195] Neither of these claims justifies rejecting a severability clause, or otherwise "interpreting" an unambiguous statute contrary to its plain meaning.

The first response presumes precisely what it is intended to justify: Judicial independence can work to reassure the legislature that its statutes will function indefinitely without constant congressional attention only if, as part of instilling that independence, the Constitution also grants courts the power to update statutes. More importantly, Professor Eskridge's second response definitively undercuts the claim that changed circumstances justify refusal to enforce severability clauses. Aside from the fact that his own research shows that Congress can and does monitor interpretive decisions,[196] severability clauses are self-consciously *not* a delegation of policymaking authority to the courts. Indeed, they are designed specifically to *remove* the severability determination from judicial control by clearly communicating congressional intent. Thus, even if Congress often delegates policymaking authority to the courts because it does not want to oversee every application of existing laws to new conditions, its passage of a severability clause tells judges, "Not this time."

None of this is to deny Eskridge's argument its due. When Congress deliberately delegates policymaking authority to the courts, or where statutory language is genuinely ambiguous, dynamic interpretation with an eye to changed or unanticipated circumstances may be appropriate. However, to the extent that we can glean insight from the structural design of the Constitution, in particular from Article I, Section 7, judges are powerless to *amend* Congress's statutes. Judicial rewriting of a statute under the guise of interpretation—which, in the severability context, involves construing a statute to mean exactly the opposite of what it says—is simply out of place in our constitutional order. In the end, it is "the prerogative of each [current] Congress to allow those laws which change has rendered nugatory to die an unobserved death if it no longer thinks their purposes worthwhile"[197] or to amend them itself to address changed circumstances.

---

[195] *Id.*

[196] *See* William N. Eskridge, Jr., *Overriding Supreme Court Statutory Interpretation Decisions*, 101 YALE L.J. 331, 338–43 (1991). Eskridge's data indicates that between 1967 and 1990, Congress overruled by statute 344 court decisions, including 124 Supreme Court opinions and 220 lower court opinions, with the number of overrides increasing over time. *Id.* at 337. In addition, the House and Senate Judiciary Committees held hearings addressing thirty to fifty percent (depending on the year) of Supreme Court statutory interpretation decisions touching on matters within those committees' jurisdiction between 1977 and 1983—a finding, he notes, "not unusual when compared with that of other committees." *Id.* at 343. These are remarkable figures—especially considering how utterly unworthy of attention the majority of interpretive decisions actually are.

[197] K Mart Corp. v. Cartier, Inc., 486 U.S. 281, 325 (1988) (Scalia, J., concurring in part).

IV. Special Problems with Non-Enforcement of Inseverability Clauses

A. *The Inseverability Clause as Structural Enforcement Mechanism*

Like severability clauses, inseverability clauses have been through Article I's exacting legislative process, and thus are valid laws due judicial enforcement. Moreover, none of the arguments for disregarding severability clauses applies in the inseverability context. Such clauses are far from boilerplate, suggesting that Congress really does know what it is doing when it includes such a directive. They pose no non-delegation threat because they deny courts any opportunity to manipulate the remaining statutory scheme in ways unimagined by the legislature. They raise no Due Process problem because they leave no remaining statutory scheme that could fail to function. And federal courts have no inherent power to leave in place a statutory scheme that Congress expressly declared it did not want.

But judicial refusal to enforce inseverability clauses raises additional concerns that are not necessarily present when courts decline to enforce severability clauses. Most legislation, as argued earlier, is the culmination of myriad deals made among competing interests. The complexities and calculations of the logrolling process, and the limits it imposes on the strength of statutes, thus mandate heightened attention to the specifics of the legislative deal. Judge Easterbrook explains:

> [Because] legislation grows out of compromises among special interests, . . . a court cannot add enforcement to get more of what Congress wanted. What Congress wanted was the compromise, not the objectives of the contending interests. The statute has no purpose. It is designed to do what it does in fact. *The stopping points are as important as the other provisions.* If the statute gave Group X twenty-five percent of what it wanted, it probably meant contending groups to keep the rest. A court cannot observe that the statute gives Group X more than it had before and then keep moving in the same direction. The compromise was that Group X would get some benefits but not more. When a court observes that Congress propelled Group X part way to its desired end, it cannot assist Group X farther along the journey without undoing the structure of the deal.[198]

When Congress includes an inseverability clause in constitutionally questionable legislation, it does so in order to insulate a key legislative

---

[198] Frank H. Easterbrook, *Foreword: The Court and the Economic System*, 98 Harv. L. Rev. 4, 46 (1984) (emphasis added).

deal from judicial interference. Such clauses are iron-clad guarantees—clear statements by Congress that it would not have enacted one part of a statute without the others. Legislation containing an inseverability clause can thus be conceived of as a contract among competing political interests containing a structural enforcement mechanism[199] designed to alleviate the concerns of those legislators who were willing to vote for, or a President who was willing to sign into law, a particular statutory scheme *only* if credibly assured that certain limiting provisions would be secure in the enacted legislation. As a result, when courts disregard an inseverability clause, there is great risk that the resulting statutory scheme will be one that could not have been enacted, and, as we shall see, that cannot be repealed. The courts' nascent approach to inseverability clauses therefore has the potential to amplify the countermajoritarian difficulty inherent in judicial review, where unelected judges undo the will of the people.[200] In most cases, disregarding an inseverability clause will not just undo that will; it will distort it.

B. The Countermajoritarian Consequences of Ignoring Inseverability Clauses

Consider the following example.[201] Following a national crisis involving the travel industry, Congress and the President rapidly move to design comprehensive legislation designed to improve airport security. Although the Democrat-controlled Senate supports nationalizing airport security, the Republican House and President wish to leave airport security in private hands. After extensive negotiations, the competing legislative interests and the President agree to nationalize airport security, but, as a key compromise designed to alleviate the Republicans' fears, they also bar the new federal airport security employees from appealing or arbitrating termination decisions. In order to ensure that the judiciary cannot abrogate this compromise, Congress includes, and the President insists upon, an inseverability clause that would render the entire legislative scheme void if any part of it is declared unconstitutional.

After working for several weeks as a federal airport security agent, John Doe is fired without cause. Doe's initial attempt to appeal the decision is rebuffed on grounds that Congress has explicitly precluded him from doing so. However, he successfully argues on appeal that the "no

---

[199] *Cf.* Friedman, *supra* note 78, at 914 (proposing a similar interpretation of the function of inseverability clauses).

[200] *See* ALEXANDER M. BICKEL, THE LEAST DANGEROUS BRANCH 16–23 (2d ed. 1986) (coining the phrase "countermajoritarian difficulty" and describing the apparent tension between judicial review and the democratic process).

[201] As will quickly become clear, this hypothetical is based loosely on the debate surrounding the structure and enactment of the airport security act crafted in the wake of September 11, 2001. *See* Aviation and Transportation Security Act, Pub. L. No. 107-71, 115 Stat. 597 (2001).

appeals" provision relied upon by the trial court violates the Due Process Clause by denying him the right to a hearing before depriving him of his property interest in continued federal employment. At that point, the appeals court must determine whether to honor the statute's inseverability clause. Despite the clause's unmistakable clarity, the court holds that, because Congress would have enacted *some* form of airport security reform legislation even without the no appeals clause, and because national security could be compromised by invalidating certain aspects of the new legislative regime for providing airport safety, the provision should be severed and the remaining statutory scheme left in place. Following the court's decision, airport security remains nationalized and federal airport security workers are able to litigate termination decisions fully—a result that neither House Republicans nor the President would have allowed to become law.

C. Could Congress Repeal the Remaining Statutory Structure?

Those tempted to favor an atextual approach to inseverability clauses might argue that a judicial decision to disregard an inseverability clause threatens to inflict little long-term harm because Congress can effectively override the judgment by repealing the remaining statutory scheme.[202] After all, Congress pays careful attention to judicial decisionmaking and it overrules statutory precedents with some frequency.[203] Considering the critical eye it casts upon ordinary statutory decisions, Congress would be quick to overrule a precedent that abrogates an important legislative compromise. And legislative inaction in the wake of a court's decision to disregard an inseverability clause could therefore be interpreted as acquiescence in the decision, representing a new, majoritarian consensus supporting the remaining statutory structure.[204] Any countermajoritarian consequences of disregarding an inseverability clause would be short-lived and non-threatening.

This argument has no more than surface appeal. To begin with, the Court has (in other interpretive cases) generally rejected the idea that legislative silence should shape judicial treatment of a statute,[205] and it

---

[202] *See, e.g.*, Robert L. Glicksman, *Severability and the Realignment of the Balance of Power Over the Public Lands: The Federal Land Policy and Management Act of 1976 After the Legislative Veto Decisions*, 36 HASTINGS L.J. 3, 92 (1984) ("[I]f the court misconstrues the legislature's intent [with regard to the severability determination], Congress is free to enact new legislation to broaden or to narrow the scope of the [decision]."); *see also* Note, *Severability of Legislative Veto Provisions: A Policy Analysis*, 97 HARV. L. REV. 1182, 1196 (1984) ("If Congress decides that the [severed] statute should not survive . . . Congress is free to repeal the statute.").

[203] *See generally* Eskridge, *supra* note 196.

[204] *See* William N. Eskridge, Jr., *Overruling Statutory Precedents*, 76 GEO. L.J. 1361, 1402 n.202 (1988) (collecting statutory interpretation opinions in which the Court has proffered a legislative acquiescence rationale for its decisions).

[205] *See* Harrison v. PPG Indus., 446 U.S. 578, 592 (1980) ("In ascertaining the meaning

did so specifically in the severability context in *Alaska Airlines*.[206] The fact that Congress "hasn't barked" simply does not signify legislative acquiescence in the Court's decisions.[207] But far more important, Article I's bicameralism and presentment requirements erect high barriers before an attempt to overcome the abrogation of a legislative deal rendered by a judicial decision refusing to enforce an inseverability clause. Our hypothetical airport security bill illustrates how.

In the wake of the appeals court's decision to disregard the bill's inseverability clause (which resulted in a nationalized airport security regime that allows employees to appeal termination decisions), disappointed House Republicans (backed by the President), initiate an attempt to repeal the remaining statutory structure and restore a privatized airport security regime. After their legislation passes the House, it stalls in the Senate, where Democrats refuse to repeal nationalization—which, of course, was precisely what they had set out to achieve in the first place (having only reluctantly accepted the no-appeals provision along the way). Thus, even though airport security could never have been nationalized in the absence of a no-appeals provision, Article I, Section 7 prevents repeal of the remaining statutory structure.

Although this unfortunate outcome is most likely to occur when control of the House, Senate, and Presidency is divided, it is important to

---

of a statute, a court cannot, in the manner of Sherlock Holmes, pursue the theory of the dog that did not bark.") (referencing Arthur Conan Doyle, *The Silver Blaze*, in THE COMPLETE SHERLOCK HOLMES (1938)). *But see* Chisom v. Roemer, 501 U.S. 380, 396 n.23 (1991) ("Congress' silence in this regard can be likened to the dog that did not bark."). There are two accepted exceptions to the Court's approach to legislative silence. First, the Court has held that "[w]here Congress includes particular language in one section of a statute but omits it in another section of the same Act, it is generally presumed that Congress acts intentionally and purposely in the disparate inclusion or exclusion." Russello v. United States, 464 U.S. 16, 23 (1983) (quoting United States v. Wong Kim Bo, 472 F.2d 720, 722 (5th Cir. 1972)). Similarly, because "Congress is presumed to be aware of both the language and the judicial interpretation of pertinent, existing law when it passes legislation[, w]here Congress knows how to say something but chooses not to, its silence is controlling." Haas v. IRS, 48 F.3d 1153, 1156 (11th Cir. 1995) (applying the *Russello* rule to language in two provisions that were part of neither the same title nor the same statute). The second exception is the traditional canon of statutory interpretation *expressio unius est exclusio alterius*—the expression of one thing is the exclusion of another. *See, e.g.*, Leatherman v. Tarrant City Narcotics Intel. & Coord. Unit, 507 U.S. 163, 168 (1993) ("The Federal Rules do address in Rule 9(b) the question of the need for greater particularity in pleading certain actions, but do not include among the enumerated actions any reference to complaints alleging municipal liability under § 1983. Expressio unius est exclusio alterius.").

[206] *Alaska Airlines*, 480 U.S. at 686 ("Congress' silence is just that—silence").

[207] *See* Camps Newfound/Owatonna, Inc. v. Town of Harrison, 520 U.S. 564, 617 (1997) (Thomas, J., dissenting) ("[E]ven more troubling, the 'preemption-by-silence' rationale virtually amounts to legislation by default, in apparent violation of the constitutional requirements of bi-cameralism and presentment. Thus, even were we wrongly to assume that congressional silence evidenced a desire to pre-empt some undefined category of state laws, and an intent to delegate such policy-laden categorization to the courts, treating unenacted congressional intent as if it were law would be constitutionally dubious.").

recognize that—even in the increasingly rare case in which those institutions are controlled by the same party[208]—the distribution of policy preferences between the political parties, and within the lawmaking branches, can render it impossible to overturn a judicial decision to disregard an inseverability clause.[209] And while it is also true that a court's inquiry into legislative history *might* reveal the compromise reached by the parties in crafting this hypothetical legislation—thereby leading the court to honor the statute's inseverability clause—there is no reason to risk that it will not. There simply is no need to resort to legislative history to confirm what is readily apparent on the face of a statute containing an inseverability clause: The remaining statutory scheme could not have been adopted in the absence of a key provision later declared unconstitutional. It is hard to justify a presumptive approach to inseverability clauses that, at the very least, contains room for such irreversible countermajoritarian errors.[210] As with severability clauses, courts are duty-bound to enforce unambiguous inseverability directives.

V. Severability in the Absence of Explicit Legislative Guidance

To this point, this Article has focused exclusively on how courts should treat unambiguous legislative instructions regarding severability. In a single word, its answer is: dispositively. But Congress does not always tell the courts what to do. Even when it knows that legislation is likely to face a constitutional challenge, it sometimes fails to enact a severability

---

[208] The same party that controlled the White House controlled forty-two of the sixty Congresses elected between 1832 and 1952. By contrast, just eight of the twenty-five Congresses elected between 1952 and 2002 (and only four elected since 1972) were controlled by the same party that controlled the Presidency. Author's calculations based on MORRIS FIORINA, DIVIDED GOVERNMENT 7 (2d ed. 1996) and BUREAU OF THE CENSUS, U.S. DEP'T OF COMMERCE, STATISTICAL ABSTRACT OF THE UNITED STATES: 2000, at 281 (2000). Although both houses of the 107th Congress (elected in November of 2000) were initially controlled by the same party that won the Presidency, Senator James Jeffords's decision to leave the Republican Party during the summer of 2001 put the Senate under Democratic control and divided control of the federal government. For that reason, I treat it as part of one of the seventeen divided governments since 1952. For more on the increasingly common phenomenon of divided government, see generally DAVID R. MAYHEW, DIVIDED WE GOVERN: PARTY CONTROL, LAWMAKING, AND INVESTIGATIONS, 1946–1990 (1991); DIVIDED GOVERNMENT: CHANGE, UNCERTAINTY, AND THE CONSTITUTIONAL ORDER (Peter F. Galderisi ed., 1996); THE POLITICS OF DIVIDED GOVERNMENT (Gary W. Cox & Samuel Kernell eds., 1991).

[209] Sarah A. Binder, *The Dynamics of Legislative Gridlock, 1947–96*, 93 AM. POL. SCI. REV. 519 (1999) (arguing that legislative gridlock is predicated on the distribution of policy preferences within political parties and across Congress, and not just upon party control itself).

[210] Such errors are irreversible not only because Article I, Section 7 can paralyze Congress's corrective attempts, but also because the extra robust form of *stare decisis* for statutory precedents makes it highly unlikely that the Supreme Court will eventually reverse its initial decision. For an illuminating discussion of the "super strong" *stare decisis* accorded to statutory precedents, see generally Eskridge, *supra* note 196.

or inseverability clause.[211] The remaining question is how courts should approach severability in the absence of explicit statutory guidance.

This Part advocates the establishment of a clear statement rule favoring severability: "Statutes are severable unless the legislature clearly states otherwise." Although commentators have criticized the Court's increasing reliance on clear statement rules, their concerns have generally focused on the norm-based nature and application of such rules.[212] Because a clear statement rule in the severability context does not implicate a norm-based inquiry, such concerns are largely inapposite here. Furthermore, a clear statement rule for severability offers the same advantages used to support the implementation of similar rules in the quasi-constitutional context, most notably that such rules establish a clear interpretive background against which future Congresses can legislate.

A. *Toward a Clear Statement Rule Favoring Severability*

As a preliminary matter, a rule favoring severability in the absence of contrary legislative guidance is well-supported by precedent. As noted earlier, the Court's earliest severability cases readily assumed that partially unconstitutional statutes should be severed.[213] In the intervening 200 years, only a handful of cases have explicitly held otherwise, and each was decided between 1928 and 1938.[214] To the extent that early practice and precedent shed light on the original understanding, they certainly suggest the propriety of a rule favoring severability.

Some might respond that the nature of statutes has changed too much to justify continued adherence to this approach. Although early federal statutes, like that confronted by the Justices in *Marbury*, were largely *ena-*

---

[211] *See, e.g.*, Ethics Reform Act of 1989, Pub. L. 101-194, 103 Stat. 1716 (codified in scattered sections of 5 U.S.C.) (partially invalidated in United States v. Nat'l Treasury Employees Union, 513 U.S. 454, 480 (1995)); Low-Level Radioactive Waste Policy Amendments Act of 1985, Pub. L. 99-240, 99 Stat. 1842 (codified at 42 U.S.C. § 2021-2022) (partially invalidated in New York v. United States, 505 U.S. 144, 188 (1992)); Sentencing Reform Act of 1984, Pub. L. 98-473, 98 Stat. 1987 (codified as amended at 18 U.S.C. § 3551-3586 and 28 U.S.C.S. §§ 991–998) (challenged repeatedly until upheld in Mistretta v. United States, 488 U.S. 361, 412 (1989)).

[212] *See, e.g.*, William N. Eskridge, Jr. & Philip P. Frickey, *Quasi-Constitutional Law: Clear Statement Rules as Constitutional Lawmaking*, 45 VAND. L. REV. 593 (1992).

[213] *See supra* Part II.A.

[214] *See* Elec. Bond & Share Co. v. SEC, 303 U.S. 419, 434 (1938) ("[A severability clause] reverses the presumption of inseparability—that the legislature intended the Act to be effective as an entirety or not at all."); Carter v. Carter Coal Co., 298 U.S. 238, 312 (1936) ("[T]he rule is against the mutilation of a statute; and if any provision be unconstitutional, the presumption is that the remaining provisions fall with it."); R.R. Ret. Bd. v. Alton R.R. Co., 295 U.S. 330, 362 (1935) ("[U]nless the act operates as an entirety it shall be wholly ineffective . . ."); Utah Power & Light Co. v. Pfost, 286 U.S. 165, 184 (1932) ("[T]he common law presumption [is] that the legislature intends an act to be effective as an entirety."); Williams v. Standard Oil Co., 278 U.S. 235, 241 (1928) ("In the absence of [a severability clause], the presumption is that the legislature intends an act to be effective as an entirety.").

bling (e.g., establishing a federal judicial structure pursuant to a specific constitutional authorization), more recent statutes, and particularly those rooted in the New Deal, are increasingly *regulatory* (e.g., fixing prices for a good and controlling the conditions under which it is produced).[215] As a result, scholars often point to the inescapable tension between "the eighteenth-century Madisonian constitutional engine of limited, internally checked government and the realities of our sprawling contemporary structures."[216] Professor Strauss elaborates:

> Laissez faire and minimal government have been replaced by welfare economics and pervasive government; the implicit assumption that the common law provides a prepolitical baseline of individual relations has been replaced by a disposition to regard all law, common and statutory, in terms of both the social ends it seeks and those it achieves. These recent developments . . . have moved government even more profoundly into the lives of citizens, regulating not only activity in the economic market-

---

[215] As Professors Eskridge and Ferejohn have observed, "[m]ost of the laws adopted in the first Washington administration were short measures addressing particular issues of maritime regulation, taxation and licensing, federal-state relations, and the mechanics of the new federal government." William N. Eskridge, Jr. & John Ferejohn, *Super-Statutes*, 50 DUKE L.J. 1215, 1223 n.22 (2001). Even the "super-statutes" of the early federal Congresses—like the Federal Judiciary Act of 1789, ch. 20, 1 Stat. 73, and the Act establishing the first Bank of the United States, Act of Feb. 25, 1791, ch. 10, 1 Stat. 191—bear little resemblance to the complex regulatory laws so common today. When the first great regulatory statutes emerged in the late nineteenth century, they were quite limited in scope. John Noyes explains:

> Although principles of administrative law had developed before 1916 and a significant bureaucracy was [then] in place . . . , most early agencies performed only routine tasks. The creation of the Interstate Commerce Commission in 1887 *signaled* the development of administrative agencies with broad regulatory powers. Yet although the ICC came to assume great adjudicatory and rule-making powers affecting the railroad industry, its abilities and desire to regulate the railroad industry developed gradually. The Act creating the Interstate Commerce Commission initially provided only weak powers, designed to combat discriminatory trade practices. Furthermore, the courts limited ICC abilities to act forcefully until the mid-1890s. [Only a]fter that [did] the Congress beg[i]n to bolster the Commission through new delegations of powers and duties.

John E. Noyes, *Implied Rights of Action and the Use and Misuse of Precedent*, 56 U. CIN. L. REV. 145, 182 n.176 (1987) (citations omitted). Other examples of early regulatory legislation include the Sherman Antitrust Act, Act of July 2, 1890, ch. 647, 26 Stat. 209 (codified at 15 U.S.C. §§ 1–2), and the Pure Food and Drugs Act of 1906, Pub. L. No. 59-384, 34 Stat. 768, but even these rested largely on delegations of quasi-legislative authority to administrative agencies, rather than upon complex, statutorily intrinsic mechanisms.

[216] Peter L. Strauss, *Sunstein, Statutes, and the Common Law: Reconciling Markets, the Communal Impulse, and the Mammoth State*, 89 MICH. L. REV. 907 (1991) (reviewing CASS SUNSTEIN, AFTER THE RIGHTS REVOLUTION: RECONCEIVING THE REGULATORY STATE (1990)).

> place but also what might seem to be more private preferences for such tastes as risk, beauty, recreation, and social milieu.[217]

Indeed, the kinds of deals that Congress makes in crafting this kind of legislation are likely to differ substantially from those it made prior to the modern era of greater federal intervention into the national economy (and private decisionmaking). Not only are modern regulatory statutes more likely than non-regulatory statutes to include interdependent provisions, the invalidation of one of those provisions is likely to unbalance a comprehensive modern regulatory scheme more profoundly than it would the kind of non-regulatory scheme generally passed before the turn of the twentieth century. Congressmen willing to vote for a statutory provision that imposes regulatory burdens on one constituency, for instance, would likely reconsider their support for that provision if they knew that the courts would hold unconstitutional another provision imposing similar burdens on other constituencies.[218] In the regulatory state, perhaps, the rule favoring severability laid down by the Court's early severability precedents, even as modified by more recent cases like *Alaska Airlines*, may no longer be desirable.

While the contours of federal "statutorification" may have changed over time, the core constitutional principle of judicial deference to legislative supremacy has not. In our system of separated powers, judicial respect for the legislature spurs numerous constitutional restraints upon the judiciary's ability to act. Article III limits the reach of judicial pronouncements to "cases" and "controversies"[219] and precludes courts from issuing advisory opinions.[220] Several traditional canons of statutory interpretation, each with arguable constitutional pedigree, similarly suggest judicial restraint: "A statute must be construed, if fairly possible, so as to

---

[217] *Id.* at 908.

[218] *See, e.g.*, 149 CONG. REC. S2096, S2152 (daily ed. Mar. 20, 2002) (statement of Sen. Nelson) ("I am pleased with the Snowe-Jeffords provision in [McCain-Feingold], which addresses some of the problems created by so-called issue ads funded by special interest groups and corporations. This provision will hold these groups more accountable for their ads by imposing strict broadcasting regulations and increasing disclosure requirements, effectively putting light where the sun does not shine in issue advocacy. Unfortunately, as many of my colleagues have pointed out, this provision is arguably the most susceptible to being struck down as unconstitutional by the Supreme Court. *If the Shays-Meehan bill had a non-severab[ility] clause that would protect it from selective dissection by the Supreme Court—which we unsuccessfully tried to include in the McCain-Feingold bill last year—I would be much more inclined to support this bill. It now seems likely that parts of this bill will be struck down in court, creating, in effect, an off-balance piece of legislation that will penalize some groups—the political parties—while giving 'issue advocacy' groups more influence.* This will alter the very basis of our political system and give disproportionate power to the least accountable groups around. I cannot support any legislation that will not only not fix our current problems but will create new ones by putting candidates of all parties at the mercy of these shadow groups, while at the same time taking away much of their ability to respond.") (emphasis added).

[219] Lujan v. Defenders of Wildlife, 504 U.S. 555 (1992).

[220] *See* Hayburn's Case, 2 U.S. (2 Dall.) 409 (1792).

avoid not only the conclusion that it is unconstitutional but also grave doubts upon that score"[221] and "Statutes are presumed constitutional"[222] are but two examples of courts' inclinations in this regard.[223] And as Alexander Bickel famously observed, a court's pronouncement that certain laws are constitutional—against a backdrop where it has the power to invalidate unconstitutional laws—serves an important legitimizing function in our constitutional democracy.[224]

Each of these notions supports holding statutes severable; toward preserving all provisions of a statute except for those that are truly unconstitutional; and toward avoiding more searching inquiry than necessary to decide the particular case in front of the court. These are reasons why, as Justice Hughes famously declared, "the cardinal principle of statutory construction is to save and not to destroy."[225] They are also the reasons why the most appropriate rule for severability would be a simple clear statement rule: "Statutes are severable unless the legislature clearly states otherwise."[226] Such a rule rests the decision with regard to severability with the people, respecting both our desires and the process that enacted them. If we are unsatisfied with the substantive statutory outcome, our remedy is at the ballot-box; severability, as a general matter, does not implicate the kinds of underenforced constitutional norms that might otherwise solicit more aggressive review.[227]

---

[221] United States v. Jin Fuey Moy, 241 U.S. 394 (1916); *see also* NLRB v. Catholic Bishop of Chi., 440 U.S. 490, 500 (1979) ("[A]n Act of Congress ought not be construed to violate the Constitution if any other possible construction remains available."). For criticism of the avoidance canon, see Frederick Schauer, Ashwander *Revisited*, 1995 SUP. CT. REV. 71, 74 ("[I]t is by no means clear that a strained interpretation of a federal statute that avoids a constitutional question is any less a judicial intrusion than the judicial invalidation on constitutional grounds of a less strained interpretation of the same statute.").

[222] Bush v. Vera, 517 U.S. 952, 991 (1996); *see also* Fairbank v. United States, 181 U.S. 283, 285 (1901) ("The presumptions are in favor of [a statute's] constitutionality, and before a court is justified in holding that the legislative power has been exercised beyond the limits granted, or in conflict with restrictions imposed by the fundamental law, the excess or conflict should be clear.").

[223] Adrian Vermeule perceptively notes several tensions between the canon of constitutional avoidance and the traditional presumption in favor of severability, most important among them the fact that applying the avoidance canon *frustrates* congressional attempts to exercise its powers as desired, while presuming severability *allows* Congress to exercise its maximal constitutional authority. *See* Vermeule, *supra* note 15, at 1961–64. The argument here is not that severability should be presumed *because* of the avoidance canon—an argument that Vermeule is right to reject—but simply to note that the perception of judicial restraint (mistakenly) underlying the Court's use of the avoidance canon supports a presumption of severability, which really does reflect judicial respect for legislative supremacy.

[224] BICKEL, *supra* note 200, at 29–33.

[225] NLRB v. Jones & Laughlin Steel Corp., 301 U.S. 1, 30 (1937).

[226] John Copeland Nagle advocates a similar clear statement rule, but expresses doubts I do not share. *See* Nagle, *supra* note 14, at 254–57.

[227] *See* United States v. Carolene Prods., 304 U.S. 144, 153 n.4 (1938). In the equal protection context, or where constitutionally troubling underinclusiveness is at issue, the choice confronted by a court is not whether to *retain or invalidate* the remaining statute, but whether to *extend or remove* the benefit or penalty improperly distributed by the law. In

B. Cautions Against a Clear Statement Rule Favoring Severability

Nevertheless, several prudential considerations may point toward a clear statement rule *against* severability: "Statutes are *in*severable unless the legislature clearly states otherwise." Just as the insights of public choice theory into the complexities of legislative deal-making demonstrate how a court's refusal to heed an inseverability clause is likely to leave in place legislation that never could have been adopted, so these insights show how a strong rule favoring severability might similarly frustrate legislative intent. In some cases, a legislative deal may not have been sufficiently concrete to warrant an inseverability clause, yet its absence still may have prevented passage of the entire measure.[228] A court that blindly adheres to a preference for severability might thereby miss the less readily apparent deals that make a statute what it is.

Even so, this probably will occur with only the greatest infrequency. As noted, Congress almost always considers severability, and it usually includes a severability clause in constitutionally questionable legislation. In very rare instances, Congress includes an inseverability clause. This suggests that in the vast majority of cases, Congress intends its statutes to be severable and that when it does not, it says so. Against a backdrop where Congress almost always favors severability, this miniscule level of risk can hardly justify a rule favoring inseverability in the face of the constitutional considerations outlined above.

Ironically, Congress's preference for severability could also be used to justify a general rule *against* severability. On this theory, a clear statement rule disfavoring severability would make severability clauses more popular in Congress, which would in turn limit judicial discretion to decide future cases. The surest way to ensure that Congress addresses severability is to discipline it into doing so: If the courts, for lack of a severability clause, wholly invalidate a statute Congress clearly intended to be severable and announce that they will continue to do so in the future, Congress will learn its lesson: It will tell the courts what to do. This, after all, is the moral of *Warren* and its progeny. Dissatisfied with the courts' invalidation of partially unconstitutional statutes under a qualified presumption of severability, legislatures *began* including severability clauses in constitutionally questionable legislation. Now, they will *always* do so.

---

such cases, the constitutional harm that prompted the court's decision may be thought to support a particular outcome to the court's inquiry. *See generally* Caminker, *supra* note 111. Courts may legitimately take into account these quasi-constitutional considerations so long as Congress has not made its preferred outcome clear in the statutory text or structure.

[228] Professor Eskridge's case study of the Civil Rights Act of 1964 and the Supreme Court's decision in *United Steelworkers v. Weber* helps illustrate how difficult it is to reconstruct how legislators would have voted on legislation lacking a key provision—and just how close these reconstructed "votes" can be. *See* ESKRIDGE, *supra* note 152, at 14–31.

This notion depends on a presumption that the courts will respect unambiguous severability directives, a dangerous presumption given the judiciary's current disposition and the fact that Congress's ever-increasing passage of severability clauses has given rise to the argument that such clauses are somehow meaningless. Even if one was willing to overlook this problem, a clear statement rule *favoring* severability would accomplish the same thing. When Congress actually intends a statute to be inseverable, against a legal background where severability is the default rule in the absence of a clear statement to the contrary, it will include an inseverability clause: Congress is presumed to know the rules of statutory interpretation and can be expected to legislate accordingly.[229] The gain in certainty produced by a clear statement rule disfavoring severability—in a world where the courts treat unambiguous severability and inseverability directives dispositively—would therefore be minimal at best.

One final argument merits attention. Partially as a consequence of the increasingly "activist"[230] approach taken by the Supreme Court (both over the long-run, as evidenced by the transformation of the Court's role under Chief Justice Warren,[231] and the recent short-run, as evidenced by the Rehnquist Court's willingness to invalidate federal legislation),[232] but also due to features that inhere in the institution of judicial review, the executive and legislative branches of the federal government are too often willing to stretch (or compromise) constitutional principles. Congress occasionally passes legislation that even supporters acknowledge poses serious constitutional concerns and presidents sometimes support legislation they believe to be constitutionally dubious, all because they sense that the courts are available as the ultimate arbiter of constitutional disputes.[233]

---

[229] McNary v. Haitian Refugee Ctr., 498 U.S. 479, 496 (1991) ("It is presumable that Congress legislates with knowledge of our basic rules of statutory construction . . . .").

[230] "Activist" here denotes *soley* the practice of invalidating legislation on constitutional grounds, *not* to imply improper judicial lawmaking (which is how the term is most often employed). For a consideration of the difference between these distinct judicial practices and a historical analysis of twentieth century "judicial activism," see Brzonkala v. Va. Polytechnic Inst. & State Univ., 169 F.3d 820, 889–98 (4th Cir. 1999) (Wilkinson, C.J., concurring).

[231] *See generally* MORTON J. HORWITZ, THE WARREN COURT AND THE PURSUIT OF JUSTICE (1998).

[232] *See* Neal Devins, *Congress As Culprit: How Lawmakers Spurred on the Court's Anti-Congress Crusade*, 51 DUKE L.J. 435 (2001) (noting that the Rehnquist Court invalidated twenty-seven laws between 1995 and 2001).

[233] *See* Joel Mowbray, *The Bush Way of Compromise*, WASH. TIMES, Apr. 12, 2002, at A23 (noting that President Bush's statement accompanying his signing of the McCain-Feingold campaign finance reform bill expressed "grave concerns" about the legislation's constitutionality); *see also* Letter from President Franklin D. Roosevelt to Rep. Samuel B. Hill (July 6, 1935) *in* 4 PUBLIC PAPERS AND ADDRESSES OF FRANKLIN D. ROOSEVELT 297–98 (1938) ("Manifestly, no one is in a position to give assurance that the proposed act will withstand constitutional tests, for the simple fact that you can get not ten but a thousand differing legal opinions on the subject. But the situation is so urgent and the benefits of the legislation so evident that *all doubts should be resolved in favor of the bill, leaving to the*

This ever-increasing willingness of the lawmaking branches to leave constitutional inquiry to the courts raises a serious concern: It not only tends to undermine the democratic process,[234] but also it diminishes the security with which citizens' rights are protected. Part of the genius of the Constitution's separation of powers is its division of governmental power among multiple centers of authority that can prevent liberty-infringing legislation from becoming law.[235] If two branches of government abdicate that responsibility to the third, the Constitution's conscientious design to create multiple layers of protection from state authority is weakened dramatically.[236]

Though it is not obvious, some might argue that a rule disfavoring severability would promote these inherently libertarian constitutional values. If Congress were faced with a real consequence for including unconstitutional provisions in legislation—the risk that its entire statutory scheme would be invalidated—it might more carefully adhere to constitutional norms. Rather than pushing against constitutional strictures for political gain, both Congress and the president might respect constitutional limitations more fully. I certainly share these concerns, but a rule disfavoring severability would do little to assuage them. So long as the legislature can enact severability clauses, and so long as courts enforce them under a clear statement rule favoring severability, Congress will be fully able to avoid any consequences from a rule favoring wholesale invalidation of partially unconstitutional legislation.

In the final analysis, Congress's overwhelming preference for severability and the long-recognized need for courts to exercise restraint in judicial review militate strongly in favor of a clear statement rule supporting severability.[237]

---

courts, in an orderly fashion, the ultimate question of constitutionality.") (emphasis added).

[234] *See* JAMES BRADLEY THAYER, JOHN MARSHALL 103–04, 106–07 (1901), *quoted in* BICKEL, *supra* note 200, at 22 ("The exercise of [judicial review], even when unavoidable, is always attended with a serious evil, namely, that the correction of legislative mistakes comes from the outside, and the people thus lose the political experience, and the moral education and stimulus that comes from fighting the question out in the ordinary way, and correcting their own errors. The tendency of a common and easy resort to this great function, now lamentably too common, is to dwarf the political capacity of the people, and to deaden its sense of moral responsibility. It is no light thing to do that."); *see also* BICKEL, *supra* note 200, at 21 (arguing that judicial review may "have a tendency over time seriously to weaken the democratic process").

[235] *See supra* text accompanying notes 179–186.

[236] Indeed, it bears noting that our constitutional design anticipates that the primary source of protection against governmental abuse will be political—not judicial, as we too often assume in the wake of *Brown v. Board of Education*, 347 U.S. 483 (1954). For a discussion of the constitutional design and the Framers' primary emphasis on non-judicial political/structural constraints, see generally Larry D. Kramer, *Foreword: We the Court*, 115 HARV. L. REV. 4 (2001), and especially pages 71–73.

[237] Of course, it would also be possible for courts to adopt a rule neither favoring nor disfavoring severability in the absence of a clear congressional statement. Courts would presumably then resort to their usual interpretive tools to address the issue. This would be the worst of all possible worlds. As Justice Scalia has noted, for the Court "to display un-

VI. CONCLUSION

The Court's current severability jurisprudence is largely outdated, and it fails to account for the constitutional requirements within which it must be crafted and the more prudential concerns that ought to animate it. I have argued here that, when confronted with unambiguous directives either to sever or to invalidate a statute entirely, federal courts are bound by constitutional norms (supported by pragmatic considerations) to give full effect to these statutory provisions. When Congress fails to address severability, principles of judicial restraint point toward holding statutes severable in the absence of a countervailing clear statement. Though *McConnell v. FEC* failed to fulfill its potential to correct eighty years of fundamentally misguided doctrinal development, the Court should seize its next available opportunity to remedy the crucial defects in its existing approach to severability.

The preceding discussion should also help demonstrate precisely why severability matters so much—not only in practice, but at the core of constitutional theory. In the modern debate over interpretive practice, severability is the single issue that most starkly pits the opposing theoretical camps against each other. Textualists have in severability and inseverability clauses prototypically unambiguous statutory directives that solicit no difficult interpretation. Indeed, they solicit no "interpretation" at all: Unlike the overwhelming majority of statutes, severability and inseverability clauses leave no gaps to be filled by a court exercising otherwise traditional interpretive powers. They contain no ambiguous terms susceptible to competing understandings among which a court can, and indeed must, choose. If the legislature's plain textual statement cannot trump judicial lawmaking here, as opponents of textualism insist it should not, then legislative action cannot trump judicial supremacy at all. Given these stakes, scholars and judges have paid far too little attention to severability doctrine and theory.

---

certainty regarding the current background rule makes all unspecifying new legislation a roll of the dice." TRW, Inc. v. Andrews, 534 U.S. 19, 38 (2001) (Scalia, J., concurring in the judgment). In deciding statutory cases, the Supreme Court's most important function (aside from resolution of the narrow statutory issue before it) is to provide Congress with a clear background of interpretive norms against which it can legislate with predictability: It should not introduce further uncertainty into an already tangled legislative process by "eschewing clear rules that might [also] avoid litigation." U.S. Airways v. Barnett, 535 U.S. 391, 412 (2002) (Scalia, J. dissenting); *see also* Kansas v. Crane, 534 U.S. 407, 424 (2002) (Scalia, J., dissenting) ("It is irresponsible to leave the law in such a state of utter indeterminacy.").

ARTICLE.

RESURRECTING RICO: REMOVING IMMUNITY FOR WHITE-COLLAR CRIME*

MICHAEL GOLDSMITH**

In this Article, Professor Michael Goldsmith argues that the recent spate of corporate scandals suggests the need to reconsider the potential civil application of federal racketeering laws to white-collar crime. He first provides an analysis of the dual purpose of RICO as a criminal and civil enforcement tool against both traditional organized crime and white-collar crime. Professor Goldsmith then argues that the courts and Congress have dismantled RICO's civil enforcement mechanism through heightened pleading requirements, artificially restrictive readings of the statutory "pattern" and "enterprise" elements, and legislative reform denying relief to many victims of securities fraud. Professor Goldsmith analyzes the Enron and Arthur Anderson scandals as exemplars of the need for RICO reform. Professor Goldsmith explains how application of recent Supreme Court cases should eliminate heightened pleading requirements for RICO actions, and he proposes legislation to effect RICO reform.

A decade ago, I published an article in these pages entitled *Judicial Immunity for White-Collar Crime: The Ironic Demise of Civil RICO.*[1] The article argued that a series of judicially imposed restrictions on civil applications of the Racketeer Influenced and Corrupt Organizations law (RICO)[2] threatened to nullify its potential effectiveness in combating "enterprise criminality."[3] These restrictions, which contravened Supreme

---

* It has often been noted that the RICO statute's title may refer to a 1931 gangster movie in which the lead gangster was named Rico. *See, e.g.*, Parnes v. Heinold Commodities, Inc., 548 F. Supp. 20, 21 n.1 (N.D. Ill. 1982) (noting the connection to the movie *Little Caesar* and Edward G. Robinson's portrayal of "Rico"). However apt the allusion to Robinson's character may have been in 1970 when Congress enacted RICO, the allusion is even more appropriate today. At the end of the movie, Robinson's dying character uttered his famous last words: "Mother of mercy—is this the end of Rico?" *See* Joseph E. Bauerschmidt, Note, "Mother of Mercy—Is this the end of Rico?"—*Justice Scalia Invites Constitutional Void-for-Vagueness Challenge to RICO "Pattern,"* 65 NOTRE DAME L. REV. 1106, 1106, n.1 (1990).

** Professor of Law, Brigham Young University. B.S., Cornell University, 1972; J.D., Cornell University 1975. The author has served as vice-chair of the U.S. Sentencing Commission and vice-chair of the ABA Criminal Justice Section RICO Committee. He expresses his profound appreciation to his research assistants, Pam Mazaheri and Benjamin McMurray, who both worked tirelessly on this project. Their efforts proved invaluable from beginning to end.

[1] Michael Goldsmith, *Judicial Immunity for White-Collar Crime: The Ironic Demise of Civil RICO*, 30 HARV. J. ON LEGIS. 1 (1993).

[2] Organized Crime Control Act of 1970, Pub. L. No. 91-452, tit. IX, 84 Stat. 922, 941–48 (1970) (codified as amended at 18 U.S.C. §§ 1505, 1961–1968, 2516–2517 (2000)).

[3] Goldsmith, *supra* note 1, at 24–38.

Court directives, had the potential to immunize white-collar offenders from both civil and criminal RICO liability.[4]

Judicial activism curtailing civil RICO has continued unabated in the ensuing years. Ironically, even the more strictly constructionist Supreme Court eventually embraced some of these revisionist restrictions.[5] RICO fared no better in Congress, which enacted legislation restricting its potential civil application by excepting securities fraud from RICO's scope.[6]

The recent spate of corporate scandals, however, suggests the need to reconsider the potential civil application of federal racketeering laws to white-collar crime. The far-reaching—and devastating[7]—effects of Global Crossing, Enron, WorldCom, and other financial debacles have arguably created a legislative climate more receptive to dealing seriously with corporate crime.[8] These corporate scandals may also temper federal judicial skeptics who have imposed heightened pleading standards upon RICO complaints against so-called "legitimate businesses."[9] More importantly, a trilogy of recent Supreme Court decisions (none of which concerned RICO) reflects the Court's renewed determination to ensure that district judges properly defer to the pleading party in deciding Rule 12(b)(6) motions to dismiss.[10] Taken together, these decisions should

---

[4] *Id.* at 41.

[5] *See* Reves v. Ernst & Young, 507 U.S. 170, 179 (1993) (adopting the rule articulated in *Bennett v. Berg*, 710 F.2d 1361, 1364 (8th Cir. 1983), that a defendant must participate in the operation or management of the enterprise itself to be liable under RICO section 1962(c)); *see also infra* notes 124–134 and accompanying text.

[6] Pub. L. No. 104-67, 109 Stat. 758 (1995) (codified as amended at 18 U.S.C. § 1964(c) (2000)).

[7] The Enron situation and other scandals have caused billions of dollars in losses to shareholders and cost tens of thousands of jobs. *See* John A. Byrne, *Fall from Grace*, BUS. WK., Aug. 12, 2002, at 51. The effect has also been felt worldwide, perhaps illustrated most notably by a nineteen-year low in the Tokyo stock market. Yuri Kageyama, *Summit Tackles Corporate Scandals: U.S. Corruption Is a New Challenge for Embattled 21-Nation Economic Group*, THE STATE (Columbia, S.C.), Oct. 24, 2002, at 13.

[8] *See* Brian Kim, Recent Development, *Sarbanes-Oxley Act*, 40 HARV. J. ON LEGIS. 235, 238 (2003). A series of newspaper articles called for legislative reform after the downfall of Enron. *See, e.g.*, Editorial, *Funny Auditing's Fallout*, L.A. TIMES, Feb. 3, 2002, at M4; Editorial, *Accountability for Accountants*, N.Y. TIMES, June 4, 2002, A18. *See also infra* note 12 and accompanying text.

[9] *See, e.g.*, Sedima, S.P.R.L. v. Imrex Co., 741 F.2d 482, 487 (2d Cir. 1984) (holding that a plaintiff must allege a racketeering injury and a prior criminal conviction to state a claim under RICO), *rev'd*, 473 U.S. 479 (1985). In *Sedima*, the Second Circuit voiced concern that RICO had "led to claims against such respected and legitimate 'enterprises' as the American Express Company, E.F. Hutton & Co., Lloyd's of London, Bear Stearns & Co., and Merrill Lynch." *Id.* at 487. Soon after the case was decided, E.F. Hutton pleaded guilty to massive fraud charges. *See* Andy Pasztor et al., *Hutton Unit Pleads Guilty in Fraud Case*, WALL ST. J., May 3, 1985, at A1. More recently, Merrill Lynch escaped a possible fraud prosecution by agreeing to pay a $100 million fine. Patrick McGeehan, *$100 Million Fine for Merrill Lynch*, N.Y. TIMES, May 22, 2002, at A1.

[10] *See* Swierkiewicz v. Sorema N.A., 534 U.S. 506 (2002); Crawford-El v. Britton, 523 U.S. 574 (1998); Leatherman v. Tarrant County Narcotics Intelligence & Coordination Unit, 507 U.S. 163 (1993).

constrain federal judges from legislating their own restrictive brand of RICO reform.

Congress and the courts have long disregarded the origin and purpose of RICO as an aid to combating organized and white-collar crime, in both the criminal and civil arenas. Consequently, although the corporate scandals have produced widespread demands for legislative reform,[11] ensuing proposals have largely overlooked RICO as a source of potential relief. In typical fashion, most reform proposals have called for creating new federal crimes, raising penalties, and increasing SEC oversight.[12] At best, however, these recommendations are necessary but not sufficient solutions; at worst, they are problematic in their own right. For example, the Sarbanes-Oxley Act of 2002, while containing measures that are certainly needed to begin combating corporate corruption, authorizes no civil relief for fraud victims and imposes lighter penalties for the destruction of corporate records than does RICO for the crimes it covers.[13] RICO, by comparison, could be a powerful weapon against corporate crime, if it enjoyed the scope and force Congress originally intended.

This Article, therefore, seeks to resurrect RICO. Part I examines the origin and dual purpose of RICO as a criminal and civil enforcement tool against both traditional organized crime and white-collar crime. Part II explains how and why federal courts and Congress, respectively, disman-

---

[11] *See, e.g.*, President's Address Before a Joint Session of Congress on the State of the Union, 38 WEEKLY COMP. PRES. DOC. 133 (Jan. 29, 2002).

[12] *See, e.g.*, Sarbanes-Oxley Act, Pub. L. No. 107-204, § 802, 116 Stat. 745, 800–01 (2002) (to be codified as amended at 18 U.S.C. §§ 1519–1520) (new criminal liability for altering and destruction of documents); § 807, 116 Stat. at 804 (to be codified as amended at 18 U.S.C. § 1348) (new criminal liability for defrauding shareholders of publicly traded companies); §§ 902–906, 116 Stat. at 805–06 (to be codified as amended at 18 U.S.C. §§ 1341, 1343, 1349–1350 and 29 U.S.C. § 1131) (enhancing penalties for white-collar crime); § 3(b)(2), 116 Stat. at 749 (to be codified as amended at 15 U.S.C. § 78u) (granting the SEC authority to conduct investigations into violations of the rules of the newly created Public Company Accounting Oversight Board); § 1103, 116 Stat. at 807–08 (to be codified as amended at 15 U.S.C. § 78u-3(c)) (authorizing the SEC under certain circumstances to seek a temporary freeze of "extraordinary payments" by a public corporation). Other legislative proposals include the Justice for Victims of Corporate Fraud Act, which would "permit certain funds assessed for securities laws violations to be used to compensate employees who are victims of excessive pension fund investments in the securities of their employers," H.R. 5496, 107th Cong. pmbl. (2002); the Consumer and Shareholder Protection Association Act of 2002, which would have established "a public-purpose, nonprofit, democratically controlled, membership Association of consumers and shareholders" with "a mandate to inform and represent consumers, shareholders, and the public interest, and to further the effective and vigorous oversight of corporate entities," S. 3143, 107th Cong. § 2(b)(2) (2002); and the Corporate Accountability in Bankruptcy Act, which would have provided "that bonuses and other extraordinary or excessive compensation of corporate insiders and wrongdoers may be included in the bankruptcy estate," S. 2901, 107th Cong. pmbl. (2002).

[13] *Compare* 18 U.S.C. § 1520(b) (Supp. 2003) (Sarbanes-Oxley provision for a ten-year maximum penalty for destruction of corporate audit and review records) *with* 18 U.S.C. § 1963(a) (2000) (providing a twenty-year penalty for investment of unlawful income obtained through a pattern of racketeering activity or control over an enterprise through a pattern of racketeering activity).

tled RICO's civil enforcement mechanism through heightened pleading requirements and artificially restrictive readings of the statutory "pattern" and "enterprise" elements, and legislative reform denying relief to many victims of securities fraud. Part III identifies how the consequences of these actions in the context of the Enron/Arthur Andersen case render relief under civil RICO problematic, if not impossible. Finally, Part IV proposes legislative reform that would restore RICO to its rightful place in commercial fraud litigation. This reform would impose liability on violators by eliminating judicially imposed obstacles to the targeting of perpetrator enterprises and corrupt professionals who facilitate their crimes. Part IV also explains how the aforementioned Supreme Court trilogy should serve, by eliminating heightened pleading requirements for RICO actions, to reverse the trend of judicial activism that has undermined civil RICO enforcement efforts since the statute's inception.

I. The Origin and Purpose of RICO

Congress enacted RICO as Title IX of the Organized Crime Control Act of 1970.[14] Although conceived in a context principally concerned with traditional organized crime,[15] Congress consciously crafted the statute to encompass a broader range of "enterprise criminality."[16] Thus, Congress defined "racketeering activity" broadly to include a spectrum of

---

[14] Pub. L. No. 91-452, tit. IX, 84 Stat. 922, 941–48 (1970) (codified as amended at 18 U.S.C. §§ 1505, 1961-1968, 2516–2517 (2000)).

[15] *See* G. Robert Blakey & Brian Gettings, *Racketeer Influenced and Corrupt Organizations (RICO): Basic Concepts—Criminal and Civil Remedies*, 53 Temp. L.Q. 1009, 1013–17 (1980). Professor Blakey is widely recognized as the principal author of RICO. *See, e.g.*, 1 Kevin P. Roddy, RICO in Business and Commercial Litigation 1–9 (1995).

[16] The law's full title is "Racketeer Influenced and Corrupt Organizations." § 901, 84 Stat. at 941. Although the problem of organized crime initially prompted Congress to consider new legislation, Senator McClellan, principal sponsor of the Organized Crime Control Act of 1970, questioned whether it "follow[s] . . . that any proposals for action stemming from that examination be limited to organized crime." 116 Cong. Rec. 18, 913–14 (1970). The legislative record is replete with references to corrupt businesses. *See* 113 Cong. Rec. 17,998 (1967) (statement of Sen. Hruska, RICO co-sponsor, mentioning infiltration and corruption of brokerage houses and accounting firms); 113 Cong. Rec. 17,950 (1967) (statement of Rep. McClory, RICO co-sponsor, observing that "business racketeers" and "criminal cartels employ staffs of attorneys, accountants, and business consultants" to "protect themselves from suit and prosecution"); 116 Cong. Rec. 592 (1970) (statement of Sen. McClellan, RICO co-sponsor, providing list of corrupted businesses that included accounting, banking, insurance, and securities firms). Accordingly, Senator McClellan stressed that Congress "has a duty not to engage in piecemeal legislation. Whatever the limited occasion for the identification of a problem, the Congress has the duty of enacting a principled solution to the entire problem." *Id. See also* United States v. Cauble, 706 F.2d 1322, 1330 (5th Cir. 1983) ("RICO's purpose is 'the imposition of enhanced criminal penalties and new civil sanctions to provide new legal remedies for all types of organized criminal behavior, that is, enterprise criminality—from simple political corruption to sophisticated white-collar crime schemes to traditional Mafia-type endeavors.'"); Michael Goldsmith, *RICO and Enterprise Criminality: A Response to Gerard E. Lynch*, 88 Colum. L. Rev. 774, 775 (1988).

offenses ranging from those commonly identified with organized crime to others more characteristic of white-collar crime.[17]

Similarly, Congress consciously chose to define other crucial elements in terms that transcend organized crime[18] and authorized expansive civil remedies—including treble damages, counsel fees, and equitable relief.[19] Based on the antitrust model,[20] Congress created these civil remedies to supplement scarce prosecutorial resources by encouraging victims to serve as "private attorneys general."[21] The antitrust experience demonstrates that treble damages are required both to deter potential violators and to provide sufficient incentive for fraud victims to assume the costs and risks of litigation.[22] From a common sense standpoint, Congress must have recognized that RICO victims would be more willing to effect pri-

---

[17] *See* 18 U.S.C. § 1961(1) (2000). Activities typical of traditional organized crime include "murder, kidnapping, gambling, arson, robbery, bribery, extortion, dealing in obscene matter, or dealing in a controlled substance." *Id.* § 1961(1)(A). Other predicates included in the original RICO statute that are less commonly associated with organized crime and more characteristic of white-collar crime include embezzlement, mail fraud, wire fraud, securities fraud, and financial institution fraud. *See id.* § 1961(1)(B), (D). Note that the original, draft definition of "criminal activity" did not include mail fraud, wire fraud, interstate theft-fraud provisions, or securities fraud. *See* 115 Cong. Rec. 6995–96 (1969). Congress added these predicates later in the legislative process, an action endorsed by the American Bar Association. *See* 18 U.S.C. § 1961(1)(B) (2000); *Measures Relating to Organized Crime: Hearings Before the Subcomm. on Criminal Laws and Procedures of the Senate Comm. on the Judiciary*, 91st Cong. 268 (1969) (testimony of Rufus King representing the ABA).

The Supreme Court has recognized that "[t]he occasion for Congress' action was the perceived need to combat organized crime. But Congress for cogent reasons chose to enact a more general statute, one which, although it had organized crime as its focus, was not limited in application to organized crime." H.J. Inc. v. Northwestern Bell Tel. Co., 492 U.S. 229, 248 (1989). The Court has also stated that the relevant consideration is whether "the predicate acts involve conduct that is 'chargeable' or 'indictable,' and 'offense[s]' that are 'punishable,' under various criminal statutes," not whether the acts are characteristic of organized crime. Sedima, S.P.R.L. v. Imrex Co., 473 U.S. 479, 488 (1985) (citing 18 U.S.C. § 1961(1) (1982)).

[18] *See* 18 U.S.C. § 1961(4) (2000) ("'[E]nterprise' includes any individual, partnership, corporation, association or other legal entity, and any union or group of individuals associated in fact although not a legal entity."). *H.J., Inc.*, 492 U.S. at 244 ("[T]he argument for reading an organized crime limitation into RICO's pattern concept . . . finds no support in the Act's text, and is at odds with the tenor of its legislative history.").

[19] 18 U.S.C. § 1964 (2000).

[20] *See, e.g.*, Agency Holding Corp. v. Malley-Duff & Assoc., Inc., 483 U.S. 143, 150 (1987); *Sedima*, 473 U.S. at 489.

[21] Rotella v. Wood, 528 U.S. 549, 557 (2000); *see also* Holmes v. Sec. Investor Prot. Corp., 503 U.S. 258, 269–70 (1992). Senator Hruska, one of RICO's original sponsors, said, "[T]he criminal provisions are intended primarily as an adjunct to the civil provisions, which I consider as the more important feature of the bill." 115 CONG. REC. 6993 (1969); *see also infra* note 23.

[22] *See* Michael Kent Block et al., *The Deterrent Effect of Antirust Enforcement*, 89 J. POL. ECON. 429, 440–44 (1981) (showing that the deterrent effect of the Justice Department's antitrust enforcement efforts come largely from the increased likelihood they create of private treble damages remedies in private actions); Michael Goldsmith, *Civil RICO Reform: The Basis for Compromise*, 71 MINN. L. REV. 827, 846–47 (1987) (citing testimony and analyses addressing inadequacies of "actual" damage remedies).

vate enforcement against white-collar, rather than organized crime, violators.[23]

Finally, none of the central types of RICO violations is limited to conventional organized crime: section 1962(a) makes it illegal for a principal to use proceeds derived from a pattern of racketeering activity to acquire, establish, or operate an enterprise affecting interstate commerce;[24] section 1962(b) makes it unlawful to acquire or maintain an interest in such an enterprise through a pattern of racketeering activity;[25] and section 1962(c) essentially prohibits conducting enterprise affairs through a pattern of racketeering activity.[26]

By centering these prohibitions on the enterprise and pattern elements rather than on individual actions, RICO revolutionized the way in which modern law conceptualized criminality. Common law judges defined crimes almost exclusively as isolated violations of the law committed by individuals.[27] Prior to 1970, statutory formulations followed this practice.[28] Consequently, recidivist activity and organizational misconduct were marginalized or ignored, and conventional evidentiary principles made a defendant's associational affiliations and prior pattern of wrongdoing inadmissible at trial.[29] These evidentiary restrictions limited the utility of criminal sanctions, because illicit enterprises routinely

---

[23] It makes sense that if private plaintiffs were bringing suit, the targets of those suits would be entities with which private individuals had dealings and were familiar, and it is more likely that such entities would be white-collar professional enterprises than illicit criminal groups.

[24] 18 U.S.C. § 1962(a) (2000).

[25] *Id.* § 1962(b).

[26] *Id.* § 1962(c).

[27] *Cf.* GEORGE FLETCHER, RETHINKING CRIMINAL LAW 115–16 (1978) ("The critical feature of this basic pattern of liability is that the commission of the crime be objectively discernible at the time it occurs. The assumption is that a neutral third-party observer could recognize the activity as criminal even if he had no special knowledge about the offender's intention.").

[28] The principal exception to this proposition involved accomplice and conspiracy liability. These doctrines, however, operated narrowly and were subject to numerous limitations. *See generally* MODEL PENAL CODE AND COMMENTARIES § 2.06 and cmts. at 295–328. Traditional criminal law proved so problematic that, in the notorious "Lucky" Luciano conspiracy prosecution, District Attorney Thomas E. Dewey had to push through what was christened the "Dewey Law" in order to permit joinder of a number of crimes in a single indictment so that he could present a holistic picture of the Luciano vice syndicate. G. Robert Blakey, *The RICO Civil Fraud Action in Context: Reflections on Bennett v. Berg*, 58 NOTRE DAME L. REV. 237, 298 n.151 (1982).

[29] *See* 1A Wigmore on Evidence § 58.2, at 1212 (1961); *see, e.g.*, *In re* Wing Y, 136 Cal. Rptr. 390, 395–96 (Ct. App. 1977); Kansas v. Thurtell, 29 Kan. 148, 149 (1883) (holding that evidence of defendant's membership in a gang of horse thieves was inadmissible); State v. Carlson, 268 N.W.2d 553, 555–59 (Minn. Sup. Ct. 1978). Under these old conspiracy rules, "the jury [could] not consider the defendant's related offenses," so evidence of other crimes was admissible only "if offered to prove 'motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident.'" Barry Tarlow, *RICO: The New Darling of the Prosecutor's Nursery*, 49 FORDHAM L. REV. 165, 168 n.4 (1980).

survived successful prosecutions of their leadership.[30] RICO changed this by making enterprise and pattern the core of each enforcement effort. By shifting the focus from prosecuting individual violators to attacking all forms of sustained enterprise criminality, Congress dramatically altered how cases are investigated,[31] prosecuted,[32] and sanctioned.[33]

RICO achieved its intended impact against traditional organized crime; indeed, it became the federal government's principal weapon in prosecuting organized crime nationwide.[34] Successful prosecutions eliminated criminal organizations, rather than merely incarcerating a few leaders,[35] and organized crime has declined dramatically as a result.[36]

---

[30] *See* PRESIDENT'S COMMISSION OF JUSTICE TASK FORCE, REPORT: ORGANIZED CRIME 4 nn.46 & 82 (1967); 115 CONG. REC. 9567 (1969) (statement of Sen. McClellan) ("Constant references have been made to the frustration resulting when the only consequence of a conviction is that organized crime . . . organizations are run by a new leader, and the organizations which are the real threat are not affected.").

[31] For example, using RICO as a theory of investigation has allowed law enforcement agencies to conduct prolonged surveillance in eavesdropping operations without triggering the federal eavesdropping law requirement that investigators terminate surveillance upon achieving their investigative objective. 18 U.S.C. § 2518(5) (2000); *cf.* United States v. Rastelli, 653 F.Supp. 1034, 1051 (E.D.N.Y. 1986) (holding that "[t]he present alleged RICO conspiracy involving multiple co-defendants and various allegations of different racketeering activity is an example of . . . a widespread conspiracy" justifying more extensive surveillance) (citing Scott v. United States, 436 U.S. 128, 140 (1978)). *See also Organized Crime: 25 Years After Valachi: Hearings Before the Senate Permanent Subcomm. on Investigations of the Comm. on Governmental Affairs*, 100th Cong. 178, 180 (1988) (statement of James Kossler, Supervisory Special Agent, Federal Bureau of Investigation, New York Division) (providing "breakdown of the enterprise theory of investigation.").

[32] *Cf.* KEVIN F. O'MALLEY ET AL., 2B FEDERAL JURY PRACTICE AND INSTRUCTIONS §§ 56.04, 56.07 (5th ed. 2000). *See also, e.g.*, United States v. Crockett, 979 F.2d 1204, 1211 (7th Cir. 1992) (allowing broad evidence of murder and extortion by an organization to be admitted at trial).

[33] Russello v. United States, 464 U.S. 16, 20–22 (1983) (forfeiture of profit and proceeds as an "interest" in the enterprise); United States v. Horak, 833 F.2d 1235, 1241 (7th Cir. 1987) (forfeiture of position in corporation); United States v. Rubin, 559 F.2d 975, 992 (5th Cir. 1977) (forfeiture of union office), *vacated on other grounds*, 439 U.S. 810 (1978); Organized Crime Control Act of 1970 § 901(a), 18 U.S.C. § 1963(a) (2000).

[34] *See Oversight on Civil RICO Suits: Hearings Before the Senate Comm. on the Judiciary*, 99th Cong. 109–11 (1985) (testimony of Stephen S. Trott, Assistant Attorney General, United States) [hereinafter *RICO Oversight Hearings*]; Tarlow, *supra* note 29, at 168–70.

[35] *See, e.g.*, United States v. Salerno, 868 F.2d 524, 527–28 (2d Cir. 1989) (convicting the leaders of a nationwide criminal society which operates through local organizations known as 'families'); *Organized Crime: 25 Years After Valachi: Hearings Before the Senate Permanent Subcomm. on Investigations of the Comm. on Governmental Affairs*, 100th Cong. 178, 180 (1988) (statement of Thomas S. Sheer, Assistant Dir. in Charge, New York Office, Federal Bureau of Investigations); 136 CONG. REC. S718–19 (daily ed. Feb. 5, 1990) (Sen. Cranston expressing opposition to the RICO Reform Act of 1989, S. 438, 101st Cong. (1989)); Brian Goodwin, *Civil versus Criminal RICO and the "Eradication" of La Cosa Nostra*, 28 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 279, 299–311 (2002); Ronald Reagan, *Declaring War on Organized Crime*, N.Y. TIMES, Jan. 12, 1986, § 6, at 26 ("The mob's internal structure has been badly weakened by prosecutions; and its methods of operation in legitimate spheres are becoming increasingly exposed."); *see also supra* note 32.

[36] *See Federal Law Enforcement Priorities: Hearing Before the S. Comm. on the Judiciary*, 104th Cong. 20 (Feb. 1995) (statement of Louis J. Freeh, Director of the Federal

Prosecutors also applied RICO successfully against other types of enterprise criminality, including a wide array of white-collar criminals. The statute, for example, proved especially effective in political corruption[37] and major fraud cases.[38] This success, however, did not carry over into the civil arena. Although RICO contains one set of prohibitions that does not distinguish between its criminal and civil applications,[39] the courts have systematically undermined these prohibitions in civil litigation.

II. THE DEMISE OF CIVIL RICO

RICO's enhanced criminal penalties initially made the statute the "new darling of the prosecutor's nursery."[40] However, since 1985, federal courts have effectively rewritten RICO to impose their own restrictive gloss on its statutory text.[41] When plaintiffs' lawyers seized upon RICO's treble damage provision to sue established businesses for fraud,[42] outraged defendants attacked RICO as an organized-crime law run amok.[43]

Bureau of Investigation); Selwyn Raab, *The Mob in Decline—A Special Report: A Battered and Ailing Mafia is Losing its Grip on America*, N.Y. TIMES, Oct. 22, 1990, at A1; Goodwin, *supra* note 35, at 281 (explaining that organized crimes' influence over the enterprises of New York was "seemingly" eliminated as a result of RICO).

[37] Indeed, its application against corrupt legislators promoted one defense attorney to observe "you know as well as I do that Congress never would have passed [RICO] if [they] ever thought they were going to use it against governors and people like that." Marro and Shannon, *Are Prosecutors Going Wild Over RICO?* LEGAL TIMES WASH., Oct. 8, 1979, at 32. *See also, e.g.*, United States v. Jannotti, 729 F.2d 213, 217 (3d Cir. 1984) (prosecuting corrupt legislators in federal ABSCAM investigation); United States v. Walsh, 700 F.2d 846, 851 (2d Cir. 1983) (same); United States v. Thompson, 685 F.2d 993, 994 (6th Cir. 1982) (holding that "The Office of Governor" was an enterprise under RICO); United States v. Marubeni, 611 F.2d 763, 764 (9th Cir. 1980) (prosecution of a private party for bribing government officials); United States v. Frumento, 563 F.2d 1083, 1090–91 (3d Cir. 1977) (affirming the RICO conviction of a state employee who misused his authority). *But see* United States v. Mandel, 415 F. Supp. 997, 1018–19 (D. Md. 1976) (dismissing RICO charge against state governor for failure to allege enterprise).

[38] *Part I, Hearings before the Subcomm. on Criminal Justice of the House Comm. on the Judiciary*, 99th Cong. 636 (1985) (Statement of the National Association of Attorneys General); *see e.g.*, United States v. Porcelli, 865 F.2d 1352, 1357–58 (2d Cir. 1989); United States v. Clark, 646 F.2d 1259 (8th Cir. 1981); Castro v. United States, 248 F. Supp. 2d 1170 (S.D. Fla. 2003); United States v. Cianci, 210 F. Supp. 71 (D.R.I. 2002).

[39] 18 U.S.C. § 1962 (2000) (containing statutory prohibitions); 18 U.S.C. §§ 1963–1964 (subjecting violators of these prohibitions to criminal and/or civil sanctions, respectively).

[40] *See* Tarlow, *supra* note 29.

[41] Examples of such restrictions include restrictive accrual rules in determining statutes of limitation, heightened pleading requirements, unusually strict standing requirements, denial of equitable relief, and rigid guidelines for finding statutory elements such as conspiracy and pattern. *See* Goldsmith, *supra* note 1, at 19–20.

[42] The Supreme Court observed in 1985 that "of the 270 known civil RICO cases at the trial court level, 40% involved securities fraud, 37% [involved] common-law fraud in a commercial or business setting, and only 9% [involved] 'allegations of criminal activity of a type generally associated with professional criminals.'" Sedima, S.P.R.L. v. Imrex Co., Inc., 473 U.S. 479, 499 n.16 (1985) (citing ABA SECTION OF CORP., BANKING, & BUS. LAW, REPORT OF THE AD HOC CIVIL RICO TASK FORCE 55–56 (1985)).

[43] *See* Philip A. Lacovara & Geoffrey F. Aronow, *The Legal Shakedown of Legitimate*

Eventually, these arguments took hold,[44] and RICO fell from darling to dunce.

Initially, RICO's breadth had created strong incentives for potential plaintiffs to initiate civil filings. For example, RICO provides that a "pattern" requires only two acts of racketeering activity within ten years of each other.[45] As most commercial transactions involve at least two mailings or interstate telephone calls, this low threshold seemingly allowed plaintiffs to convert common law fraud and breach of contract claims into civil racketeering actions.[46] Predictably, civil RICO lawsuits began to inundate federal courts.[47]

Seeking to reduce the litigation flood, courts imposed their own limitations on civil RICO.[48] Many courts required RICO plaintiffs to meet heightened specificity requirements in their pleadings.[49] These arcane restrictions were reminiscent of long-abandoned common law pleading standards, which authorized dismissal for failure to allege factual minutia.[50] Moreover, many courts, in derogation of RICO's liberal construc-

---

Business People: The Runaway Provisions of Private Civil RICO, 21 NEW ENG. L. REV. 1, 3 (1985).

[44] *See, e.g.*, L. Gordon Crovitz, *The RICO Monster Turns Against its Master*, WALL ST. J., Jan. 15, 1992, at A13; *Review & Outlook: Look Who's Saving RICO*, WALL ST. J., May 24, 1991, at A10.

[45] Organized Crime Control Act of 1970 § 901(a), 18 U.S.C. § 1961(5) (2000).

[46] *See Sedima*, 473 U.S. at 501 (Marshall, J., dissenting); Ill. Dep't of Revenue v. Phillips, 771 F.2d 312, 313 (7th Cir. 1985); Feinberg v. Katz, No. 95 Civ. 45 (CSH), 2002 U.S. Dist. LEXIS 13771, at *32 (S.D.N.Y. July 26, 2002); Dorian v. Harich Tahoe Dev., No. C-94-3387 DLJ, 1990 U.S. Dist. LEXIS 21627, at *17 (N.D. Cal. Jan. 16, 1996); 5 CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1251.1, at 352–53 (1990).

[47] The full extent of this litigation flood has been the subject of dispute. Estimates range from 4.2% to 17% of all federal filings. *See* Blakey, *supra* note 28, at 870–71.

[48] *Overview of Case Law*, 37 RICO L. REP. A-1, A-33 (2003); *see also* Goldsmith, *supra* note 1, *at* 18–38.

[49] *See, e.g.*, Miranda v. Ponce Fed. Bank, 948 F.2d 41, 44 (1st Cir. 1991) (holding that "to avert dismissal under Rule 12(b)(6), a civil RICO complaint must, at a bare minimum, state facts sufficient to portray (i) specific instances of racketeering activity within the reach of the RICO statute and (ii) a causal nexus between that activity and the harm alleged"); Hecht v. Commerce Clearing House, Inc., 897 F.2d 21, 25–26 (2d Cir. 1990) (holding that to withstand a motion to dismiss for failure to state a claim, "a RICO civil conspiracy complaint, at the very least, must allege specifically such an agreement [between the defendants to commit at least two predicate acts]"); Elliott v. Foufas, 867 F.2d 877, 881 (5th Cir. 1989) (holding that "to avoid dismissal for failure to state a claim, a plaintiff must plead specific facts, not mere conclusory allegations, which establish the existence of an enterprise"); Schiffels v. Kemper Fin. Servs., No. 91-C1735, 1993 U.S. Dist. LEXIS 6283, at *31 (N.D. Ill. May 11, 1993) (stating that "allegations of conspiracy are not measured under Rule 9(b)'s strict pleading standard, but instead are measured under the more liberal notice pleading standard of Rule 8(a)" but that "courts require that the allegations contain supportive factual allegations"). *But see* Rose v. Bartle, 871 F.2d 331, 355–56 (3d Cir. 1989) (holding that "[the 12(b)(6)] standard of review does not distinguish between RICO and non-RICO claims" and that "'[u]nder the modern federal rules [of civil procedure], it is enough that a complaint put the defendant on notice of the claims against him'").

[50] *See* Richard L. Marcus, *The Puzzling Persistence of Pleading Practice*, 76 TEX. L. REV. 1749, 1753 (1998). Historically, "pleadings were expected to do so many duties" that

tion clause,[51] interpreted various provisions restrictively to preclude some civil RICO suits entirely.[52] Given these rulings, which contradict modern pleading doctrine and create difficult-to-satisfy elements of proof, most civil RICO claims have been dismissed pre-trial for failure to state a claim.[53] This dismissal rate stands in stark contrast to the relatively rare dismissals of general civil litigation.[54]

Modern pleading doctrine disfavors dismissal motions because they may deny plaintiffs the opportunity to present their evidence.[55] A district court may not dismiss for failure to state a claim unless it finds "beyond doubt that the plaintiff can prove no set of facts in support of his claim [that] would entitle him to relief."[56] In applying this standard, a district judge must construe the complaint liberally, accept the plaintiff's allegations as true, and draw all reasonable inferences in his or her favor.[57] These established principles favoring notice pleading date back to the adoption of the Federal Rules of Civil Procedure.[58]

Under these principles, after setting forth the nature of defendant's illicit activity and identifying which statutory violations occurred, a RICO plaintiff need only generally allege the existence of an enterprise

---

a vast assortment of "strict rules" were developed to govern their adequacy. 5 WRIGHT & MILLER, *supra* note 46, § 1202, at 69 (1990). Historically, "[t]he complaint not only gave notice of the nature of the plaintiff's case but also was required to state facts constituting the cause of action."

[51] Organized Crime Control Act of 1970, Pub. L. No. 91-452, tit. IX, § 904(a), 84 Stat. 922, 947 (1970) ("The provisions of this title shall be liberally construed to effectuate its remedial purposes."); *see also* Craig W. Palm, Note, *RICO and the Liberal Construction Clause*, 66 CORNELL L. REV. 167, 167–69 (1980).

[52] *See infra* Part II of this Article.

[53] *See, e.g.*, Pamela H. Bucy, *Private Justice*, 76 S. CAL. L. REV. 1, 22 (presenting results from a survey of all federal appellate decisions in RICO civil cases rendered between 1999 and 2001 indicating affirmation of trial court dismissal or summary judgment for defendants in "[a]lmost 70%" of cases). *See also Hearing on H.R. 1046 Before the Subcomm. on Crime of the Comm. on the Judiciary*, 101st Cong. 613, 630–31 (1989) (statement of Ronald Goldstock) (reporting a sixty-five percent dismissal rate).

[54] *See* 5A WRIGHT & MILLER, *supra* note 46, § 1357, at 321–25 (1990) ("The motion to dismiss is viewed with disfavor and is rarely granted."); *see also* Marcus, *supra* note 50, at n.34 (citing a 1975 study that showed a six percent dismissal rate and a 1988 study that showed a three percent dismissal rate, both for the federal district courts in the District of Maryland and the Eastern District of Pennsylvania).

[55] Even in a fact-pleading regime, dismissal at the pleading stage "does not provide a reliable method for determining whether a defendant has violated the plaintiff's rights because it requires the plaintiff to marshal evidence before conducting discovery." Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 COLUM. L. REV. 433, 436 (1986). If this is true in fact pleading systems, it is even more so for notice pleading, which requires an even thinner recitation of the underlying facts. A complaint need only contain "a short and plain statement of the claim showing that the pleader is entitled to relief." FED. R. CIV. P. 8(a)(1).

[56] Conley v. Gibson, 355 U.S. 41, 45–46 (1957).

[57] *See, e.g.*, Viqueria v. First Bank, 140 F.3d 12, 16 (1st Cir. 1998); Whisman v. Rinehart, 119 F.3d 1303, 1308 (8th Cir. 1997); *see also* 5A WRIGHT & MILLER, *supra* note 46, § 1357, at 304.

[58] 5 WRIGHT & MILLER, *supra* note 46, § 1202, at 68–76.

and pattern of racketeering activity.[59] Applied properly, Rule 8(a), which provides general rules of pleading, does not require detailed allegations of the enterprise or pattern of racketeering elements.[60] These are matters of proof that are properly addressed at summary judgment (after completion of the discovery process) or at trial.[61]

Despite these clear principles, however, many courts have dismissed RICO claims for failure to allege a proper pattern of racketeering activity, a proper enterprise, or both.[62] These dismissals have stemmed from a combination of heightened pleading requirements and unduly narrow judicial interpretations of the pattern and enterprise elements. Though designed to curtail frivolous RICO claims, such judicial intervention has broadly undermined the statute's utility as a weapon against commercial fraud.

A. The Pattern Element

The meaning of "pattern of racketeering activity" first became an important issue as a result of a footnote in the Supreme Court's decision in *Sedima, S.P.R.L. v. Imrex, Co.*[63] The Court asserted "that, in its private civil version, RICO is evolving into something quite different from the original conception of its enactors."[64] Despite this "increasing divergence,"[65] however, the Court declined to rewrite the statute: "It is true

---

[59] *See* Seville Indus. Mach. Corp. v. Southmost Mach. Corp., 742 F.2d 786, 789–90 (3d Cir. 1984) (holding that a RICO complaint satisfied the rules of pleading where it "identified the four entities it believed were the enterprises that had been marshaled against it").

[60] Rule 8(a) requires only three things: "(1) a short and plain statement of the grounds upon which the court's jurisdiction depends . . . , (2) a short and plain statement of the claim showing that the pleader is entitled to relief, and (3) a demand for judgment for the relief the pleader seeks." FED. R. CIV. P. 8(a); *see also* 5 WRIGHT & MILLER, *supra* note 46, § 1251.1, at 344 (stating that some cases "take the position that the basic Rule 8 standards should be applied to RICO cases and do not demand, other than in the context of Rule 9(b), that the elements of an action under the statute be pleaded with particularity").

[61] *See, e.g.*, Swistock v. Jones, 884 F.2d 755, 758 (3d Cir. 1989) (reversing a dismissal for failure to allege pattern where plaintiffs would "have the opportunity to have their pattern allegations threshed out in discovery" and where "many of these issues [would] then be susceptible to resolution via summary judgment"); *Seville*, 742 F.2d at 790 (stating that "it is enough that a complaint put the defendant on notice of the claims against him" and that "[i]t is the function of discovery to fill in the details, and of trial to establish fully each element of the cause of action").

[62] *See* 36 RICO L. REP. B7–B9, B27–B39 (2002).

[63] 473 U.S. 479 (1985). Before this case reached the Supreme Court, the Second Circuit held that in order to state a claim, the complaint must allege a "racketeering injury," which was one "different in kind from that occurring as a result of the predicate acts themselves, or not simply caused by the predicate acts, but also caused by an activity which RICO was designed to deter." Sedima, S.P.R.L. v. Imrex Co., Inc., 741 F.2d 482, 496 (2d Cir. 1984). The Second Circuit also held that a prior criminal conviction was a prerequisite to a civil RICO action. *Id.* The Supreme Court rejected these limitations. *Sedima*, 473 U.S. at 488, 495.

[64] *Sedima*, 473 U.S. at 500.

[65] *Id.*

that private civil actions under the statute are being brought almost solely against [respected businesses], rather than against the archetypal, intimidating mobster. *Yet this defect—if defect it is—is inherent in the statute as written, and its correction must lie with Congress.*"[66] The Supreme Court attributed the "'extraordinary' uses to which civil RICO has been put" to "the breadth of the predicate offenses, in particular the inclusion of wire, mail, and securities fraud, and the failure of Congress and the courts to develop a meaningful concept of 'pattern.'"[67] The Court provided an explanatory footnote suggesting how to narrow the statute's application via the pattern element:

> [T]he definition of a "pattern of racketeering activity" differs from the other provisions in § 1961 in that it states that a pattern "*requires* at least two acts of racketeering activity," . . . not that it "means" two such acts. The implication is that while two acts are necessary, they may not be sufficient. Indeed, in common parlance, two of anything do not generally form a "pattern." The legislative history supports the view that two isolated acts of racketeering activity do not constitute a pattern.[68]

This text quickly became perhaps the most litigated footnote in Supreme Court history, as courts struggled to define the "pattern" element.[69] A variety of standards emerged, at least one of which routinely produced pretrial dismissals for failure to state a proper pattern.[70]

---

[66] *Id.* at 499 (emphasis added) (citation omitted).

[67] *Id.* at 500.

[68] *Id.* at 497 n.14.

[69] Concurring *in H.J. Inc. v. Northwestern Bell Tel.*, Justice Scalia wrote that footnote 14 "promptly produced the widest and most persistent Circuit split on an issue of federal law in recent memory." H.J. Inc. v. Northwestern Bell Tel. Co., 492 U.S. 229, 251 (1989) (Scalia, J., concurring).

[70] Four major views emerged. The least restrictive standard held that a small number of related racketeering acts could be enough to allege pattern. *See, e.g.*, United States v. Ianniello, 808 F.2d 184, 192 (2d Cir. 1986); R.A.G.S. Couture, Inc. v. Hyatt, 774 F.2d 1350, 1355 (5th Cir. 1985). Under this standard, two or three letters or phone calls regarding a single fraudulent transaction could satisfy the pattern element.

Another line of cases held that "multiple episodes" of a single scheme would satisfy the pattern element. *See, e.g.*, Morgan v. Bank of Waukegan, 804 F.2d 970, 975–76 (7th Cir. 1986); Temporaries, Inc. v. Md. Nat'l Bank, 638 F. Supp. 118, 123–24 (D. Md. 1986) ("If there is only one scheme, then the scope of the scheme itself can be scrutinized for factors indicating continuity"); Soper v. Simmons Int'l, Ltd., 632 F. Supp. 244, 253 (S.D.N.Y. 1986) (quoting *Graham*); Graham v. Slaughter, 624 F. Supp. 222, 225 (N.D. Ill. 1985) ("while a RICO claim must involve different criminal episodes, . . . an open-ended scheme may include a sufficient number of independent episodes to satisfy the 'continuity' factor of *Sedima*"). Some courts have also required that multiple episodes must project a threat of continuity. *See, e.g.*, Ghouth, 642 F. Supp. at 1337.

The third line of cases looked for "continuity plus relationship," which the Court highlighted in *Sedima*, to find a pattern. *See Sedima*, 473 U.S. at 496 n.14. Some courts held that the presence of these factors was enough for a pattern. *See, e.g.*, Sun Sav. & Loan Ass'n v. Dierdorff, 825 F.2d 187, 191–94 (9th Cir. 1987); Roeder v. Alpha Ins., 814 F.2d

Properly understood, however, the Supreme Court's cautionary footnote concerned matters of proof rather than pleading. Nothing in the *Sedima* opinion suggested a shift from conventional standards for evaluating a claim under Rules 8(a) and 12(b)(6). On the contrary, in rejecting efforts to impose new pleading requirements on civil RICO complaints, *Sedima* emphasized that "a violation of § 1962(c) requires (1) conduct (2) of an enterprise (3) through a pattern (4) of racketeering activity [and while] [t]he plaintiff must, of course, *allege* each of these elements to state a claim . . . the statute requires no more than this."[71]

Thus, to satisfy modern federal pleading rules, a section 1962(c) complaint need only allege that the defendant, employed by or associated with an enterprise, conducted the affairs of such enterprise through a pattern of racketeering activity.[72] The Supreme Court's call for more discerning analysis of the pattern element simply meant that the statute requires more rigorous *proof* of this element. At the *pleading* stage, however, courts still had to draw all inferences and view the complaint in the light most favorable to the pleading party.[73] The liberal notice pleading standards of the Federal Rules of Civil Procedure merely require plaintiffs to allege the essence of their claim.[74] RICO plaintiffs could properly be put to their proof only upon completion of pre-trial discovery. At that point, failure to produce evidence of the requisite pattern warrants adverse summary judgment.[75]

A subsequent Supreme Court decision bears out this analysis. In *H.J. Inc. v. Northwestern Bell Telephone*,[76] the Supreme Court rejected the

---

22, 30–31 (1st Cir. 1987). Other courts have taken a multi-factored approach to determine whether "continuity plus relationship" exists. *See, e.g.*, Med. Emergency Serv. Assocs. v. Foulke, 844 F.2d 391, 394–95 (7th Cir. 1988) (setting forth a variety of factors to consider such as number and variety of predicate acts, their duration, presence of separate schemes, number of victims, nature of injury); Barticheck v. Fidelity Union Bank, 832 F.2d 36, 38–39 (3d Cir. 1987). Morgan v. Bank of Waukegan, 804 F.2d 970, 975 (7th Cir. 1986).

The most restrictive view was the Eighth Circuit's "multiple schemes" approach, which routinely resulted in pre-trial dismissal. *See* Edwards v. First Nat'l Bank, 872 F.2d 347, 351 (10th Cir. 1989); Superior Oil Co. v. Fulmer, 785 F.2d 252, 254–57 (8th Cir. 1986); *see also infra* note 78 and accompanying text. Under this view, in order to satisfy the pattern element, a plaintiff had to allege two separate schemes, each one having a unique objective. The United States Supreme Court ultimately repudiated this view in *H.J. Inc. v. Northwestern Bell Tel. Co.*, 492 U.S. 229, 234–35 (1989).

[71] *Sedima*, 473 U.S. at 496–97 (emphasis added).

[72] *See* FED. R. CIV. P. app. of forms 3–13. If the claim is grounded in fraud, plaintiff must also detail the underlying fraud with particularity. FED. R. CIV. P. 9(b).

[73] *See* Haroco v. Am. Nat'l Bank & Trust Co., 747 F.2d 384, 404 (reversing a dismissal for failure to state a claim because there are "no grounds for demanding that a civil RICO plaintiff essentially plead evidence and prove the case in the complaint"), *aff'd on other grounds*, 473 U.S. 606 (1985).

[74] *See supra* notes 55–61 and accompanying text. For example, the Appendix of Forms accompanying the Federal Rules of Civil Procedure contains numerous examples of concise pleadings that set forth only the essence of plaintiff's claim. *See, e.g.*, FED. R. CIV. P. app. of forms 3–13.

[75] *See* FED. R. CIV. P. 56(c), (e).

[76] 492 U.S. 229, 249–50 (1989).

most restrictive judicially imposed definition of pattern and sustained the plaintiff's complaint based on a looser standard for the proper allegation of pattern. *H.J. Inc.* concerned the Eighth Circuit's holding that multiple schemes are essential to a pattern of racketeering activity.[77] Under this view, a single illicit scheme could never qualify as a RICO pattern regardless of its duration, number of victims, or the extent of loss.[78] The Supreme Court overruled this narrow interpretation as inconsistent with RICO's statutory text, legislative history, and underlying policies.[79] Instead, Justice Brennan's majority opinion established a flexible test for satisfying the pattern element: "[w]hat a plaintiff or prosecutor must prove is continuity of racketeering activity, or its threat, *simpliciter.*"[80] Because the plaintiffs had alleged multiple bribes, albeit in a single scheme, the Court held that they "may be able to prove that the multiple predicates constitute 'a pattern of racketeering activity.'"[81]

Even under this relaxed standard, however, many courts continued to dismiss RICO complaints for failure to allege a proper pattern of racketeering activity. Rather than viewing the complaint in the light most favorable to the pleading party, these courts read artificial temporal or other requirements into the pattern element and assumed that the plaintiff could not adduce evidence at trial to satisfy these requirements.[82] Consequently, dismissals for failure to allege a proper pattern are probably as common today as before *H.J., Inc.*[83]

---

[77] *Id.* at 234.

[78] *See id.*; *see also* Michael Goldsmith, *RICO and "Pattern": The Search for "Continuity Plus Relationship,"* 73 CORNELL L. REV. 971, 988 (1988).

[79] *See H.J. Inc.*, 492 U.S. at 241; *see also* 5A Wright & Miller, *supra* note 46, § 1376, at 590.

[80] *H.J. Inc.*, 292 U.S. at 241. The Court elaborated on this standard as follows:

> "Continuity" is both a closed- and open-ended concept, referring either to a closed period of repeated conduct, or to past conduct that by its nature projects into the future with a threat of repetition . . . . A party alleging a RICO violation may demonstrate continuity over a closed period by proving a series of related predicates extending over a substantial period of time. Predicate acts extending over a few weeks or months and threatening no future criminal conduct do not satisfy this requirement: Congress was concerned in RICO with long-term criminal conduct. Often a RICO action will be brought before continuity can be established in this way. In such cases, liability depends on whether the threat of continuity is demonstrated.

Id. at 241–42.

[81] *Id.* at 250.

[82] The following examples illustrate some of the reasons courts have dismissed complaints for failure to allege a proper pattern of racketeering activity: the racketeering activity was conducted for less than one year; the racketeering activity did not constitute a regular part of the defendant's business; the racketeering activity ended before the suit was commenced; there was only one victim; there was no pattern as to each victim; there was not a close enough relationship between the predicate acts; there was no "significant societal" injury; or there was no organized-crime connection. *See* Goldsmith, *supra* note 1, at 20–22.

[83] *See* G. Robert Blakey & John Robert Blakey, *The Racketeer Influenced Corrupt Or-*

B. The Enterprise Element

The enterprise element has encountered similar judicial resistance.[84] In contrast to the pattern element, RICO defines enterprise by providing a non-exhaustive list of qualifying examples.[85] Although this definition clearly includes a "group of individuals associated in fact,"[86] the First Circuit held in *United States v. Turkette* that a RICO enterprise does not reach a group organized solely for illicit purposes.[87] Without elaborating, the court reasoned that the RICO statute dictated this outcome to avoid an evidentiary overlap between its enterprise and pattern elements; the court maintained that this overlap would render the enterprise element redundant.[88] This remarkable, albeit incoherent, analysis easily could be interpreted to mean that the Organized Crime Control Act of 1970 did not apply to organized crime.[89]

The Supreme Court promptly reversed. Responding to the First Circuit's puzzling conclusion that applying RICO to illicit groups threatened to merge the pattern and enterprise elements, the Supreme Court explained that "[w]hile the proof used to establish these separate elements may in particular cases coalesce, proof of one does not necessarily establish the other."[90] Justice White's majority opinion distinguished between the types of evidence that ordinarily would be used to prove these separate elements.[91]

---

ganizations Act (RICO), *in* WHITE-COLLAR CRIME 1999 D-29 n.80 (ABA Center for Continuing Legal Educ. ed., 1999) ("Circuit court opinions indicate that district courts regularly grant dismissals for a failure to show a 'pattern,' and they are regularly affirmed on appeal." They cite extensive post-*H.J. Inc.* authority.).

[84] *See* Goldsmith, *supra* note 1, at 22–38.

[85] *See* Organized Crime Control Act of 1970 § 901(a), 18 U.S.C. § 1961(4) (2000) ("'enterprise' *includes* any individual, partnership, corporation, association, or other legal entity, and any union or group of individuals associated in fact although not a legal entity") (emphasis added).

[86] *Id.*

[87] United States v. Turkette, 632 F.2d 896, 899 (1st Cir. 1980), *rev'd*, 452 U.S. 576 (1981).

[88] *Id.*

[89] The absurdity of the First Circuit's result is manifest in the colloquy between the Justices and Turkette's attorney during oral argument before the Supreme Court. Chief Justice Burger asked, "But is it not possible that a corporation could be organized and do nothing but deal in stolen goods or stolen securities . . . and be wholly illegitimate? Would you say that . . . a corporation, so organized and totally illegitimate would not be an enterprise within the meaning of the statute?" United States v. Turkette, No. 80-808, 1981 U.S. TRANS LEXIS 66, at *31 (U.S. Apr. 27, 1981). Chief Justice Burger asked counsel whether "some group [might] . . . set up a specialty of collecting illegal debts. That is, usurious debts, gambling debts . . . and that's all they do. You say they are not covered? . . . You mean they can't be prosecuted under this statute?" *Id.* at *36. *See also* United States v. Turkette, 452 U.S. 576, 583 (1981) ("That a wholly criminal enterprise comes within the ambit of the statute does not mean that a 'pattern of racketeering activity' is an 'enterprise.'").

[90] *Turkette*, 452 U.S. at 583.

[91] *Id.* ("[Enterprise] is proved by evidence of an ongoing organization, formal or informal, and by evidence that the various associates function as a continuing unit. [Pattern]

Subsequent criminal cases elaborated further on the proof required to establish a RICO enterprise comprised of a group of individuals associated in fact.[92] Because RICO's statutory prohibitions do not distinguish between their potential civil and criminal applications,[93] judicial interpretations of the enterprise element arising in criminal prosecutions also govern civil RICO cases. These decisions, which required proof of commonality of purpose, structure, and some continuity of membership, made sense because they ensured that RICO would not otherwise apply to a handful of ad hoc violators who committed a few crimes together.[94] In an effort to limit the scope of civil RICO, however, federal courts transformed these evidentiary elements of proof to pleading requirements for civil RICO complaints, requiring particularity in pleading the nature of the RICO enterprise.[95]

Rule 8(a) of the Federal Rules of Civil Procedure, which lays out the general rules of pleading, does not impose this level of specificity.[96] Nor does Rule 9(b), which requires plaintiffs alleging fraud to plead with particularity, impose such a heightened requirement.[97] On the contrary, Rule 9(b) requires claimants only to particularize the nature of the un-

---

is proved by evidence of the requisite number of acts of racketeering committed by the participants in the enterprise.").

[92] *See* United States v. Riccobene, 709 F.2d 214, 222–24 (3d Cir. 1983); United States v. Bledsoe, 674 F.2d 647, 664–65 (8th Cir. 1982) (holding that proof of an enterprise requires proof of three characteristics: "common or shared purpose," "continuity of both structure and personality," and "'ascertainable structure' distinct from that inherent in the conduct of a pattern of racketeering activity" (quoting United States v. Anderson, 626 F.2d 1358, 1372 (8th Cir. 1980)).

[93] *See supra* note 39.

[94] As the Eighth Circuit observed in *Bledsoe*, "unless the inclusion of the enterprise element requires proof of some structure separate from the racketeering activity and distinct from the organization which is a necessary incident to the racketeering, the Act simply punishes the commission of two of the specified crimes within a 10-year period." *Bledsoe*, 674 F.2d at 664.

[95] *See, e.g.*, Richmond v. Nationwide Cassel L.P., 52 F.3d 640, 645 (7th Cir. 1995); Frank v. D'Ambrosi, 4 F.3d 1378, 1386 (6th Cir. 1993); Bill Buck Chevrolet, Inc. v. GTE Fla., Inc., 54 F. Supp. 2d 1127, 1135 (M.D. Fla. 1999); *see generally* Thomas S. O'Neill, Note, *Functions of the RICO Enterprise Concept*, 64 NOTRE DAME L. REV. 646 (1989).

[96] FED. R. CIV. P. 8(a).

[97] FED. R. CIV. P. 9(b). *E.g.*, Bulkmatic Transp. Co. v. Pappas, No. 99 Civ. 12070 (RMB) (JCF), 2001 U.S. Dist. LEXIS 6894, at *9 (S.D.N.Y. May 11, 2001) (citing Stevelman v. Alias Research Inc., 174 F.3d 79, 84 (2d Cir. 1999)); Shields v. Citytrust Bancorp, Inc., 25 F.3d 1124, 1127–28 (2d Cir. 1994) (requiring a securities fraud complaint to "(1) specify the statements that the plaintiff contends were fraudulent, (2) identify the speaker, (3) state where and when the statements were made, and (4) explain why the statements were fraudulent"); Int'l Motor Sports Group, Inc. v. Gordon, No. 98 Civ. 5611 (MBM), 1999 U.S. Dist. LEXIS 12610, at *10 (S.D.N.Y. Aug. 16, 1999) (balancing the heightened pleading requirements of Rule 9(b) with the notice pleading requirements of Rule 8(a) and with the Rule 8(f) requirement that "all pleadings shall be so construed as to do substantial justice"); Madara v. Singular Music Publ'g Co., Inc., Civ. Action No. 84-0006, 1987 U.S. Dist. LEXIS 3348, at *8 (E.D. Pa. Apr. 28, 1987) ("Plaintiffs [in fraud case] need only allege the time, place, and contents of the false representation, as well as the identity of the parties committing the fraud and how they benefited.") (citing Seville Indus. Mach. Corp. v. Southmost Mach. Corp., 742 F.2d 786, 791 (3d Cir. 1984)).

derlying fraud—when and where did who do what to whom.[98] Rule 9(b) does not require complainants to specify every element of their claim.[99] Indeed, Congress has declined to enact proposals adding such a requirement to RICO.[100]

Yet heightened specificity is often required as a result of decisions requiring RICO plaintiffs to allege the enterprise's commonality of purpose and structure, as well as some continuity of membership, in support of complaints based on association-in-fact enterprises.[101] As business conspiracies are often complex and involve an intricate array of entities, relatively few victims can satisfy such demanding pleading requirements without extensive pre-trial discovery.[102] Consequently, dismissals for fail-

---

[98] 2 JAMES WM. MOORE ET AL., MOORE'S FEDERAL PRACTICE ¶ 9.03[1][b] (3d ed. 2002). Even these requirements apply flexibly. For example, in *Seville Industrial Machinery Corp. v. Southmost Machinery Corp.*, 742 F.2d 786 (3d Cir. 1984), the Third Circuit observed:

> Rule 9(b) requires plaintiffs to plead with particularity the "circumstances" of the alleged fraud in order to place the defendants on notice of the precise misconduct with which they are charged, and to safeguard defendants against spurious charges of immoral and fraudulent behavior. It is certainly true that allegations of "date, place or time" fulfill these functions, but nothing in the rule requires them. Plaintiffs are free to use alternative means of injecting precision and some measure of substantiation into their allegations of fraud.

Id. at 791.

[99] *See* 5 WRIGHT & MILLER, *supra* note 46, § 1298, at 624. *See, e.g.*, SEC v. Davis, 689 F. Supp. 767 (E.D. Ohio 1988); Bethlehem Steel Corp. v. Fischbach & Moore, Inc., 641 F. Supp. 271 (E.D. Pa. 1986); SEC v. Platt, 565 F. Supp 1244 (D. Okla. 1983) (noting that fraud claims do not require pleading "evidence").

[100] The RICO Reform Act of 1989 would have required plaintiffs to plead their complaints with particularity. *See* S. 438, 101st Cong. § 5 (1989); *see also* S. REP. NO. 101-269, at 9 (1990) (endorsing particularity requirement "for all civil RICO damage actions and for all elements of each RICO claim whether or not fraud is involved"). This bill was never considered by the full Senate. Bill Summary and Status for the 101st Congress, S. 438, *available at* http://thomas.loc.gov/cgi-bin/bdquery/D?d101:1:./temp/~bd0qNu:@@@L&summ2=m&l/bss/d101query.htmll.

[101] *See, e.g.*, Stachon v. United Consumers Club, Inc., 229 F.3d 673, 676–77 (7th Cir. 2000) (affirming dismissal for failure of the allegations to show continuity of membership "as an ongoing RICO organization"); Shapo v. O'Shaughnessy, 246 F. Supp. 2d 935, 962 (N.D. Ill. 2002) (finding failure to plead adequately an association-in-fact where plaintiff "only alleges in a conclusory fashion that the association-in-fact enterprise 'functioned in a hierarchical decision-making structure as a continuing unit from at least 1989 to 1996'"); *In re* SmithKline Beecham Clinical Labs., Inc., 108 F. Supp. 2d 84, 94–97 (D. Conn. 1999) (dismissing a RICO claim on both enterprise and association-in-fact enterprise theories for failure to allege facts establishing that members operated with continuity of structure and that there was a common purpose for each member); Moll v. U.S. Life Title Ins. Co., 654 F. Supp. 1012, 1032 (S.D.N.Y. 1987) (finding plaintiffs failed to allege facts to establish how individuals joined together in a group functioning as a "continuing unit" with "continuity of structure").

[102] At least one court has acknowledged the difficulty in holding civil RICO plaintiffs to heightened pleading requirements. Schiffels v. Kemper Fin. Servs., Inc., No. 91-C1735, 1993 U.S. Dist. LEXIS 6283, at *34 n.14 (N.D. Ill. May 11, 1993) (deeming it "immaterial that at the pleading stage, a plaintiff cannot point to a specific telephone conversation or face-to-face meeting" while noting that "'[t]he courts have recognized that the nature of

ure to allege a proper association-in-fact RICO enterprise routinely occur.[103]

These dismissals have proven to be especially problematic because they preclude the most viable option for overcoming another restrictive doctrine known as the person-enterprise distinction.[104] Congress based civil RICO's treble damages provision on the premise that it would deter future racketeering activity by encouraging victims to file suits.[105] This premise necessarily contemplates litigating against a solvent violator. A series of strained interpretations, however, has defeated this policy by holding that a section 1962(c) RICO complaint may not name a defendant as both the alleged *violator*—that is, the "person" in the statute's language[106]—and the RICO *enterprise*.[107] This result aims to avoid imposing RICO liability vicariously upon companies that may have been victimized themselves by employee criminality.[108] Such protection, however, is unnecessary, because vicarious liability does not apply to mere victims.[109] Moreover, this doctrine creates an internal inconsistency within the RICO act: although section 1962(c) prohibits conducting enterprise affairs through a pattern of racketeering activity, the judicially imposed person-enterprise doctrine mandates dismissal of RICO complaints alleging that an institutional violator conducted its own affairs through such a pattern.[110] This result often insulates perpetrator entities from liability.[111] Thus, in such cases, RICO is powerless to reach the en-

---

conspiracies often makes it impossible to provide details at the pleading stage and that the pleader should be allowed to resort to the discovery process and not be subjected to a dismissal of his complaint'" (quoting 5 WRIGHT & MILLER, *supra* note 46, § 1233)).

[103] *E.g.*, Richmond v. Nationwide Cassel L.P., 52 F.3d 640 (7th Cir. 1995); *Stachon*, 229 F.3d at 676; *Shapo*, 246 F. Supp. 2d at 962.

[104] *See generally* Goldsmith, *supra* note 1, at 24–28 (discussing origins of person-enterprise doctrine).

[105] *See supra* notes 20–23 and accompanying text.

[106] Organized Crime Control Act of 1970 § 901(a), 18 U.S.C. § 1962(c) (2000).

[107] *See* Goldsmith, *supra* note 1, at 24–25 & nn.138–40 (discussing early cases that created and adopted the person-enterprise distinction).

[108] *See* Yellow Bus Lines v. Drivers, Chauffeurs & Helpers Local Union 639, 883 F.2d 132, 140 (D.C. Cir. 1989); Bennett v. United States Trust Co., 770 F.2d 308, 315 (2d Cir. 1985).

[109] *See* Cox v. Adm'r U.S. Steel & Carnegie, 17 F.3d 1386, 1406 (11th Cir. 1994) ("Although we have rejected the 'non-identity' rule [i.e., the person-enterprise distinction], we have expressed concern that enterprises that are merely victims of the RICO violations perpetrated by their employees should not be held liable for the acts of their employees under *respondeat superior*."); D & S Auto Parts, Inc. v. Schwartz, 838 F.2d 964, 967 (7th Cir. 1988) ("An employer may be vicariously liable only for employee action taken within scope of employment, that is, with the intent to benefit the employer."); *cf.* KATHLEEN F. BRICKEY, CORPORATE CRIMINAL LIABILITY 83 (1984) (noting intent to benefit principal is a prerequisite of vicarious liability); WAYNE R. LAFAVE, CRIMINAL LAW 707–08 (2003) (noting exception to vicarious corporate criminal liability if employee's acts are "undertaken solely to advance [his] own interests or interests of parties other than the corporate employer").

[110] *See, e.g.*, United States v. Computer Scis. Corp., 689 F.2d 1181, 1190 (4th Cir. 1982); *see also* Goldsmith, *supra* note 1, *at* 24–25.

[111] *See* Goldsmith, *supra* note 1, at 24–28. *See also* Gasoline Sales, Inc. v. Aero Oil

tity most responsible for the violation and in the best position to pay a judgment.

To avoid this outcome, plaintiffs have employed other theories of RICO liability. For example, because courts declined to apply the person-enterprise distinction to conventional organized crime prosecutions against associated-in-fact Mafia enterprises,[112] plaintiffs began to allege that defendant business entities comprised an associated-in-fact enterprise analogous to an organized crime family.[113] This litigation strategy enjoyed some initial success,[114] but succumbed when courts began to insist that the components of the associated-in-fact business enterprise must be alleged with particularity.[115]

Plaintiffs have also turned to RICO section 1962(a), which prohibits investing racketeering proceeds in an enterprise.[116] This provision held

---

Co., 39 F.3d 70, 72–74 (3d Cir. 1994) (affirming dismissal of RICO claim despite accepting that the officers of the defendant corporation engaged in "a widespread fraudulent scheme"); *cf.* Bd. of County Comm'rs v. Liberty Group, 965 F.2d 879, 881, 884–85, 887 (10th Cir. 1992) (applying the person-enterprise distinction to remand for new trial determination, even though jury below found racketeering violations by the corporate defendant).

[112] *See Yellow Bus Lines*, 883 F.2d at 140 (noting an exception to the person-enterprise distinction "for the institution that functions as both 'perpetrator' and 'victim'"); Goldsmith, *supra* note 1, at 34. As applied to an alleged association-in-fact enterprise, the person-enterprise distinction "cannot survive scrutiny . . . because it would preclude the quintessential organized crime prosecution in which a mobster is prosecuted for conducting the affairs of a Mafia family of which he is a member." II ARTHUR F. MATTHEWS ET AL., CIVIL RICO LITIGATION § 6.03[B] at 6-43 (2d ed. 1992).

[113] *See, e.g.*, Commercial Cleaning Servs., L.L.C. v. Colin Serv. Sys., Inc., 271 F.3d 374, 379 (2d Cir. 2001) (alleging an associated-in-fact enterprise consisting of "employment placement services, labor contractors, newspapers . . . and 'various immigrant networks that assist fellow illegal immigrants in obtaining employment, housing and illegal work permits'"); N.Y. Auto. Ins. Plan v. All Purpose Agency & Brokerage, Inc., No. 97 Civ. 3164 (KTD), 1998 U.S. Dist. LEXIS 15645, at *12–13 (S.D.N.Y. Oct. 6, 1998) (alleging an orchestrated scheme to obtain fraudulently reduced automobile insurance premium rates).

[114] *See, e.g.*, Atlas Pile Driving Co. v. DiCon Fin. Co. 886 F.2d 986, 995 (8th Cir. 1989) (sustaining civil application due to concern for preserving associated-in-fact enterprise theory to prosecute criminal gangs); *see also* Goldsmith, *supra* note 1, at 34. In criminal RICO cases, the government has successfully alleged an associated-in-fact enterprise consisting of multiple corporations. *See* United States v. Perholtz, 842 F.2d 343, 353 (D.C. Cir. 1988).

[115] *See, e.g.*, Chang v. Chen, 80 F.3d 1293, 1299–1301 (9th Cir. 1996); Montesano v. Seafirst Commercial Corp., 818 F.2d 423, 427 (5th Cir. 1987) ("[P]laintiffs must plead specific facts, not mere conclusory allegations, which establish the [associated-in-fact] enterprise."); Eva v. Midwest Nat'l Mortgage Banc, Inc., 143 F. Supp. 2d 862, 875 (N.D. Ohio 2001) (requiring plaintiff to allege "facts which suggest that the behavior of the defendants is coordinated, such that they function as a continuing unit" and "some minimal level of organizational structure between the entities involved"); Flanagan v. Polites, No. 86 C 4944, 1989 U.S. Dist. LEXIS 12954, at *23–*24 (N.D. Ill. Oct. 27, 1989) ("Prior allegations in the complaint as to the alleged fraudulent activities of these individual defendants, even if the defendants at times acted in concert, do not establish the existence of an association in fact with a definite structure.") (citing H.G. Gallimore, Inc. v. Abdula, 652 F. Supp. 437, 445 (N.D. Ill. 1987)).

[116] *See, e.g.*, Scott v. Boos, 215 F.3d 940, 942 (9th Cir. 2000).

potential because some courts had ruled that its terms did not contemplate a person-enterprise distinction.[117] Thus, a corporate perpetrator may be accused of investing racketeering proceeds in its own enterprise operations. Courts have thwarted whatever promise this provision offered, however, by reading a new element into section 1962(a) RICO claims. As this provision outlaws investing racketeering proceeds in an enterprise, almost all courts reasoned that civil RICO required plaintiffs to establish injury from a defendant's *investment* of racketeering proceeds in the enterprise.[118] Despite the Supreme Court's decision in *Sedima* rejecting a comparable injury requirement for racketeering claims,[119] these decisions held that complaints alleging injury stemming solely from the pattern of racketeering activity itself did not make out a 1962(a) claim.[120] Because most victims incur injuries not from a defendant's *investment* of racketeering proceeds in an enterprise but from the racketeering activity itself,[121] the "investment injury" rule often foreclosed the final option for overcoming the person-enterprise doctrine. Cases in which a racketeering victim has successfully established an investment injury are virtually nonexistent.[122]

Remarkably, this line of authority contravenes Supreme Court and statutory directives that RICO must be liberally construed.[123] Eventually, however, even the Supreme Court succumbed to the revisionist tide of the lower courts.[124] In *Reves v. Ernst & Young*,[125] the Supreme Court considered whether RICO section 1962(c) applies only to persons involved in the operation or management of the enterprise. In pertinent part, this pro-

---

[117] *See* James D. Higgason, Jr., Note, *Enterprise Liability in Private Civil RICO Actions*, 45 WASH. & LEE L. REV. 1447, 1458–64 (1988); *see also* Haroco, Inc. v. Am. Nat'l Bank & Trust Co., 747 F.2d 384, 402 (7th Cir. 1984) (noting that even though the defendant "may not be held liable under section 1962(c) . . . [it] might be held liable under section 1962(a) if it received the proceeds from the alleged racketeering activities").

[118] *See* Goldsmith, *supra* note 1, at 29–31 (collecting cases).

[119] Sedima, S.P.R.L. v. Imrex Co., 473 U.S. 479, 495 (1985) (rejecting racketeering injury limitation).

[120] Ouaknine v. MacFarlane, 897 F.2d 75, 82–83 (2d Cir. 1990).

[121] *See id.*; Goldsmith, *supra* note 1, at 29–31 (collecting cases).

[122] For a rare example of a court declining to reject a 1962(a) claim based on the absence of an investment injury, see St. Paul Insurance Co. v. Williamson, 224 F.3d 425, 444 (5th Cir. 2000).

[123] *See supra* note 51; *see also Sedima*, 473 U.S. at 497–98.

[124] Although the Supreme Court has never adopted the person-enterprise distinction, its decision in *Cedrick Kushner Promotions, Ltd. v. King* assumed the validity of the distinction. 533 U.S. 158, 161–63 (2001). Remarkably, the Solicitor General endorsed the distinction as "legally sound and workable." *Id.* at 162 (quoting Brief for United States as Amicus Curiae at 11). The Supreme Court uncritically accepted this assessment, noting that "12 Courts of Appeals have interpreted the statute as embodying some such distinctness requirement without creating discernible mischief in the administration of RICO." *Id.* The Solicitor General's views may change when defense counsels inevitably argue that the distinction also applies to organized crime prosecutions, and the Supreme Court may reconsider its conclusory statement once the difficulty of applying RICO to corporate corruption becomes more apparent.

[125] 507 U.S. 170, 183 (1993).

vision made it illegal for anyone "employed by or associated with any enterprise . . . to conduct or participate, directly or indirectly, in the conduct of such enterprise's affairs through a pattern of racketeering activity."[126] Because this prohibition reaches persons conducting enterprise affairs "directly or indirectly,"[127] the statute would appear to encompass outside professionals not involved in the management or operation of the enterprise.[128]

A divided Supreme Court, however, interpreted RICO narrowly in holding that an outside accounting firm did not conduct enterprise affairs when its fraudulent audits facilitated a company fraud.[129] Justice Blackmun's majority opinion reasoned that the word "conduct" implies "some degree of direction" and that the word "participate" anticipates "some part in that direction."[130] Over Justice Souter's dissent, which argued that Justice Blackmun's dictionary-based analysis ignored RICO's liberal construction clause and statutory context,[131] the Court sustained summary judgment for the defense and established an "operation or management" test for section 1962(c) liability.[132] Although the Court rejected the stricter view that section 1962(c) requires "*significant control* over or within an enterprise,"[133] *The Washington Post* reported that, under this decision, "[p]eople who lose money in thrifts and other businesses that go belly up because of wrongdoing can no longer use [RICO] to sue lawyers, accountants or other advisers who played key roles in the enterprise."[134]

* * *

---

[126] Organized Crime Control Act of 1970, § 901(a), 18 U.S.C. § 1962(c) (2000).

[127] *Id.*

[128] *See, e.g.*, Bank of Am. Nat'l Trust & Sav. Assn. v. Touche Ross & Co., 782 F.2d 966, 970 (11th Cir. 1986), *overruled* by *Reves*, 507 U.S. at 177–79.

[129] Reves v. Ernst & Young, 507 U.S. 170, 185–86 (1993).

[130] *Id.* at 178–79.

[131] *Id.* at 188–89 (Souter, J., dissenting).

[132] *Id.* at 185.

[133] *Id.* at 179 n.4 (quoting Yellow Bus Lines Inc. v. Drivers, Chauffeurs & Helpers Local Union 639, 913 F.2d 948, 954 (D.C. Cir. 1990) (emphasis added by court)).

[134] Joan Biskupic, *Supreme Court Limits Use of Racketeering Law*, WASH. POST, Mar. 4, 1993, at A1. *See also, e.g.*, Azrielli v. Cohen Law Offices, 21 F.3d 512, 521–22 (2d Cir. 1994) (applying *Reves* to affirm dismissal of a RICO claim against an attorney because he did not meet the "operation or management" test, but leaving other claims against him intact); Manley v. Stark & Stark, P.C., Civil No. 97-524 (AET), 1999 U.S. Dist. LEXIS 22082, at *29–30 (D.N.J. Aug. 10, 1999) (quoting language from Univ. of Md. v. Peat, Marwick, Main & Co., 996 F.2d 1534 (3d Cir. 1993), that under the *Reves* test as applied in the Third Circuit, "not even action involving some degree of decisionmaking constitutes participation in the affairs of an enterprise" to preclude a RICO claim against defendant attorneys); Bowdoin Constr. Corp. v. R.I. Hosp. Trust Nat'l Bank, 869 F. Supp. 1004, 1009–10 (D. Mass. 1994) (dismissing a RICO claim against a lawyer for a developer because he had only provided legal advice and did not meet the operation or management test); Morin v. Turpin, 832 F. Supp. 93, 98 (S.D.N.Y. 1993) (applying *Reves* to preclude a RICO suit against attorneys).

Taken together, these decisions potentially shielded perpetrator entities and outside professionals from RICO liability. Yet various business groups have long sought even more restrictive legislative reform. Led by the accountants' professional association, these well-financed groups mounted a massive lobbying effort maintaining that RICO fostered legalized extortion.[135] Interestingly, despite a spate of stock scandals, the chairman of the Securities and Exchange Commission (SEC) endorsed these efforts.[136] He testified before Congress that RICO's treble damage provision promoted litigation abuse, that the SEC needed no help from private attorneys general, and that RICO threatened to override pre-existing remedies under the securities laws.[137] Congress embraced this view and enacted legislation precluding securities-based RICO claims.[138]

In part, these efforts to restrict RICO stemmed from concerns that the statute's breadth fostered frivolous filings.[139] RICO opponents, however, have exaggerated the problem of abusive RICO litigation.[140] Moreover, to the degree that plaintiffs' counsel have employed RICO in bad

---

[135] *See RICO Oversight Hearings*, *supra* note 34, at 243–50 (statement of Ray J. Groves, Chairman, American Institute of Certified Public Accountants); *see also* Philip A. Lacovara & Geoffrey F. Aronow, *The Legal Shakedown of Legitimate Business People: The Runaway Provisions of Private Civil RICO*, 21 NEW ENG. L. REV. 1, 1 n.** (1985) ("This article is based on the submissions that the authors have made to Congress as counsel to the American Institute of Certified Public Accountants in seeking a congressional amendment of the civil RICO statute."). The accountants argued that RICO's structure and breadth fostered excessive and abusive litigation against respected legitimate businesses. *Id.* at 4–6, 9–23; *see also* Editorial, *The RICO 'Reformers' Are Back Again*, N.Y. TIMES, Sept. 24, 1990, at A18; John Conyers Jr., Editorial, *Don't Water Down the Antifraud Law*, N.Y. TIMES, Dec. 27, 1987, § 4, at 13; Howard Kurtz, *Businesses Fight RICO Antifraud Law*, WASH. POST, Oct. 16, 1985, at A21.

[136] *See Common Sense Legal Reform Act: Hearings on H.R. 10 Before the Subcomm. on Telecomms. and Fin. of the House Comm. on Commerce*, 104th Cong. 191–95 (1995) (statement of Arthur Levitt, Chairman, Securities and Exchange Commission).

[137] *See id.*

[138] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, § 107, 109 Stat. 737, 758 (1995) (codified at 18 U.S.C. § 1964(c) (2000)).

[139] *See* Norm Abrams, *A New Proposal For Limiting Private Civil RICO*, 37 UCLA L. REV. 1, 12 (1989); Ethan M. Posner, *Clarifying a "Pattern" of Confusion: A Multi-Factor Approach to Civil RICO's Pattern Requirement*, 86 MICH. L. REV. 1745, 1769 (1988).

[140] *See* G. Robert Blakey & Thomas A. Perry, *An Analysis of the Myths that Bolster Efforts to Rewrite RICO and Various Proposals for Reform: "Mother of God—Is This the End of RICO?,"* 43 VAND. L. REV. 851, 869–80 (1990); Michael Goldsmith & Penrod Keith, *Civil RICO Abuse: The Allegations in Context*, 1986 BYU L. Rev. 55, 66–84; Michael Goldsmith & Mark Jay Linderman, *Civil RICO Reform: The Gatekeeper Concept*, 43 VAND. L. REV. 735, 745–62 (1990); Bruce Haber, *Congress Responds to Sedima: Is There a Contract Out on Civil RICO?*, 19 Loy. L.A. L. REV. 851, 867–70 (1986); *see also RICO Oversight Hearings*, *supra* note 34, at 141 (statement of Stephen S. Trott). Admittedly, plaintiffs' counsel would naturally attempt to convert ordinarily commercial disputes into racketeering claims and RICO opponents understandably disliked being named as defendants. However, courts routinely rejected frivolous claims pre-trial. *See* Goldsmith & Linderman, *supra*, at 759–60. Moreover, when a plaintiff states a proper claim, no reason exists to give business defendants immunity from having to defend at trial. Ironically, some of the most prominent businesses supposedly unfairly tarnished as RICO defendants were subsequently exposed as corrupt enterprises. *See supra* note 9.

faith, another remedy exists short of restrictive reform. Rule 11 of the Federal Rules of Civil Procedure authorizes sanctions for abusive litigation, and courts have, on occasion, imposed severe penalties in cases of RICO abuse.[141] Rather than rely on Rule 11 and seek reform tailored to legitimate concerns of undue breadth, however, RICO opponents have sought to eviscerate its civil remedy.[142]

Ultimately, judicial and congressional revisions combined to weaken RICO and create a climate conducive to corporate corruption. Unless corrected, these restrictions will impede recovery by victims of both recent and future corporate criminality.

III. RICO RESTRICTIONS: IMPLICATIONS FOR RACKETEERING VICTIMS

The Enron/Arthur Andersen case exemplifies the difficulties victimized investors and employees likely will encounter when they attempt recovery under civil RICO. The corporate debacles of 2001–02 share numerous common characteristics: overstated profits, understated debts, conflicts of interest within senior management, outside accountants disinclined to ask the right questions, and an overall reluctance to disclose painful answers.[143] Prior to the recent scandals, companies such as Enron enjoyed elite status.[144] Enron ranked fifth among the *Fortune* 500 list of largest U.S. corporations.[145] *Newsweek* characterized Enron as "one of America's most admired companies, and a perennial favorite on 'best places to work' lists."[146] Wall Street analysts rated "its stock and bonds as the greatest thing since money was invented."[147]

---

[141] Goldsmith & Keith, *supra* note 140, at 94–95 (citing examples of court-imposed sanctions for bad faith RICO litigation).

[142] *See* Michael Goldsmith, *Civil RICO Reform: The Basis for Compromise*, 71 MINN. L. REV. 827, 848–58, 884–911 (1987) (critiquing then-pending "reform" proposals and including appendix of such bills).

[143] Arthur Levitt, former Chairman of the Securities Exchange Commission, described Enron's collapse as occurring against a backdrop of "obsessive zeal by too many American companies to project greater earnings from year to year." *The Fall of Enron: How Could It Have Happened?: Hearing Before the Senate Comm. on Governmental Affairs*, 107th Cong. 26 (2002). Levitt went on to say that in attempting to project greater earnings, companies "bend the rules . . . tweak the numbers, and let obvious and important discrepancies slide . . . analysts more often overlook dubious accounting practices . . . auditors are more occupied with selling other services and making clients happy than detecting potential problems . . . and . . . directors are more concerned about not offending management than with protecting shareholders." *Id.* at 26–27.

[144] The *Wall Street Journal*'s headlines documented the rapid and seemingly steady growth of Enron. *See, e.g.*, *Enron Corp.'s Profit Rises 22% as Pipelines Show Steady Income*, WALL ST. J., Oct. 11, 1996, at C17; Rebecca Smith, *Enron's Net Increased 30% in 2nd Quarter*, WALL ST. J., Jul. 25, 2000, at A3; Rebecca Smith, *Enron's Net Soars, Despite Telecom Loss, Gaining 40% Amid Strong Energy Units*, WALL ST. J., Jul. 13, 2001, at A2.

[145] *FORTUNE Five-Hundred Largest U.S. Corporations*, FORTUNE, Apr. 15, 2002, at F1.

[146] Allan Sloan, *Who Killed Enron?*, NEWSWEEK, Jan. 21, 2002, at 18.

[147] *Id.* at 18. Enron's innovative and ground-breaking effect on the market was likened to Elvis's effect on entertainment, and its awe-inspiring effect on analysts was likened to

In one year, however, Enron declined more than ninety-nine percent in value[148] and terminated approximately 20,000 employees. Enron employees also lost most of their retirement savings because company policy kept them from selling Enron stock from their 401(k) plans.[149] This policy, however, did not bind senior managers, who continued falsely touting Enron stock even as they sold most of their shares at huge profits.[150] During the period in which the company sank into bankruptcy, chairman Ken Lay made $250 million in stock-option profits.[151] Investments in off-the-books partnerships, which allowed Enron to overstate its net worth by $1.2 billion, generated multi-million dollar profits for other senior officials.[152] These investments violated company conflict-of-interest rules, and Enron's failure to disclose the partnerships violated generally accepted accounting principles.[153] This occurred because the Arthur Andersen accounting firm, operating under its own conflict of interest stemming from its lucrative contracts with Enron, failed to function as an independent auditor.[154] Federal prosecutors eventually convicted Arthur Andersen for obstruction of justice for destroying potentially incriminating documents.[155]

The financial disaster left Enron victims with few options. Some pragmatic employees, apparently having literally lost the shirts off their backs, posed for Playboy.[156] Others began a Web site, "laydoff.com," which

---

Jesus Christ's effect on his followers. *See* Brian O'Reilly, *The Power Merchant*, FORTUNE, Apr. 17, 2000, at 148.

[148] Kenneth N. Gilpin, *Enron's Collapse: The Investors; Plenty of Pain to Go Around for Small Investors, Funds, Workers, and Creditors*, N.Y. TIMES, Dec. 4, 2001, at C8. At its peak, Enron had a market value of about $68 billion, but as of Dec. 3, 2001, the market value was only $344 million.

[149] *Retirement Insecurity: 401(k) Crisis at Enron: Hearing before Senate Comm. On Governmental Affairs*, 107th Cong. (2002) (statement of Deborah G. Perrotta, Frmr. Sr. Administrative Assistant at Enron).

[150] Allan Sloan, *Is the Boss Dumping Stock?*, NEWSWEEK, Mar. 11, 2002, at 30.

[151] Allan Sloan, *Enron and Fuzzy Math*, NEWSWEEK, Feb. 4, 2002, at 22.

[152] *Enron's Fastow Charged In 78-Count Federal Indictment*, CHI. TRIB., Nov. 1, 2002, at 2N; Julie Mason, *3 Execs Contradict Skilling Testimony; Ex-Chief OK'd Deals, They Say*, HOUS. CHRON., Mar. 1, 2002, § A, at 1; *Enron: A Simple Question Of Right And Wrong*, USA TODAY, Jan. 22, 2002, at 12A; Editorial, *Clinton-Gore: See No Evil*, WASH. TIMES, July 22, 2002, at A18. On April 17, 2001, Enron released its first-quarter results to investors and reported $425 million in earnings. Yet since mid-1999, Enron had engaged in several partnerships that it was using as a tool to make the company seem more profitable than it really was. In the very same quarter that Enron reported $425 million in earnings, its partners kept almost $504 million in debt off its books. Floyd Norris, *Fun-House Accounting: The Distorted Numbers at Enron*, N.Y. TIMES, Dec, 14, 2001, at C1.

[153] Norris, *supra* note 152, at C1.

[154] Editorial, *Accountants on Trial*, N.Y. TIMES, June 19, 2002, at A22; Editorial, *Insider Outsiders*, WASH. POST, May 22, 2002, at A36.

[155] Kurt Eichenwald, *Andersen Guilty in Effort to Block Inquiry on Enron*, N.Y. TIMES, June 16, 2002, at A1.

[156] *See* Peter Carlson, *Playboy's 'Women of Enron,' Cashing in on the Bare Market*, WASH. POST, July 16, 2002, at C08; Thomas S. Mulligan, *They Have Nothing to Hide; Ten 'Women of Enron' Find a Temp Job Posing for a Pictorial in Playboy*, L.A. TIMES, June 27, 2002, § 5 at 1.

sold T-shirts emblazoned with the phrase, "I got layd by Enron."[157] Unfortunately, however, given the extent to which Congress and the courts have "reformed" RICO, its treble damages provision offers many victims no greater potential for restitution. This is especially troublesome as studies based on the antitrust model demonstrate that victims often require a treble damages remedy to recover their *actual* damages.[158] Moreover, treble damages serve a deterrent function which is vital to restoring corporate integrity.[159] Thus, absent remedial reform, RICO will neither compensate victims adequately nor deter future corporate criminality. To be effective, such reform must encompass both legislative and judicial solutions.

IV. Remedial RICO Reform

A. Legislative Solutions

Initially, the Private Securities Litigation Reform Act of 1995[160] stands as the most formidable obstacle to Enron victims and others similarly situated (both present and future). This legislation exempted securities-based frauds from RICO section 1964(c), which provides treble damages plus counsel fees for racketeering victims.[161] Accordingly, although the consequences of Arthur Andersen's actions dwarfed Meyer Lansky's accounting crimes when he cooked the books for the mob,[162] civil RICO bars relief for the victims of securities fraud.[163]

Fortunately, this aspect of RICO reform is easy to remedy (if the political will exists). Congress should repeal RICO's securities exemption. Further, this repeal should operate retroactively so that injured investors could sue for relief. Retroactive application would be fully constitutional.[164]

---

[157] Although the Web site www.laydoff.com no longer exists, its content has been preserved on sites such as http://csf.colorado.edu/forums/pkt/2002I/msg00557.html.

[158] *See* A. Douglas Melamed, *Damages, Deterrence, and Antitrust—A Comment on Cooter*, 60 Law & Contemp. Probs., 93, 115–21 (1997) (citing National Cooperative Production Amendments of 1993, Pub. L. No. 103-42, 107 Stat. 117 (1993); National Cooperative Resarch Act of 1984, Pub. L. No. 98-462, 99 Stat. 1815 (1984)).

[159] Michael Goldsmith, *Civil RICO Reform: The Basis for Compromise*, 71 Minn. L. Rev. 827, 847 n.86 (1987) (citing Judith A. Morse, *Treble Damages Under RICO: Characterization and Computation*, 61 Notre Dame L. Rev. 526, 533 n.38 (1986)).

[160] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, 109 Stat. 737.

[161] 18 U.S.C. § 1964(c) (2002).

[162] *See* Hank Messick, Lansky 2–3, 25–26, 198–203 (1971).

[163] At least from a civil standpoint, this suggests that mob accountants would be well advised to concentrate on securities work rather than fudge the numbers for wholly illicit organizations.

[164] *See* Galvan v. Press, 347 U.S. 522, 531 (1954) (affirming long-established principle that the *Ex Post Facto* Clause of the Constitution generally applies to criminal prosecutions); *In re* Extradition of McMullen, 769 F. Supp. 1278, 1292 (S.D.N.Y. 1991).

The right to sue, however, matters little if judicially imposed impediments to recovery remain in place. Today, such impediments may block Enron victims from using RICO to attain relief. For example, the person-enterprise distinction precludes Enron victims from alleging that Enron conducted its own affairs through a pattern of racketeering activity in violation of section 1962(c). Moreover, although courts have justified the person-enterprise doctrine by noting that it does not apply to RICO section 1962(a),[165] the investment-injury rule forecloses recovery under that provision. Finally, even if the person-enterprise doctrine did not exist, the *Reves* decision renders problematic efforts to impose section 1962(c) liability upon outside accountants, such as Arthur Andersen.

Some of these obstacles require legislative remedies. Others call for judicial reform consistent with recent Supreme Court precedent, modern pleading practice under the Federal Rules of Civil Procedure, and RICO's statutory text.

Therefore, in addition to repealing the securities exemption to its civil remedy, RICO requires three legislative reforms. First, Congress should eliminate the judicially imposed person-enterprise doctrine[166] by amending RICO to impose liability on perpetrator entities that conduct their own affairs through a pattern of racketeering activity. Otherwise, for example, neither Enron nor Arthur Andersen could be accused, civilly or criminally, of conducting its own business through a pattern of racketeering activity. Without legislative action, institutional entities may face RICO liability only if they act in concert with others, thereby creating an associated-in-fact enterprise.[167] As the person-enterprise rule reflects judicial concerns that corporations should not face vicarious RICO liability for crimes committed by low-level employees,[168] any such reform should limit liability to perpetrator enterprises. Thus, companies that merely served as passive instruments of racketeering activity or were themselves victimized by low-level employees would remain immune from civil RICO liability.

Second, Congress should overturn the *Reves* decision[169] by explicitly imposing liability upon outsiders who facilitate or promote the enterprise's pattern of racketeering activity. Even prior to Enron and other recent corporate scandals, corrupt accountants and attorneys played pivotal roles in many major financial frauds.[170] Thus, outside professionals should not enjoy automatic RICO immunity.

---

[165] *See, e.g.*, New Beckley Mining Corp. v. Int'l Union, United Mine Workers, 18 F.3d 1161, 1163 (4th Cir. 1994); B.F. Hirsch, Inc. v. Enright Refining Co., 617 F. Supp. 49, 51 (D.N.J. 1985).

[166] *See supra* notes 104–111 and accompanying text.

[167] *See supra* notes 112–115 and accompanying text. Even then, complex rules make it extremely difficult for plaintiffs to plead and prove the existence of such an enterprise.

[168] *See supra* notes 105–111 and accompanying text.

[169] 507 U.S. 170 (1993).

[170] *See, e.g.*, G. Robert Blakey & Kevin P. Roddy, *Reflections on* Reves v. Ernst &

Third, Congress should repeal the judicially imposed investment-injury doctrine. The doctrine contravenes Supreme Court precedent and frustrates legislative policy. Certainly from the racketeering victim's standpoint, its economic losses are not diminished because the violator chose not to invest fraudulently acquired proceeds in a manner that caused further harm.

The appendix contains a unified proposed text to implement these reforms. Taken together, these measures would provide recovery for fraud victims and deter future corporate criminality. Legislative reform, however, will accomplish little if courts continue to hold RICO complaints to a heightened pleading standard. This trend conflicts with three Supreme Court decisions aimed at restoring judicial deference to civil pleadings. Thus far, however, few courts have even considered this trilogy's impact on RICO litigation. Properly understood, the trilogy eliminates judicially imposed heightened pleading standards for civil RICO complaints and provides a judicial framework more conducive to remedial enforcement efforts.

B. The Supreme Court Trilogy Rejects Heightened Pleading Standards

Three relatively recent decisions, *Leatherman*, *Crawford-El*, and *Sorema N.A.*, represent the Court's renewed insistence that Rule 8(a) provides a low threshold for surviving motions to dismiss. From its inception, Rule 8(a) encountered judicial resistance.[171] Its liberal notice pleading provisions proved too revolutionary for many federal courts, which continued dismissing complaints for lack of factual detail.[172] In 1957, this practice finally prompted the Supreme Court to issue a decision emphasizing that Rule 8(a) "meant what it said":[173]

> The Federal Rules of Civil Procedure do not require a claimant to set out in detail the facts upon which he bases his claim. To the contrary, all the Rules require is a "short and plain statement of the claim" that will give the defendant fair notice of what the plaintiff's claim is and the grounds upon which it rests.[174]

This directive notwithstanding, some courts developed heightened pleading standards for certain judicially disfavored claims. For example,

---

Young: *Its Meaning and Impact on Substantive, Accessory, Aiding Abetting and Conspiracy Liability Under RICO*, 33 AM. CRIM. L. REV. 1345, 1354–62 (1996) (detailed analysis of outside professional culpability in multi-billion dollar savings and loan scandals of 1980s).

[171] *See* Marcus, *supra* note 50, at 1750.

[172] *See id.* at 434.

[173] Leatherman v. Tarrant County Narcotics Intelligence and Coordination Unit, 507 U.S. 163, 168 (1993) (discussing 1957 decision).

[174] Conley v. Gibson, 355 U.S. 41, 47 (1957).

the Fifth Circuit required civil rights plaintiffs, in cases involving the defense of sovereign immunity, to "state with factual detail and particularity the basis for the claim which necessarily includes why the defendant-official cannot successfully maintain the defense of immunity."[175] Eliminating this practice required the Supreme Court to issue a trilogy of decisions reinforcing its commitment to notice pleading principles.

Leatherman v. Tarrant County Narcotics Intelligence and Coordination Unit,[176] the first case in the trilogy, concerned the Fifth Circuit's imposition of a heightened pleading requirement upon civil rights complaints. In rejecting this practice, the Supreme Court reinforced the traditional standard for reviewing motions to dismiss. Foremost, judges deciding such motions "must accept as true all the factual allegations in the complaint."[177] Chief Justice Rehnquist's opinion for a unanimous court observed "that it is impossible to square the 'heightened pleading standard' . . . with the liberal system of 'notice pleading' set up by the Federal Rules."[178] The Court recognized that Rule 9(b) requires particularized pleading in certain actions, but this provision does not encompass civil rights actions: *Expressio unius est exclusio alterius*.[179]

Therefore, even if valid policies supported a heightened pleading standard for civil rights actions, the Court stressed "that is a result which must be obtained by the process of amending the Federal Rules, and not by judicial interpretation."[180] Rather than affirming a heightened pleading standard, *Leatherman* directed that "federal courts and litigants must rely upon summary judgment and control of discovery to weed out unmeritorious claims."[181]

This theme subsequently provided the basis for the Supreme Court's decision in *Crawford-El v. Britton*,[182] the second part of the trilogy. In rejecting the D.C. Circuit's requirement that civil rights plaintiffs must adduce clear and convincing evidence of a defendant's state of mind, the Supreme Court characterized summary judgment as the "ultimate screen to weed out truly insubstantial lawsuits prior to trial."[183] Justice Stevens's majority opinion acknowledged that serious and widespread problems attendant to civil rights litigation had prompted the D.C. Circuit's decision to impose a higher burden of proof in such cases.[184] But his opinion rejected judicial legislation as a vehicle for addressing these concerns:

---

[175] Judge v. City of Lowell, 160 F.3d 67, 73 (1st Cir. 1998); Elliott v. Perez, 751 F.2d 1472, 1473 (5th Cir. 1985).

[176] 507 U.S. 163, 167 (1993).

[177] *Id.* at 164.

[178] *Id.* at 168.

[179] The inclusion of specific things implies the exclusion of those not mentioned. *Id.*

[180] *Id.*

[181] *Id.*

[182] 523 U.S. 574 (2002).

[183] *Id.* at 600.

[184] *Id.* at 584–85, 590, 593 n.14, 596.

"Neither the text of [the civil rights law] or any other federal statute, nor the Federal Rules of Civil Procedure provide any support for imposing the clear and convincing burden of proof on plaintiffs."[185] To the degree that civil rights suits create special problems for federal courts, "questions regarding pleading . . . are most effectively resolved either by the rulemaking process or the legislative process."[186] In short, judges may not develop their own procedural or evidentiary requirements that override or constrain federal legislation.

Most recently, this perspective proved decisive in *Swierkiewicz v. Sorema N.A.*,[187] the last case in the trilogy, an employment discrimination case in which the Supreme Court rejected the Second Circuit's attempt to impose a heightened pleading standard upon employment discrimination plaintiffs. Responding to arguments that conclusory discrimination claims unduly burden courts and employers, the Supreme Court stated: "[w]hatever the practical merits of this argument, the Federal Rules do not contain a heightened pleading standard for employment discrimination suits."[188] Citing *Leatherman*, Justice Thomas' opinion for a unanimous court reminded the judiciary that a particularized pleading requirement may be imposed only by amending the Federal Rules of Civil Procedure.[189]

Significantly, the *Sorema* Court did not ground its decision solely on separation-of-powers considerations.[190] Justice Thomas also returned to basic principles of modern federal civil procedure. He observed,

> The liberal notice pleading of Rule 8(a) is the starting point of a simplified pleading system which was adopted to focus litigation on the merits of a claim. . . . The Federal Rules reject the approach that pleading is a game of skill in which one misstep by counsel may be decisive to the outcome and accept the principle that the purpose of pleading is to facilitate a proper decision on the merits.[191]

Finally, Justice Thomas emphasized that the Second Circuit's heightened pleading requirement was akin to an evidentiary burden that has no place at the pleading stage.[192] Under Rule 8(a), plaintiffs need not establish any likelihood of success on the merits. On the contrary, "it may appear . . .

---

[185] *Id.* at 594.
[186] *Id.* at 595.
[187] 534 U.S. 506, 511 (2002).
[188] *Id.* at 514.
[189] *Id.*
[190] *Id.*
[191] *Id.*
[192] *Id.* at 510–11.

that a recovery is very remote and unlikely, [but] that is not the test . . . ."[193]

Indeed, the correct test has not changed since the adoption of the Federal Rules of Civil Procedure; it is the same test that federal courts often ignored when the rules first took effect and have since ignored in most civil RICO litigation. A complaint may be dismissed only if it demonstrates "beyond doubt that the plaintiff can prove no set of facts in support of his claim [that] would entitle him to relief."[194] If *Leatherman*, *Crawford-El*, and *Sorema N.A.* mean what they say, the trilogy should finally restore the liberal notice-pleading model envisioned under the Federal Rules of Civil Procedure. Nothing about these cases suggests their reasoning applies only to civil rights actions. The Supreme Court could not have made plainer that, except as provided in Rule 9(b), all complaints must be gauged by notice pleading standards.[195] Judicial compliance with this mandate would revitalize civil RICO litigation, as courts could no longer freely substitute their own views concerning the statute's proper scope for those of Congress.

Thus far, however, *Leatherman* and its progeny have had almost no effect on civil RICO rulings. Most decisions dismissing RICO complaints make no reference to these cases,[196] and only a few have even noted the issue of whether *Leatherman* precludes heightened pleading standards in civil RICO cases.[197] More often, courts have indicated, without explanation, that a heightened pleading standard still governs RICO complaints.[198] These decisions either ignore *Leatherman* entirely or attempt to reconcile *Leatherman* by looking to the district court's authority to require plaintiffs to submit a "RICO Case Statement" detailing every aspect of each RICO claim.[199]

---

[193] *Id.* at 515 (quoting Scheuer v. Rhodes, 416 U.S., 232, 236 (1974)).

[194] 2 JAMES WM. MOORE ET AL., MOORE'S FEDERAL PRACTICE § 12.34[1][a], at 12–56 (3d ed. 1998) (quoting Conley v. Gibson, 355 U.S. 41, 45–46 (1957)).

[195] *See supra* notes 176–193 and accompanying text.

[196] Jeffrey W. Stempel, *Contracting Access to the Courts: Myth or Reality? Boon or Bane?*, 40 ARIZ. L. REV. 965, 990 n.104 (1998) (finding post-*Leatherman* "particularized pleadings claims involving RICO, antitrust, prisoners' claims, constitutional rights, and civil rights actions generally. In other words, courts are still playing favorites.").

[197] *See, e.g.*, Jacobs v. Port Neches Police Dep't, No. 1:94-CV-767, 1996 U.S. Dist. LEXIS 911, at *41–*43 (E.D. Tex. June 26, 1996); Ellis v. Welch, No. 92 C 4633 , 1994 U.S. Dist. LEXIS 2692, at *16–*18 (N.D. Ill. Mar. 8, 1994).

[198] *See, e.g.*, Warden v. McLelland, 288 F.3d 105, 114 (3d Cir. 2002); Slaney v. Int'l Amateur Athletic Fed'n, 244 F.3d 580, 598 (7th Cir. 2001), *cert. denied*, 534 U.S. 828 (2001); Langadinos v. Am. Airlines, Inc. 199 F. 3d 68, 73 (1st Cir. 2000).

[199] *See* Sanville v. Bank of Am. Nat'l Trust & Save As'n, 18 Fed. Appx. 500, 501 (9th Cir. 2001); Efron v. Embassy Suites (P.R.), Inc., 223 F.3d 12, 20 (1st Cir. 2000); Bozeman v. Rochester Tel. Corp., No. 99-7252, 2000 U.S. App. LEXIS 812, 2 (2d Cir. Jan. 19, 2000); Atkins v. Hibernia Corp., 182 F.3d 320, 326 (5th Cir. 1999). For a sample RICO Case Statement, see Michael Goldsmith & Penrod Keith, *supra* note 140, at 105–07 (containing the first RICO Case Statement). Such statements are widely employed. *See, e.g.*, A. Darby Dickerson, *Curtailing Civil RICO Long Reach: Establishing New Boundaries for Venue and Personal Jurisdiction under 18 U.S.C. § 1965*, 75 NEB. L. REV. 476, 548 n.78

Indeed, cases challenging RICO Case Statements as inconsistent with *Leatherman* have generally failed.[200] For example, *Northland Insurance Co. v Shell Oil Co.* contains the most comprehensive analysis sustaining the district court's authority to order RICO Case Statements.[201] Rejecting plaintiffs' argument that RICO Case Statements violate *Leatherman* by imposing heightened standards, *Northland Insurance* ruled that such statements stem from various federal provisions authorizing district courts to implement appropriate measures to manage complex cases.[202] *Northland Insurance* maintained that, among other provisions of federal law,[203] Federal Rule of Civil Procedure 12(e), which authorizes district courts to order plaintiffs to provide more definite statements of their claims, provides authority for RICO Case Statement orders. The court also reasoned that *Leatherman* does not apply to RICO claims because "RICO cases often involve allegations of fraud, which . . . require pleading with specificity pursuant to Federal Rule 9(b)."[204] Finally, the court found "it . . . significant that the Supreme Court relied upon the RICO Case Statement for a complete statement of the claim in *NOW v. Scheidler*[205] . . . a full year after the *Leatherman* decision."[206]

Remarkably, this analysis at least implicitly acknowledged that RICO Case Statements impose heightened pleading standards. Indeed, rather than suggest otherwise, the court attempted to justify heightened scrutiny for RICO complaints. Unfortunately, its analysis sustaining RICO Statements misconstrued *dicta* in *Scheidler*, misapplied Rule 12(e), and misconceived the origin, purpose and limits of RICO Case Statements.

The Supreme Court's passing reference in *NOW v. Scheidler* to plaintiffs' having filed a RICO Case Statement hardly constitutes authority for their continued validity. As the defendants in *Scheidler* did not challenge plaintiffs' pleadings as lacking detail,[207] the Supreme Court's observation that plaintiffs filed such a statement has no bearing on whether this judicially imposed requirement survives *Leatherman*.

The RICO Case Statement arose as a well-intentioned management tool to eliminate frivolous claims and help district courts process com-

---

(1996) (noting judicial "animosity" and citing extensive authority).

[200] *See, e.g.*, Northland Ins. Co. v. Shell Oil Co., 930 F. Supp. 1069, 1073 (D.N.J. 1996) (noting that that "approximately one dozen districts have adopted and routinely employ RICO Case Statements . . . [and that] plaintiff supplies no case which has held the RICO Case Statement Order improper.").

[201] *Id. at* 1071–76.

[202] *Id.* at 1072–74. The court stressed that "the parties do not dispute that RICO claims generally involve complex issues." *Id.* at 1074.

[203] *Id.* at 1071, 1074 (citing 28 U.S.C. § 2071 (1994); FED. R. CIV. P. 83)

[204] *Id.*

[205] 510 U.S. 247 (1994).

[206] 90 F. Supp. at 1075.

[207] *See* 510 U.S. at 254 (attack based on substantive RICO requirements; no claim of lack of detail in pleadings).

plex RICO complaints.[208] RICO Case Statements require plaintiffs to particularize every aspect of their RICO claim including, for example, details pertinent to the alleged pattern of racketeering activity, the RICO enterprise, and how the defendant(s) conducted that enterprise through the alleged pattern of racketeering activity.[209] These mandates go well beyond the particularity requirements of Rule 9(b), which run only to the elements of the underlying fraud.[210] Because this level of detail is the antithesis of "notice pleading," RICO Case Statements conflict with Rule 8 of the Federal Rules of Civil Procedure when courts treat RICO Case Statements as pleadings subject to dismissal under a more demanding standard than properly applies to Rule 12(b)(6) motions to dismiss.

Pursuant to local rules or standing orders, district courts may require plaintiffs to file RICO case statements.[211] Such statements, however, should be employed only to facilitate discovery and judicial oversight during pre-trial stages of the litigation.[212] Moreover, federal law prohibits local rules that contravene other federal statutes or rules.[213] Accordingly, RICO Case Statements may neither impose heightened pleading standards nor lead to dismissal for non-compliance. Decisions such as *Northland Insurance* ignore this mandate.

RICO Case Statements impose a level of specificity upon plaintiffs that exceed the norm required of motions "for a more definite statement" under Rule 12(e). Courts may grant Rule 12(e) motions only "[i]f a pleading . . . is so vague or ambiguous that a party cannot reasonably be required to frame a responsive pleading."[214] These motions, however, are disfavored,[215] and may not substitute for requests for a detailed "bill of particulars,"—especially since a "1948 amendment [to the Federal Rules of Civil Procedure] eliminated the bill of particulars from federal practice

---

[208] *See* FED. R. CIV. P. 16(c)(12); Advisory Committee Note, 1993 Amendment ("regarding the authority of the court to make appropriate order [sic] designed either to facilitate settlement or to provide for an efficient and economical trial."), *quoted in Northland*, 930 F. Supp. at 1074 (also noting that "the federal rule providing the strongest authority for RICO Case Statements is Rule 16."); *see also* Bromm v. Premier Capital Investment Corp., 1996 U.S. District Lexis 22133 *5 (D. Neb.)

[209] *See supra* notes 200–206 and accompanying text.

[210] *See supra* notes 97–100 and accompanying text. Furthermore, not all RICO claims are based on fraud. The statute also contains numerous non-fraud predicates. 18 U.S.C. § 1961(1) (2000).

[211] *See* 930 F. Supp. at 1073 n.5 (listing sample jurisdictions using RICO case orders).

[212] *See supra* note 208 and accompanying text.

[213] 28 U.S.C.S. § 332(d)(4) (2001) and FED. R. CIV. P. 83. *See* Walter W. Heiser, *A Critical Review of the Local Rules of the United States District Court for the Southern District of California*, 33 SAN DIEGO L. REV. 555, 557, 560 (1996) (tracing history of federal legislation limiting local rules).

[214] FED. R. CIV. P. 12(e).

[215] *See* 2 MOORE ET AL., *supra* note 98, § 12.36, at 12–87 (citing authority). Professors Wright and Miller have likewise observed, "many cases have denied Rule 12(e) motions on the ground that the information requested was properly the subject of discovery . . . ." 5A WRIGHT & MILLER, *supra* note 46 § 1376, at 585–86. *See also id.* at 589–90, 600 (citing authority).

and restricted the availability of the motion for a more definite statement."[216] Congress enacted this amendment to preserve the "scheme and spirit" of notice pleading under the Federal Rules and to defer elicitation of factual detail to the discovery process.[217]

Even when granted, a motion for more definite statement may require only that a plaintiff provide sufficient information "to frame a responsive pleading."[218] Thus, courts may grant such motions only when a complaint suffers from "unintelligibility[,] not lack of detail."[219] Rather than furnish evidentiary detail, plaintiff's more definite statement need only clarify the basis upon which its claim seeks relief.[220] Finally, although Rule 12(e) authorizes courts to sanction non-compliance by "strik[ing] the pleading to which the motion was directed,"[221] most judges view dismissal as a harsh penalty "that should be used only as a last resort."[222] This trend properly reflects the interplay between Rule 12(e) and Rule 12(b)(6) motions to dismiss. A complaint could state a valid claim yet suffer from ambiguities that render a responsive pleading problematic. Therefore, dismissal is not the appropriate sanction for a somewhat indefinite complaint that otherwise satisfies notice pleading standards.[223]

At least in non-RICO cases, courts have declined to allow Rule 12(e) motions to circumvent either the "notice pleading" formula that pervades federal pleading practice or the standard of review governing dismissal motions.[224] Federal law does not permit courts to achieve this result indirectly, in RICO actions, through a judicial management tool such as a RICO Case Statement.[225] Indeed, dismissal for failure to file a sufficiently

---

[216] 5A WRIGHT & MILLER, supra note 46, § 1376, at 573. The Advisory Committee explained the basis for this amendment as follows: "The tendency of some courts freely to grant extended bills of particulars has served to neutralize any helpful benefits derived from Rule 8, and has overlooked the intended use of the rules on depositions and discovery." *Id.* at 572.

[217] *Id.* at 567–68.

[218] FED. R. CIV. P. 12(e).

[219] 5A WRIGHT & MILLER, *supra* note 46, § 1377, at 601 n.4 (citing cases).

[220] *See id.* at 601–02, 637–38.

[221] FED. R. CIV. P. 12(e).

[222] 5A WRIGHT & MILLER, *supra* note 46, § 1379, at 641 (citing extensive authority)

[223] *See id.* § 1376, at 576–80 & n.8 ("In theory pleadings that may state a claim cannot be dismissed under Rule 12(b)(6) for vagueness . . . . [T]he person to whom the pleading is addressed . . . must seek [a remedy] under Rule 12(e)").

[224] *See supra* note 222 and accompanying text.

[225] *See supra* note 213 and accompanying text. The 1988 Judicial Improvements Act requires each judicial district to appoint an advisory committee concerning potential conflicts between local rules and federal law. 28 U.S.C. § 2071, 2077(b) (2000). In 1997, one such committee report concluded that the district court local rule requiring RICO Case Statements "may be inconsistent with the federal rules." *Report to the Ninth Circuit Judicial Council from the District Court Local Rules Review Committee*, vol. 2, at 29 (1997). The committee explained:

> [The local rule] requires the plaintiff to file a RICO Case Statement along with any complaint which states a RICO cause of action. This case statement contains 20 very specific and precise bases for the RICO action which must be pleaded by

detailed RICO Case Statement renders meaningless the Supreme Court's trilogy upholding notice pleading standards under which the Federal Rules of Civil Procedure are based.

Existing judicial conflicts have also made RICO litigation vulnerable to uncertainty and forum shopping.[226] For RICO to achieve its ambitious legislative goals, similarly situated victims must receive similar results. Only remedial legislation—combined with proper adherence to the Supreme Court's trilogy enforcing Rule 8(a)—can achieve this objective.

Conclusion

In 1974, the United States Chamber of Commerce placed the economic cost of fraud at more than $41 billion annually.[227] A decade later, the Department of Justice reported that fraud losses exceeded $200 billion per year.[228] Even adjusted for inflation, however, these staggering numbers are dwarfed by the recent corporate scandals. One recent analyst has observed that "Enron was just a warm-up act for a series of disasters traceable to accounting 'gimmickry' during which . . . $4.3 *trillion* in market wealth—$15,000 for every American—simply evaporated."[229]

Moreover, during the very period in which the dramatic fraud increases occurred, Congress and the courts undermined RICO—the principal weapon responsible for the demise of traditional organized crime.

> the plaintiff. The local rule states that such a case statement will be construed by the court as an amendment to the pleadings. Since the specific and lengthy case statement is considered a pleading . . . it conflicts with Fed. R. Civ. P. 8(e) which states that no technical forms of pleading shall be required.
>
> The RICO Case Statement . . . also appears inconsistent with Fed. R. Civ. P. 8(a)(1) and Fed. R. Civ. P. 9(b). The RICO Case Statement requires extensive and detailed "fact" pleading, and not the liberal notice pleading envisioned by Rule 8(a). . . . [The] RICO Case Statement [also] goes far beyond the particularity required by Rule 9(b) . . . .
>
> In addition, . . . this local rule states that failure to comply . . . subjects the RICO cause of action to dismissal. This conflicts with case law which holds that a court may not fail to consider a complaint merely because the complainant does not comply with some technical local rules . . . . Moreover, such dismissal could conflict with Fed. R. Civ. P. 83(a)(2) (proposed) which states, "A local rule imposing a requirement of form shall not be enforced in any manner that causes a party to lose rights because of a non-willful failure to comply with the requirement."

Id. at 29–31.

[226] *Compare, e.g.*, Elliot v. Foufas, 867 F.2d 877 (5th Cir. 1989) (sustaining RICO Case Statement), *with* Commercial Cleaning Servs. v. Colin Service Systems, Inc., 271 F.3d 374, 386 (2d Cir. 2001) (rejecting RICO Case Statement).

[227] U.S. Chamber of Commerce, A Handbook on White Collar Crime: Everybody's Problem, Everybody's Loss 6 (1974).

[228] 1984 Att'y Gen. Ann. Rep. 42.

[229] Rob Walker, *Not Even Fuzzy*, N.Y. Times, Apr. 6, 2003, § 7 at 30 (reviewing Alex Berenson, THE NUMBER: How the Drive for Quarterly Earnings Corrupted Wall Street and Corporate America).

Thus far, the legislative response to the corporate scandals has failed to consider strengthening RICO to combat corporate corruption. If the political will exists, however, a series of statutory amendments could resurrect RICO. These amendments should correct the judicially imposed limitations improperly restricting civil RICO. Congress must restore RICO liability for outside professionals who facilitate RICO violations, as well as remove the person-enterprise and investment-injury doctrines. The Appendix contains proposed statutory text implementing these reforms. Once enacted, these measures would deter white-collar racketeering without fostering litigation abuse.

Standing alone, however, legislative reform will be ineffective if courts continue to disregard the "notice pleading" principles upon which the Federal Rules of Civil Procedure are based. The Supreme Court's recent trilogy directing trial courts to abide by these principles provides the required jurisprudential foundation for legislative civil RICO reform to succeed. Taken together, these measures would restore civil RICO to its rightful place in combating systemic commercial fraud.

APPENDIX

The Corporate Rico Reform Act of 2004

(a) 18 U.S.C. § 1961(3) is amended to read as follows:

"person" includes any individual, *perpetrator enterprise*, or entity capable of holding a legal or beneficial interest in property.

Commentary: This proposal adds "perpetrator enterprise" to the definition of person.[230] As such, it eliminates problematic situations in which the person-enterprise distinction precludes a perpetrator RICO enterprise from being named as a defendant person. By limiting the amendment to "perpetrator enterprises," the proposal conforms to existing principles that do not impose vicarious liability upon enterprises victimized by racketeering activity.[231] When violators use the enterprise as an instrumentality of crime, conventional principles of vicarious liability would still apply.

(b) 18 U.S.C. § 1962(c) is amended to read as follows:

It shall be unlawful for any person employed by or associated with any enterprise engaged in, or the activities of which affect, interstate or foreign commerce, to conduct [or], participate, *or aid*, directly or indirectly, in the conduct of such enterprise's affairs through a pattern of racketeering activity or collection of unlawful debt.

Commentary: This proposal legislatively overturns the Supreme Court's decision in *Reves v. Ernst & Young*.[232] It thereby permits RICO to reach corrupt outside professionals—such as lawyers and accountants—who knowingly aid and abet others violating this subsection.

(c) 18 U.S.C. § 1964(c) is amended to read as follows:

---

[230] New Jersey recently enacted a similar amendment to improve its state RICO law. *See* N.J. STAT. ANN. § 2C:41-1(b) (1999) ("'Person' includes any individual or entity *or enterprise* as defined herein holding or capable of holding a legal or beneficial interest in property.") (emphasis added).

[231] The term perpetrator in this context refers to situations in which the corporation is a "central figure in the criminal scheme." Haroco v. Am. Nat'l. Bank & Trust, 747 F.2d 384, 401–02 (7th Cir. 1984). In such cases, the corporation is essentially corrupt. Its "agents were not acting on their own; their conduct . . . was 'authorized' and 'approved.'" G. Robert Blakey & Brian J. Morris, *Threats, Free Speech, and the Jurisprudence of the Federal Criminal Law*, 2002 B.Y.U. L. REV. 829, 349 n.178.

[232] 507 U.S. 170 (1993).

Any person injured in his business or property by reason of a violation of section 1962 of this chapter, *including any racketeering activity that is an element of that violation*, may sue therefor in any appropriate United States district court and shall recover threefold the damages he sustains and the cost, including a reasonable attorney's fee. ~~except that no person may rely upon any conduct that would have been actionable as fraud in the purchase or sale of securities to establish a violation of section 1962. The exception contained in the preceding sentence does not apply to an action against any person who is criminally convicted in connection with the fraud, in which case the statute of limitations shall start to run on the date on which the conviction becomes final~~

Commentary: This proposal eliminates the investment-injury rule and other potential standing requirements by authorizing relief to anyone injured either by a violation of section 1962 or by racketeering activity comprising a part of that completed RICO violation. In addition, the proposal repeals the securities exemption that currently bars certain securities fraud victims from seeking civil RICO relief.

RECENT DEVELOPMENTS

PAYMENT FOR COLLEGE FOOTBALL PLAYERS IN NEBRASKA

The National Collegiate Athletic Association ("NCAA") describes its mission as striving "to maintain intercollegiate athletics as an integral part of the educational program and the athlete as an integral part of the student body."[1] As college athletics, and in particular programs in Division I-A,[2] have become more of a business,[3] however, the emphasis within the term "student-athlete" appears to have shifted away from "student" and toward "athlete." This shift appears most dramatically in the context of the two "revenue-producing" sports, football and men's basketball.[4] In the past twelve years, the amount of money generated by these two sports has increased nearly 300%, such that they now fund almost all other sports programs.[5] Despite this exponential increase, the maximum official compensation for student-athletes has remained largely unchanged: a full grant-in-aid, which includes tuition and fees, room, and required course-related books.[6] As a result of this stagnation, many col-

---

[1] Nat'l Collegiate Athletic Ass'n, *NCAA Mission, Values and Goals*, *at* http://www1.ncaa.org/eprise/main/Public/hr/mission.html (last visited Nov. 8, 2003).

[2] The NCAA is divided into three primary divisions (I, II, and III), generally broken down by the size of the school. The largest schools tend to be in Division I because of its more demanding requirements for event attendance and the number of different sports in which a school must field teams. *See* Nat'l Collegiate Athletic Ass'n, *What's the Difference Between Divisions I, II, and III?*, *at* http://www.ncaa.org/about/div_criteria.html (last visited Nov. 6, 2003). In addition, while athletes in Divisions I and II may receive athletically related financial aid, Division III offers no athletic scholarships. *Id.* In football, Division I schools are further divided into Division I-A schools (the top 117 programs) and Division I-AA (the "non-major" scholarship football programs). *Id.*

[3] *See generally* ANDREW ZIMBALIST, UNPAID PROFESSIONALS: COMMERCIALISM AND CONFLICT IN BIG-TIME COLLEGE SPORTS (1999).

[4] *See* Daniel L. Fulks, *Revenues and Expenses of Division I and II Intercollegiate Athletic Programs 2001*, at 22 tbl. 3.2, *available at* http://www.ncaa.org/library/research/i_ii_rev_exp/2002/d1_d2_revenues_expenses.pdf. In 2001, NCAA Division I-A schools averaged $17.25 million in revenue. *Id.* Combined, football and men's basketball averaged to contribute more than $14.5 million, which represents 84% of all revenues. *Id.* Football alone generates 63% of all revenues at Division I-A schools. *Id.* Some studies have pointed out that while, on average, Division I schools may generate a profit, a number of programs—including highly competitive schools such as the University of Michigan and Ohio State University—are unprofitable or barely break even. *See* Michael Sokolove, *Football is a Sucker's Game*, N.Y. TIMES, Dec. 22, 2002, § 6 (Magazine), at 36 (citing ZIMBALIST, *supra* note 3.

[5] *See* Fulks, *supra* note 4.

[6] 2003-2004 NCAA DIVISION I MANUAL §§ 15.02.5, 15.1.1, *available at* http://www.ncaa.org/library/membership/division_i_manual/2003-04/2003-04_d1_manual.pdf [hereinafter NCAA BYLAWS]. At certain schools, this amount can be as high as $40,000 per year. *See, e.g.*, Duke University Financial Aid Office, *Duke University—Facts & Figures, 2003-2004 Cost of Attendance*, *at* http://dukefinancialaid.duke.edu/facts.asp (last visited Nov. 8, 2003).

legiate athletic teams have turned to extra "unofficial" compensation to gain an advantage on the competition, acting in direct violation of NCAA bylaws.[7] Reports of athletic scandals, ranging from illegal gifts during recruiting to altering an athlete's grades to preserve his or her eligibility, have become an annual occurrence, with the offending schools including well-known institutions like the University of Michigan, Florida State University, and the University of Alabama.[8] This past spring, in response to these scandals, the Nebraska legislature passed Nebraska Legislative Bill 688,[9] which calls for additional compensation of football players at the University of Nebraska-Lincoln (hereinafter "Nebraska-Lincoln" or "the University").[10]

The bill's intent is to legitimize and regulate any payments made to Nebraska football players, in an attempt to prevent "under the table" payments made in violation of NCAA rules.[11] By creating a system of official, aboveboard compensation, the bill aims to allow the players to share in the vast revenues they help create.[12] Implementation of the bill, however, remains problematic. While the legislation purports to be economically fair by providing "big-time" college athletes with a portion of the revenue they generate, were the bill implemented it would be unlikely to withstand legal challenges, most notably under Title IX of the Education Amendments of 1972 ("Title IX").[13]

The Nebraska legislation comes at a time of increasing displeasure with what are often perceived as arbitrary NCAA policies that do not allow student-athletes many of the same opportunities as ordinary students. While the NCAA has emphasized the importance of the amateur aspect of college athletics,[14] it has responded to the changing environ-

---

[7] *See, e.g.*, NCAA BYLAWS § 15.01.2 ("Any student-athlete who received financial aid other than that permitted by the [NCAA] shall not be eligible for intercollegiate athletics.").

[8] For example, the University of Michigan men's basketball team was sanctioned by the NCAA in May 2003 for numerous violations committed in the 1990s, including players' acceptance of regular payments and gifts from a prominent Michigan booster. *See* Avani Patel, *NCAA Adds to Michigan Penalty; Wolverines Lose 4 Scholarships*, CHI. TRIB., May 9, 2003, at C1. The punishment includes a two-year ban from postseason play, the loss of one scholarship (out of thirteen) for four years and three and a half years of probation. *Id. See also* Sokolove, *supra* note 4, at 36 (describing sanctions against Florida State University and the University of Alabama, as well as Gardner-Webb University).

[9] NEB. REV. STAT. §§ 85-1, 131–37 (2003).

[10] *Id.* § 85-1, 134. The Nebraska Cornhuskers have consistently been one of the Top 25 football teams in the nation. The team has won forty-seven conference championships and five national championships, including three in the 1990s. *See* MICHAEL BABCOCK, GO BIG RED: THE COMPLETE FAN'S GUIDE TO NEBRASKA FOOTBALL 24 (1998).

[11] Nebraska Senate Committee on Business and Labor, *Committee Statement—LB 688*, at 2, *available at* http://www.unicam.state.ne.us/PDF/CommitteeStatement_LB688.pdf (Nov. 7, 2003).

[12] NEB. REV. STAT. §§ 85-1, 132 (1), (10) (2003).

[13] Title IX, Education Amendments of 1972, 20 U.S.C. §§ 1681-1688 (2000).

[14] *See, e.g.*, Nat'l Collegiate Athletic Ass'n, *Purpose and Goals*, *at* http://www.ncaa.com/about/purposes.html (last visited Nov. 19, 2003); Nat'l Collegiate Athletic Ass'n,

ment with a number of measures designed to increase the money available to student-athletes. Perhaps most importantly, the NCAA now allows student-athletes, even those receiving full scholarships, to hold paying jobs during the academic year.[15] The NCAA passed this bylaw in response to numerous complaints by student-athletes, which ranged from an inability to afford a McDonald's hamburger[16] to the need to gain professional experience for a non-athletic career.[17] Students demonstrating financial need may also receive money from certain federal grants and the NCAA Special Assistance Fund.[18] While these funds have limits on who may receive them and on what they may be used,[19] the NCAA has also proposed a more general fund to assist other athletes in paying for expenses such as emergency travel and medical costs.[20] The money given from these funds does not count against the maximum allowable financial aid package.[21] Finally, beginning in 2003, students may accrue frequent-flyer miles for all athletics-related travel, which for some players could be worth more than $1,000 per year.[22]

Amateurism, at http://www1.ncaa.org/membership/enforcement/amateurism/index.html (last visited Nov. 19, 2003).

[15] NCAA BYLAWS § 15.2.6. Prior to August, 1998, scholarship athletes could not have paying part-time jobs during the academic year. *See* "DI Proposal: 199-15 Financial Aid—Division I Employment Earnings," *at* https://goomer.ncaa.org/wdbctx/LSDBi/LSDBi.LSDBi_LP_Search.D1_DisplayProposal?p_ID=90&p_HeadFoot=1&p_CallCount=1&p_BylawTerms=ThisIsADummyPhraseThatWillNotBeDuplicated&p_IntentTerms=ThisIsADummyPhraseThatWillNotBeDuplicated&p_RationaleTerms=ThisIsADummyPhraseThatWillNotBeDuplicated (last visited Nov. 6, 2003). Starting in 1998, a full-scholarship student could earn up to $2,000 per year through employment. 2002-2003 NCAA DIVISION I MANUAL §§ 15.2.6.1, 15.02.4.1(a)(6)-(8), *available at* http://www.ncaa.org/library/membership/division_i_manual/2002-03/2002-03_ncaa_d1_manual.pdf. In 2003, however, the NCAA revised this rule and removed the earnings limit. NCAA BYLAWS § 15.2.6.

[16] Dan Wetzel, *Arrogant Michigan Must Accept Truth to Move Beyond It* (Mar. 21, 2002), *at* http://cbs.sportsline.com/b/page/pressbox/0,1328,5157352,00.html.

[17] *See* Lisa Dillman, *NCAA Decides Athletes May Go to Work*, L.A. TIMES, Jan. 14, 1997, at C1; Larry Keech, *NCAA Jobs Bill Raises Questions*, GREENSBORO NEWS & REC., July 14, 1998, at C1.

[18] NCAA BYLAWS § 15.2.4.1(e). In 2000-2001, the average award per recipient from the Special Assistance Fund (including Pell Grants) was $421.91. Nat'l Collegiate Athletic Ass'n, *NCAA Membership Report 2001*, at 12–13, *available at* http://www.ncaa.org/library/membership/membership_report/2001/12-13.pdf (last visited Nov. 8, 2003).

[19] The Special Assistance Fund covers four types of expenditures: (1) clothing or other essential expenditures up to $500, (2) expendable academic course supplies, (3) medical/dental costs not covered by other insurance, and (4) costs associated with student-athlete or family emergencies. *See* Nat'l Collegiate Athletic Ass'n, *Guidelines for Use of Fund Clarified*, NCAA NEWS (Sept. 24, 2001), *available at* http://www.ncaa.org/news/2001/20010924/div1/3820n16.html.

[20] *See* Mark Alesia, *New NCAA Assistance Fund Will Benefit Larger User Group*, INDIANAPOLIS STAR, Apr. 16, 2003, at 1D.

[21] NCAA BYLAWS § 15.01.7.1.

[22] NCAA BYLAWS § 16.12.1.11(b). Members of the University of Hawaii men's basketball team and men's and women's swimming teams each accrued about 57,400 miles in the 2002-03 season; *Consumer Reports* magazine values those miles at about $.02 each, for a total value of $1,148. Jack Carey, *College Athletes Can Collect Frequent-Flier Miles*, USA TODAY, June 2, 2003, at 3C.

Despite these recent changes, the NCAA has faced increasing criticism, both in the media and in state legislatures. In July 2003, the University of Utah basketball team was placed on probation for a series of minor violations by its coach. These violations included the coach's paying for a pizza he shared with a player at a restaurant, when there would have been no violation had he brought the pizza back and served it in his home.[23] In response to the media outcry,[24] the NCAA recently amended its bylaws to remove the seemingly trivial distinction of where a meal was consumed, so long as the school makes appropriate documentation.[25]

The case of University of Colorado football player and world champion freestyle skier Jeremy Bloom has also gained the attention of NCAA critics.[26] In 2002, the NCAA informed Bloom that he could not receive any endorsement money based on his Olympic skiing success or he would be ruled ineligible to compete in college football.[27] Bloom has since filed suit, claiming the restriction is unenforceable under state law.[28] In response, NCAA President Myles Brand has defended the NCAA's treatment of student-athletes, outlining the many funds available for financially needy students and the $1 billion in scholarships spent on student-athletes each year.[29]

Nonetheless, the publicity surrounding the Bloom case has caused legislators in at least one other state to take action. Last year, two California state senators introduced a bill that would prohibit universities in the state from adhering to NCAA rules that limit the rights of athletes, including prohibitions on earning income from their name, hiring an agent, and transferring schools.[30] This proposal, along with Nebraska Legislative Bill 688, represents perhaps the strongest challenge to the NCAA in its history.

The Nebraska bill focuses on football for a number of reasons. With few exceptions, only football and men's basketball generate significant revenues for college athletic departments,[31] making these sports logical

---

[23] *See* Robyn Norwood, *Utah is Penalized for Rules Violations*, L.A. TIMES, July 31, 2003, § 4, at 5. *See also* NCAA BYLAWS § 16.12.1.5.

[24] *See, e.g.*, Rick Reilly, *Corrupting Our Utes*, SPORTS ILLUSTRATED, Aug. 11, 2003, at 154.

[25] *See* Norwood, *supra* note 23, at 5.

[26] *See, e.g.*, *Let Jeremy Bloom Play C.U. Football; NCAA Rules Needlessly Punish Skier*, ROCKY MTN. NEWS (Denver), July 31, 2002, at 42A; Irvin Muchnick, *Welcome to Plantation Football*, L.A. TIMES, Aug. 31, 2003, at I-14.

[27] *See Bloom Hopes to Make Choice by June 1*, ROCKY MTN. NEWS (Denver), May 14, 2003, at 13C. Skiers rely on endorsement deals to finance their careers, because there is little money elsewhere in the sport. *Id.*

[28] *See* Adam Thompson, *Bloom's Lawyers File Brief*, DENV. POST, July 3, 2003, at D-10.

[29] Adam Thompson, *Bloom Receives Response from Brand*, DEN. POST, Aug. 17, 2003, at C-12.

[30] S.B. 193, 2003-2004 Reg. Sess. (Cal. 2003).

[31] *See* Fulks, *supra* note 4, at 22 tbl.3.2. At Nebraska-Lincoln, the football team ac-

choices for athlete compensation. In Nebraska, however, while the men's basketball team is modestly successful,[32] few things can rival Husker football for importance in the state. The team has sold out 73,918-seat Memorial Stadium more than 250 consecutive times, which makes the stadium the third-largest "city" in the state on game days.[33] Fans routinely drive three hours or more to attend each game and are willing to trek thousands of miles to watch the Huskers in action.[34] Senator Ernie Chambers, the bill's sponsor, estimates that Nebraska-Lincoln football generated $155 million in revenue from 1994 to 2003.[35] As Chambers notes, however, the University distributed only $14 million in scholarships and aid to the players during this same period.[36]

As a result, Chambers proposed Legislative Bill 688, which the Legislature approved and Governor Mike Johanns signed into law on April 16, 2003.[37] The text of the bill describes the problems inherent in the college athletic recruiting and competition process, including "rampant" scandals, players from "impoverished families" who are "vulnerable to inducements, benefits, and other types of [illicit] compensation," and the "unduly restrictive and unreasonable" rules of the NCAA.[38] The bill goes on to assert that the University's interest in maintaining a successful football team, due to the income generated, the publicity gained, and the overall benefit to the University's image, often places substantial burdens on student-athletes, who are recruited solely for their athletic ability and not because of any academic achievements.[39]

---

counted for more than $32.6 million of the more than $47.2 million in revenues generated by the athletic department between July 1, 2001 and June 30, 2002. Office of Postsecondary Educ., U.S. Dep't of Educ., *Equity in Athletics Disclosure Website*, *at* http://www.ope.ed.gov/athletics/index.asp (last visited Nov. 8, 2003) [hereinafter *Equity in Athletics*].

[32] The Husker men's basketball team won the NIT Championship in 1996 and reached the NCAA Tournament numerous times in the past two decades, but has never been in serious contention for the National Championship, as the football team is nearly every year. *See* Univ. of Neb. Athletic Dep't, NEBRASKA BASKETBALL RECRUITING AND MEDIA GUIDE (2003); Babcock, *supra* note 10, at 24.

[33] *See* Univ. of Neb. Athletic Dep't, OFFICIAL NEBRASKA [FOOTBALL] MEDIA AND RECRUITING GUIDE 324 (2003); Ctr. for Pub. Affairs Research, Univ. of Nebraska-Omaha, *Nebraska Cities and Towns Ranked by 2000 Population*, *at* http://www.unomaha.edu/~cpar/table_2b1.pdf (noting that the third-largest city in the state, Bellevue, has a population of 44,382).

[34] *See* Terry Douglass, *Husker Fans Flock to Desert for Fiesta Bowl* (Jan. 2, 2000), HUSKERS HQ, *at* http://www.theindependent.com/stories/010200/Hus_huskernews0102.html (last visited Aug. 10, 2003).

[35] *See* Milan Simonich, *Pay for Play?*, PITTSBURGH POST-GAZETTE, Mar. 30, 2003, at D1.

[36] *Id.* The remainder of the money was used primarily to finance other sports at the University. *See* Fulks, *supra* note 4.

[37] Bill Status, L.B. 688, *available at* http://www.unicam.state.ne.us/scripts/dbBSInfo.asp?Prefix=L.B.&BillNumber=688&Suffix=&Session= (last visited Nov. 8, 2003).

[38] NEB. REV. STAT. §§ 85-1, 131(1)–(2), (4) (2003).

[39] *Id.* §§ 85-1, 132.

The bill establishes two methods for protecting the rights of these athletes against potential abuse by universities. The first alternative provides compensation "in the same manner that non-athlete students are compensated for performing various tasks while a student,"[40] most likely in the form of a stipend.[41] The bill states that the amount of the stipend should be set by the University,[42] with most state officials supporting an amount of $200–$400 per month.[43] The legislation also contains the option of granting similar stipends to other athletes competing in inter-conference athletics.[44]

While the bill does not provide examples of the "various tasks" non-athletes perform or of the compensation received by students who perform them,[45] students at schools across the nation do indeed receive benefits for serving in certain roles, such as student government.[46] Compensation varies, but often includes stipends, free cell phones, money for clothing, and even salaries in some cases.[47] Two key differences exist, however, between the payment of student government officials and student-athletes. First, all of these forms of compensation, if offered to athletes, would violate NCAA rules and lead to penalties, both for the student-athlete and for the college.[48] Second, as opponents of the bill note, student-athletes already receive a great deal of benefits, both financial and otherwise.[49] Full athletic scholarships include tuition, fees, books, and room and board.[50] In addition, athletes often receive team-related clothing and equipment and have access to medical facilities and aca-

---

[40] *Id.* §§ 85-1, 133.

[41] *Id.* §§ 85-1, 134.

[42] *Id.*

[43] *See, e.g.*, Simonich, *supra* note 35.

[44] NEB. REV. STAT. §§ 85-1, 134 (2003).

[45] Nebraska Revised Statute Section 85-1, 137 refers to part-time work "of at least twelve hours a week;" however, this section is labeled as a "preferable alternative" to Section 85-1, 134, and not an addendum or definition of the terms in Section 85-1, 133. *Id.* § 85-1, 137. In addition, as noted *supra* note 15, in 1998, the NCAA changed its bylaws, allowing full-scholarship athletes to hold part-time jobs. Thus, it does not follow that the intent of the bill is to secure the ability of athletes to gain employment.

[46] *See, e.g.*, Am. Student Gov't Ass'n, *The Student Government Salary Survey*, *available at* http://www.studentleader.com/sl_16.htm (last visited Nov. 8, 2003). The survey reports that 71% of undergraduate institutions nationwide, including 85% of public schools and 87% of schools with enrollments greater than 30,000, provide some sort of compensation for elected student leaders. *Id.*

[47] The most common type of compensation offered is an hourly wage (often near minimum wage) based on a forty-hour work week or a stipend of $200 to $400 per month. *Id.* Nebraska Governor Mike Johanns reportedly favors a pay structure for football players similar to the stipend and for approximately the same amount. *See* Simonich, *supra* note 35.

[48] *See* NCAA BYLAWS §§ 12.1.1 (defining penalities against students), 19 (defining the enforcement procedures and penalties against institutions).

[49] *See, e.g.*, John Markon, *Cornhuskers' Payment Plan is a Bad Idea*, RICH. TIMES DISPATCH, Feb. 21, 2003, at D1.

[50] NCAA BYLAWS § 15.02.5.

demic resources on campus for no charge.[51] The value of these items far exceeds the highest amount paid to a student government official.[52]

Perhaps in an attempt to avoid these conflicts, the bill also provides an alternative to athlete compensation. Section 7 of the bill proposes a limitation on the number of hours in which an athlete can participate in a sport, including practices, games, and preparation sessions.[53] The bill is silent on the exact number of hours, but it dictates that the number should be low enough so that student-athletes can have a normal academic schedule, graduate in four years, participate in campus activities, and work an average of twelve hours per week.[54] It should be noted that limitations already exist on the amount of time a Division I athlete can participate in official practices, training sessions, and games. The NCAA bylaws limit an athlete to four hours a day and twenty hours a week of "athletically-related activities"[55] during the "playing season."[56] Collegiate coaches, however, routinely avoid this limitation through "optional" sessions and captain-led practices, with no coaches present.[57] As a result of these "volandatory"[58] sessions, some student-athletes spend up to sixty hours per week focusing on their sport.[59] There is no indication that Legislative Bill 688's provisions would in any way prevent this manipulation.

Senator Chambers proposed the bill "to let the NCAA know that legislators are concerned about the treatment of athletes and that the rules relative to financial assistance must be modified . . . ."[60] This "Declaration of Independence" from NCAA regulations aims to bring about fairness and "equity of the marketplace" in the relationship between athletes and the governing institutions.[61] While the law cannot take effect until four

---

[51] *See* NCAA BYLAWS §§ 16.4 (academic resources), 16.4 (medical expenses), 16.8 (practice and competition expenses).

[52] The highest documented benefits package for a student leader is $25,000, paid to the student council president of Northeastern University in Boston (in housing, tuition, stipend, parking, and other costs). *See* Am. Student Gov't Ass'n, *supra* note 46.

[53] NEB. REV. STAT. §§ 85-1, 137 (2003).

[54] *Id.*

[55] NCAA BYLAWS § 17.02.1. This includes all required activity completed under the supervision of a coach, including practice, strength and conditioning, film sessions, and competition. *Id.*

[56] NCAA BYLAWS §§ 17.1.5.1, 17.11 (defining the football playing season).

[57] *See* Liz McCaslin & Kenan Smith, *How Long Do We Practice Anyway?*, IN THE SAAC, Fall 2003, at 1, *available at* http://www.atlantic10.org/saac/newsletter/VolIIssueI1.pdf. The Student-Athlete Advisory Committee has begun to focus on this issue and has made several recommendations to the NCAA. *See* NCAA Student-Athlete Advisory Comm., *Current Issues and Accomplishments*, *at* http://www1.ncaa.org/membership/membership_svcs/saac/d1/SAACAccomplishments (last visited Nov. 8, 2003).

[58] Muchnick, *supra* note 26.

[59] *See id.* ("Coaches book every hour of [a student-athlete's] non-classroom time.").

[60] Ernie Chambers, *Introducer's Statement of Intent for* LB 688, 98th Leg., 1st Sess. (Neb. 2003), *available at* http://www.unicam.state.ne.us/PDF/StatementOfIntent_LB688.pdf (last visited Nov. 7, 2003).

[61] *Id.*

other states home to Big 12 schools[62] pass similar laws,[63] the legislature provided this only as a protection for the University's NCAA eligibility.[64] Although the law's main purpose may be only to encourage the NCAA to provide greater protections for student-athletes, the fact remains that the current law in Nebraska is designed to pay a stipend to college football players. Indeed, the philosophy animating the law rejects amateur athletics and holds that University of Nebraska football players are entitled to legitimate, aboveboard monetary compensation for their athletic services.

Attempts to enforce this bill, or any law similar in scope, will most likely prove exceedingly difficult. Most obviously, the law contradicts a number of NCAA regulations, and in many ways, the spirit of college amateur athletics. NCAA bylaws clearly prohibit any compensation to student-athletes beyond the amount necessary for their education and student expenses.[65] The NCAA prohibits several forms of payment to student-athletes, including "[a]ny direct or indirect salary, gratuity or comparable compensation."[66] In an apparent attempt to avoid this prohibition, the Nebraska legislature changed the original wording of the bill. As introduced, the bill called for players to be paid "compensation" no less than the federal minimum wage, making them employees of the University.[67] The legislators revised the bill before passage, calling instead for a "stipend" and stating that "[n]othing in this act shall be construed to make a person a professional athlete."[68]

The NCAA bylaws, however, appear to prohibit even the revised bill. Section 12 of the bylaws defines a professional athlete as "one who receives any kind of payment, directly or indirectly, for athletics participation *except as permitted* by the [NCAA]."[69] The bylaws further state that an athlete loses his or her amateur status, and thus the ability to compete in the NCAA in a certain sport, if the athlete "[u]ses his or her athletics skill (directly or indirectly) for pay in any form in that sport."[70] The Bloom case showcases that even an athlete attempting to achieve financial gain on the basis of success in a different sport will not be permitted to play in the NCAA. Thus, an attempted payment plan based on skill in the intercollegiate sport has little chance for survival, and the bill's mandate that "football players shall be entitled to fair financial compensation

---

[62] Besides Nebraska, the states with schools in the Big 12 are Colorado, Iowa, Kansas, Missouri, Oklahoma, and Texas.

[63] NEB. REV. STAT. §§ 85-1, 136 (2003).

[64] *See* Chambers, *supra* note 60 ("This is a 'fail-safe' provision that protects the university from risking its eligibility when LB 688 becomes law.").

[65] *See* NCAA BYLAWS §§ 15, 15.1, 15.02.5.

[66] *Id.* § 12.1.1.1.1.

[67] L.B. 688, 98th Leg., 1st Sess. (Neb. 2003), Initial Reading, § 4, *available at* http://www.unicam.state.ne.us/pdf/INTRO_LB688.pdf.

[68] NEB. REV. STAT. §§ 85-1, 134–35 (2003).

[69] NCAA BYLAWS § 12.02.3 (emphasis added).

[70] *Id.* § 12.1.1(a).

for playing football"[71] would most likely result in all of the Nebraska football players' losing their amateur status and ability to compete in the NCAA.

If, however, a number of schools grow tired of the NCAA's seemingly arbitrary rules and inconsistent enforcement procedures, nothing would prevent them from leaving the organization and forming a new governing body. Since the recent expansion of the Atlantic Coast Conference,[72] some analysts believe that the five biggest conferences could separate from the NCAA,[73] taking with them large television contracts for football and men's basketball.[74] The approximately sixty schools that would comprise a new organization represent the most prolific revenue-producing athletic departments in the nation.[75] Already, the biggest conferences have formed the Bowl Championship Series for football, which is the only college championship not conducted by the NCAA.[76] If these schools severed all ties with the NCAA, they would be able to keep a larger portion of television and advertising revenues[77] and would be able to compensate their players in any manner, including the stipends proposed by the Nebraska legislature. If the California Senate succeeds in preventing its schools from following NCAA mandates,[78] other states could follow suit, thereby increasing the chance that a rival organization to the NCAA could form.

Although the formation of a new governing body is a necessary step in implementing Legislative Bill 688, it is not enough to protect the law from its largest obstacle, Title IX.[79] The University of Nebraska can leave the NCAA, receive the unanimous support of the state legislature, and be granted funds by the Governor; however, barring the repeal of Title IX, a law compensating only football players, all of whom are male, likely will not withstand legal challenges.

---

[71] NEB. REV. STAT. §§ 85-1, 133 (2003).

[72] *See Conference Call; Miami Joins Virginia Tech in Making an 11-Team ACC*, CHI. TRIB., July 1, 2003, at 15.

[73] *See, e.g.*, Gregg Doyel, *Divide and Conquer; 'Superconferences' Threaten NCAA's Power*, RICH. TIMES-DISPATCH, June 2, 2003, at C-2.

[74] *See id.* ABC is paying $930 million over seven seasons for the rights to the Bowl Championship Series. *See* STREET & SMITH'S SPORTS BUSINESS JOURNAL BY THE NUMBERS (2002) 87. CBS will pay the NCAA $6 billion over eleven years for the right to the Division I Men's Basketball Tournament. *Id.*

[75] *See Equity in Athletics*, *supra* note 31.

[76] The BCS is a collaboration of the six major conferences (Big 12, Big 10, Pacific-10, Atlantic Coast Conference, Southeastern Conference, and Big East) and Notre Dame, an independent school. *See Bowl Championship Series*, *at* http://www.bcsfootball.org. The NCAA plays no role in the BCS, other than sanctioning the games. *Id.*

[77] Currently, the NCAA spends approximately $30 million per year on Division II and III athletics. *See* NCAA 2001-2002 Approved Budget, *available at* http://www.ncaa.org/financial/2001-02_budget.pdf (last visited Aug. 10, 2003). If the top sixty schools separated from the NCAA, they would each receive, therefore, approximately $500,000 in additional revenue per year.

[78] *See supra* note 30 and accompanying text.

[79] 20 U.S.C. §§ 1681–1688 (2003).

Title IX states that "no person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance."[80] Congress passed Title IX in 1972, after a House of Representatives investigation revealed extensive discrimination against women with respect to educational opportunities.[81] In an attempt to remedy this discrimination, Congress used its most powerful weapon: federal funding. By withholding financial assistance from educational institutions that continued favoring men, lawmakers hoped "to provide individual citizens effective protection against [discriminatory] practices."[82]

The original text of Title IX made no reference to athletics or athletic programs,[83] but a 1975 regulation implementing Title IX[84] included specific requirements for both intercollegiate athletics[85] and athletic scholarships.[86] As a result, the late 1970s witnessed an incredible growth in athletic opportunities for women, with the number of female intercollegiate athletes doubling.[87]

In the original Act, Congress did not provide universities with guidelines for what actions constituted progress in ending discriminatory practices. In 1979, the Department of Health, Education, and Welfare developed a three-prong test for "effective accommodation" under the statute.[88] Under the test, a school is in compliance with Title IX if:

> (1) intercollegiate level participation opportunities for male and female students are provided in numbers substantially proportionate to their respective enrollments; or
>
> (2) if the members of one sex have been and are underrepresented among intercollegiate athletes, the institution can show a history and continuing practice of program expansion which is demonstrably responsive to the developing interest and abilities of the members of that sex; or
>
> (3) if the members of one sex are underrepresented among intercollegiate athletics, and the institution cannot show a con-

---

[80] *Id.* § 1681.

[81] 118 CONG. REC. 5804 (1972); *see also* N. Haven Bd. of Educ. v. Bell, 456 U.S. 512, 523 n.13 (1982).

[82] Cannon v. Univ. of Chi., 441 U.S. 677, 704 (1979); Cohen v. Brown Univ., 101 F.3d 155, 165 (1st Cir. 1996) [hereinafter *Cohen II*].

[83] 20 U.S.C. § 1681 (1972), *available at* http://www.dol.gov/oasam/regs/statutes/titleix.htm.

[84] 34 C.F.R. § 106.41 (2003).

[85] *Id.*

[86] *Id.* § 106.37(c).

[87] *See* Bart Barnes, *More Women Than Ever in Sports, but Report Says Bias Still Exists*, WASH. POST, July 27, 1980, at E11.

[88] 44 Fed. Reg. 71,413 (1979) [hereinafter *Policy Interpretation*].

> tinuing practice of program expansion under the second prong, the school can show that its current program nevertheless fully and effectively accommodates the interests and abilities of the members of the underrepresented sex.[89]

This three-part test has remained the official policy of the federal government since 1979, and the Office for Civil Rights [hereinafter "OCR"] of the Department of Education has twice reconfirmed its commitment to it.[90]

Over the past twenty-five years, the test has been the subject of much legal criticism. Scholars have decried it for being impractical and limited,[91] producing gender quotas,[92] and placing too great an emphasis on substantial proportionality at the expense of the other factors.[93] Despite this criticism, and the fact that the Policy Interpretation alone does not have the force of law, courts have continued to use this test as the standard for judging Title IX claims.

Title IX jurisprudence has consistently favored women's programs, even at the expense of opportunities for men.[94] In 1984, the Supreme Court attempted to limit the reach of Title IX, holding in *Grove City College v. Bell*[95] that Title IX applied only to programs that directly benefited from federal funds. As most collegiate athletic departments do not receive money directly from the federal government,[96] colleges were no longer compelled to comply with Title IX in regard to athletics, and the athletic balance once again favored male athletes and teams.[97] The change in Title IX policy proved short-lived, however, as Congress

---

[89] *Id.* In 1984, Title IX fell under the authority of the Department of Education, which adopted this language rather than formulating a new policy. Kelley v. Bd. of Trs., Univ. of Ill., 35 F.3d 265, 269 (7th Cir. 1994); Cohen v. Brown Univ., 991 F.2d 888, 895–96, n.10 (1st Cir. 1993) [hereinafter *Cohen I*].

[90] OFFICE FOR CIVIL RIGHTS, U.S. DEP'T OF EDUC., CLARIFICATION OF INTERCOLLEGIATE ATHLETICS POLICY GUIDANCE: THE THREE-PART TEST, *in* Letter from Norma V. Cantú, Assistant Secretary for Civil Rights, to Colleague (Jan. 16, 1996) *at* http://www.ed.gov/about/offices/list/ocr/docs/clarific.html [hereinafter 1996 CLARIFICATION]; OFFICE FOR CIVIL RIGHTS, U.S. DEP'T OF EDUC., FURTHER CLARIFICATION OF INTERCOLLEGIATE ATHLETICS POLICY GUIDANCE REGARDING TITLE IX COMPLIANCE, *in* Letter from Reynolds, Assistant Secretary for Civil Rights, to Colleague (July 11, 2003) *at* http://www.ed.gov/about/offices/list/ocr/title9guidanceFinal.html [hereinafter 2003 CLARIFICATION].

[91] *See, e.g.*, David Klinker, *Why Conforming with Title IX Hurts Men's Collegiate Sports*, 13 SETON HALL J. SPORT L. 73, 83 (2003).

[92] *See, e.g.*, Sara A. Elliot & Daniel S. Mason, *Gender Equity in Intercollegiate Athletics: An Alternative Model to Achieving Title IX Compliance*, 11 J. LEGAL ASPECTS SPORT 1, 10 (2001).

[93] *See, e.g.*, Mary W. Gray, *Setting a Course for College Athletics: The Concept of Substantial Proportionality in Title IX Athletics Cases*, 3 DUKE J. GENDER L. & POL'Y 165, 167–68 (1996).

[94] *See* David Hancock Moon, *Gender Inequity? An Analysis of Title IX Lawsuits in Intercollegiate Athletics*, 6 DEPAUL-LCA J. ART & ENT. L. 87 (1995).

[95] 465 U.S. 555, 563 (1984).

[96] *Cohen I*, 991 F.2d at 894.

[97] *Id.*

passed the Civil Rights Restoration Act of 1987, reinstating Title IX protection for intercollegiate athletes.[98] This Act, passed over President Reagan's veto,[99] applied Title IX's regulations to all operations of a program receiving federal funding.[100] Thus, a university that receives any federal assistance[101] must comply with Title IX in all of its programs, including athletics.

After *Grove City*, Title IX jurisprudence has consistently protected female athletic programs. In *Cohen v. Brown University*, the First Circuit held that a university's decision to downgrade two women's varsity teams constituted a *prima facie* violation of the second and third prongs of the 1979 test.[102] Because Brown also failed the first prong of the test,[103] the University's action was ruled a violation of Title IX.[104] Brown argued that because the plan also eliminated two men's teams, it provided equal, albeit incomplete, opportunities for both male and female student-athletes, thereby satisfying the third prong.[105] For the court, however, anything less than full accommodation of women's interests would result in a violation of prong three, even if men's interests were not fully accommodated.[106] According to the majority, an institution must bring itself into compliance, if necessary, by "subtraction and downgrading, that is, by reducing opportunities for the overrepresented gender while keeping opportunities stable for the underrepresented gender (or reducing them to a much lesser extent)."[107]

In 1996, the First Circuit again considered the case and again held against Brown. In this decision, the court held that even if Brown could prove that women had less interest in sports than men, "such evidence, standing alone, cannot justify providing fewer athletic opportunities for women than for men."[108] The Supreme Court denied *certiorari*,[109] and these cases remain the primary legal precedent for interpreting the three-prong test.

Few cases exist that deal with the specific question of athletically related financial assistance. In fact, no court has ever allowed recovery for the disproportionate allocation of financial assistance to athletes.[110]

---

[98] Pub. L. No. 100-259, §§ 908(2)(a), (3)(a), 908, 102 Stat. 28-29 (1988).

[99] *See* George E. Curry, *Civil Rights Veto Overridden*, CHI. TRIB., Mar. 23, 1988, at C1.

[100] Pub. L. No. 100-259, §§ 908(2)(a), (3)(a), 908, 102 Stat. 28-29 (1988).

[101] More than 2000 colleges and universities across the nation receive federal financial assistance. *See Equity in Athletics*, *supra* note 31.

[102] *Cohen I*, 991 F.2d at 896–99.

[103] At the time of the lawsuit, roughly 52% of Brown's student population was male, but men constituted approximately 63% of the school's student-athletes. *Id.* at 898.

[104] *Id.* at 900.

[105] *Id.* at 899.

[106] *Id.* at 898.

[107] *Id.* at 898 n.15.

[108] *Cohen II*, 101 F.3d at 180.

[109] Brown Univ. v. Cohen, 520 U.S. 1186 (1997).

[110] Judith Jurin Semo et al., *A Guide to Recent Developments in Title IX Litigation—February 15, 2000*, *in* NCAA, ACHIEVING GENDER EQUITY, III, 2, *available at*

This fact, however, can be attributed in large part to certain evidentiary and procedural barriers plaintiffs must overcome to prove such a case. In *Beasley v. Alabama State University*, the court found that the institution failed to allocate scholarships in a proportionate amount for one year, including denying the plaintiff a scholarship; however, the court found no other year in which such discriminatory behavior had taken place, and thus, no continuing violation as required under Title IX.[111] Likewise, in *Boucher v. Syracuse University*, while the court found evidence of disproportionate allocation of scholarship funds, the plaintiffs had graduated by the time the case reached trial, destroying their standing to sue.[112]

Similar barriers likely would not exist in a suit concerning Legislative Bill 688. A continuing violation could be easily proven if Nebraska-Lincoln followed the law and provided only football players (all males) with additional compensation. Even in the first year of implementation, most courts would likely view the law as *prima facie* evidence of an intent to commit a continuing violation. Furthermore, the finding of such a continuing violation would remove the problem of standing, as any female student-athlete could bring a claim that would likely be successful prior to her graduation.

A school complies with Title IX in its financial assistance policies by allocating total scholarship dollars in an amount proportionate to each gender's level of participation in the intercollegiate sports program.[113] Thus, if 60% of all athletes are male and 40% of all athletes are female, then men should receive 60% of the scholarship money and women 40%. A school can award as many or as few scholarships as it wishes, so long as the total amount of money awarded to men and women remains in proportion.[114] While the calculation does not include expenses for when classes are not in session, all dollars awarded during the academic year count against the total.[115] The stipend proposed by Legislative Bill 688 would increase the percentage of financial aid awarded to men and trigger a violation of Title IX.

The case law appears to indicate, moreover, that if the courts were to allow an institution to allocate funds disproportionately, it would most likely do so only to assist the underrepresented sex.[116] Since Congress

---

bruary 15, 2000, *in* NCAA, ACHIEVING GENDER EQUITY, III, 2, *available at* http://www.ncaa.org/library/general/achieving_gender_equity/current_case_law.pdf (last visited Nov. 8, 2003).

[111] 3 F. Supp. 2d 1325, 1329, 1336-37 (M.D. Ala. 1998).

[112] 164 F.3d 113, 115 (2d Cir. 1999).

[113] *See Policy Interpretation*, *supra* note 88, at 71,413. *See also* Valeria M. Bonnette, *Title IX Basics*, *in* NCAA, ACHIEVING GENDER EQUITY, II, 8, *available at* http://www.ncaa.org/library/general/achieving_gender_equity/title_ix_basics.pdf (last visited Nov. 8, 2003).

[114] *See Policy Interpretation*, *supra* note 88, at 71,415.

[115] *Id.*

[116] I was unable to find any cases in which males were claimed to be the underrepresented sex, although this issue may arise in the future.

overruled the Supreme Court decision in *Grove City*, every major court decision has favored the underrepresented gender.[117] In *Gonyo v. Drake University*,[118] the court held that the over-allocation of funds to the underrepresented sex, women, did not violate Title IX. The rulings in both *Cohen* cases also seem to favor advantages to women, even if they must come at the expense of male programs.[119] Thus, it seems unlikely that a court following a similar line of reasoning would permit "extra" compensation for male athletes. In order to avoid cutting other men's sports expenses, supporters of Legislative Bill 688 would need to lobby for the stipend not to be included in Title IX calculations, perhaps by exempting football altogether from the statute.

Such a rule could perhaps be justified by the much larger size of football rosters in comparison with other sports,[120] as well as the incredible amount of revenue generated by most football teams.[121] At many schools, the money raised by the football program helps finance almost the entire athletic department.[122] Thus, proponents of Legislative Bill 688 might argue, these football players deserve increased compensation without upsetting the balance of Title IX.

Unfortunately for the advocates of the bill, these rationales have been proposed and rejected numerous times throughout the history of Title IX. On two different occasions, Congress has declined an opportunity to exempt revenue-producing sports from Title IX calculations.[123] Thus far, Congress has enacted only one exception for football. The Javits Amendment, passed in 1974,[124] excludes legitimate, non-gender related differences in sports, such as equipment costs and event management. Therefore, a university can still spend much more in aggregate on football, a program with higher facility and equipment costs, than on women's sports teams.[125] Proponents of Title IX, however, have ada-

---

[117] *See* Semo, *supra* note 110, at 2–12.

[118] 837 F. Supp. 989, 995-96 (S.D. Iowa 1993).

[119] *Cohen I*, 991 F.2d at 906; *Cohen II*, 101 F.3d at 176.

[120] The NCAA permits Division I-A football teams eighty-five scholarships and does not place limits on roster sizes. NCAA BYLAWS § 17.11. In 2001, Nebraska-Lincoln's football team had 183 members. *See* Markon, *supra* note 49. In comparison, most basketball teams have eleven to thirteen players, soccer teams usually maintains a roster of twenty-five, and baseball teams frequently field fewer than thirty players. *See, e.g.*, Duke University Athletics Rosters, *available at* http://www.goduke.com (last visited Nov. 8, 2003).

[121] *See supra* note 4.

[122] *See Equity in Athletics*, *supra* note 31.

[123] In 1974, the Senate rejected the Tower Amendment, which would have provided the exemption in the Title IX legislation. 120 Cong. Rec. 15, 322 (1974). *See generally* "History of Title IX Legislation, Regulation and Policy Interpretation," available at http://bailiwick.lib.uiowa.edu/ge/documents/history.pdf (last visited Nov. 10, 2003). In 1977, Senator Tower again proposed the exemption, this time in a separate bill, but once again it was rejected. S. 2106, 94th Cong. (1977).

[124] Education Amendments of 1974 (Title VIII), Pub. L.
No. 93-380, § 844, 88 Stat. 484 (1974).

[125] *See Q & A with Donna Lopiano: What Title IX Means*, AUSTIN AMERICAN-

mantly stated that this represents enough of, if not too much of, a concession.[126] Thus, Congress seems unlikely, at least at this time, to grant football any further exemptions.

Without an exemption, providing additional compensation to football players would only widen the gap between male and female student-athletes. In order to comply with Title IX, the Nebraska-Lincoln Athletic Department would have two options, both of which would necessitate further cutbacks in other men's sports. The first option would be to pay female athletes in a fashion equal to their male counterparts. Thus, if football players received a total of $500,000 in stipends, universities could provide female athletes with a similar amount. The text of the bill provides for such a plan, leaving the payment of other athletes to the discretion of the University.[127] This option presents a fiscal problem: a number of athletic departments currently run deficits and are looking to cut spending.[128] Adding a stipend would increase the athletic budget by anywhere from $300,000 to $1.3 million,[129] which most schools simply cannot afford.

Thus, the most likely course of action to ensure Title IX compliance would be the second option: cutting spending on men's sports by an amount equal to the stipend, which would result in no change to the overall men's sports budget. Cutting men's expenses means eliminating non-football male sports opportunities. Since the inception of Title IX thirty years ago, more than 400 men's programs have been discontinued; most, if not all, were eliminated to comply with the statute.[130] Just recently, the Nebraska-Lincoln Athletic Department eliminated its male swimming and diving team because the school could not afford its estimated $500,000 budget.[131] Thus, a stipend of several hundred thousand

---

STATESMAN, June 24, 2002, at A10; Donna Lopiano, *Football Doesn't Need Title IX Protection*, (June 5, 2002), *at* http://www.womensportsfoundation.org/cgi-bin/iowa/issues/rights/article.html?record=131.

[126] Lopiano, *supra* note 125.

[127] NEB. REV. STAT. §§ 85-1, 134 (2003).

[128] In 2001–2002, sixty-five percent of Division I schools, including programs such as Duke, the University of Connecticut, and Marquette University, all had budget deficits. *See* Ted Hutton, *Playing the Price; Higher Costs, Lower Profit Have Many D-I Programs Seeing Red*, SUN-SENTINEL (Ft. Lauderdale), Apr. 10, 2003, at 1C. In most cases, a successful football program marked the difference between a profitable and non-profitable department; however, the University of Miami (Fla.) lost $1.5 million in 2002, despite winning the national championship. *Id.*

[129] Nebraska's Governor reportedly favors a stipend of $200–$400 per month. *See* Simonich, *supra* note 35. Paying 85 scholarship football players (the maximum allowed by the NCAA) $200 each per month over a nine-month academic year would cost a school $153,000. Matching those funds for female student-athletes would result in a total expenditure of $306,000. If all 183 Nebraska football players (including non-scholarship athletes) were paid $400 per month, the amount would be $658,800, with matching funds for women bringing the total to $1,317,600.

[130] *See* Terri Cotten, *Title IX Hearings Review Equality; Men's Sports Suffer from Changes, Some Say*, DENV. POST, Oct. 23, 2002, at D-02.

[131] *See Nebraska Cutting Men's Swimming and Diving*, http://www.theindependent.

dollars paid to football players could in fact equal an entire sport. While a number of people have blamed these cuts on Title IX, the law's supporters argue that the true culprit is football, with its excess of expenses.[132] It would become difficult, if not impossible, to counter such an argument if Legislative Bill 688 took effect.

Male participants in non-revenue sports appear to have little legal recourse against the elimination of their sports. Recently, in *Miami University Wrestling Club v. Miami University*,[133] the Sixth Circuit held that Miami (Ohio) University's decision to eliminate its wrestling team to comply with Title IX did not constitute gender discrimination. Referencing the 1996 OCR Clarification, the Court conceded that "universities and other recipients of federal funds do not have infinite money supplies."[134] Therefore, the court continued, "[i]f a university cannot afford to add sports teams in order to provide equal athletic opportunity for men and women, it may be forced to subtract in order to equalize."[135] Despite OCR's most recent clarification, which called the elimination of teams a "disfavored practice" that is "contrary to the spirit of Title IX,"[136] the economic reality discussed by the *Miami University* court remains. A bill such as Legislative Bill 688 would place further strains on finite athletic budgets and undoubtedly result in the loss of more men's sports programs, with the affected athletes having little or no legal recourse.

States also seem unlikely to rescue athletic departments by increasing their budgets or providing financial assistance. Governments across the nation are facing budget deficits, with many programs being drastically scaled back or eliminated completely.[137] While a state legislature may be willing to pass a bill such as Legislative Bill 688 that will require years to be implemented, it seems likely that many state officials would balk at the immediate distribution of more funds to student-athletes already receiving full scholarships. Immediately following the introduction of the bill, the *Omaha World-Herald* ran an editorial sharply criticizing such a proposal in a time of budget cutbacks across the state:

> How's this for irony: While the state is considering cutting its need-based scholarship programs to help Nebraska's poorest students attend college—while students are borrowing record amounts of money to cover steeply rising tuition costs—the Ne-

---

com/stories/032601/Hus_menswim26.html. (last modified Mar. 26, 2001).

[132] *See, e.g.*, Lopiano, *supra* note 125.

[133] 302 F.3d 608, 613 (6th Cir. 2002).

[134] *Id.*

[135] *Id.*

[136] 2003 CLARIFICATION, *supra* note 90.

[137] The Web site of one political organization claims that forty-three states and the District of Columbia will have budget shortfalls in FY 2003 totaling $27 billion. It is projected that Nebraska alone will lose $220 million between 2001 and 2003. *The State Fiscal Crisis in Nebraska*, *at* http://www.nebraskansforpeace.org/2002/ja02/budget.html.

> braska Legislature and the governor are worried about football players (more than half of them on full-ride scholarships) and whether they've got spending money.[138]

While people across the nation took note of the bill's passage,[139] one can surmise that their responses were muted because of the apparently slim chance that the act would be implemented. The bill has become law, however, and could be implemented, in theory, as soon as next year. Should this happen, a more vocal and hostile reaction against giving full-scholarship athletes additional money could surely be expected, no matter the revenue these athletes help generate for the school.

If approached from a purely economic perspective, Nebraska Legislative Bill 688 appears completely justified in a free-market economy. If football players are largely responsible for generating millions of dollars in revenue, why should they not receive greater compensation than athletes in sports that produce no revenue? The fact remains, though, that student-athletes are students and not professional athletes.[140] The constraints of Title IX will likely keep such a law from ever being implemented, at least to the degree desired by the bill's supporters. Perhaps, then, the bill will succeed in its other purpose: alerting the public, and in particular the NCAA, to the ever-increasing burdens of the student-athlete in "big time" college athletics. Universities must allow student-athletes to be students, by enforcing hourly limitations on athletic participation and devoting larger portions of athletic revenues to student and academic programs. By doing this, universities will eliminate the need for laws such as Legislative Bill 688, which the schools cannot afford and which could possibly turn "first-class amateur programs" into "third-rate professional sports franchises."[141]

— Greg Skidmore

---

[138] *Those Poor Football Players Fight NCAA Rules on their Court; Leave Legislative Arena for State Scholarship Programs*, OMAHA WORLD-HERALD, Feb. 17, 2003, at 6B.

[139] *See, e.g.*, *Nebraska Proposal to Pay College Athletes Stirs Issue*, USA TODAY, Feb. 21, 2003, at 6C.

[140] In the past twenty years, the Olympic Games has wrestled with a similar dilemma, balancing preserving amateurism against the compensating athletes for their unique skills. *See, e.g.*, *Changing the Rules*, N.Y. TIMES, Feb. 21, 1986, at A22. While the International Olympic Committee's decision to allow professional athletes to participate created initial excitement (*e.g.*, the Dream Team of NBA basketball players), a movement has developed to reinstate the amateur rule in order to decrease the commercialism of the Games. *See, e.g.*, Glenn Dickey, *Put the Amateurs Back in the Games*, S.F. CHRON., Aug. 16, 2000, at E3.

[141] Myles Brand, *Welfare of Student-Athletes NCAA's Top Priority*, DEN. POST, Aug. 17, 2003 (on file with author).

UNIVERSAL NATIONAL SERVICE ACT

On January 7, 2003,[1] Representative Charles Rangel (D-N.Y.), the ranking Democrat on the House Ways and Means Committee,[2] introduced the Universal National Service Act of 2003 (the bill),[3] a bill requiring that all young persons in the United Sates, including women, perform a period of military service or a period of civilian service in furtherance of the national defense and homeland security.[4] This bill, like its identical Senate counterpart introduced by Senator Ernest Hollings (D-S.C.),[5] mandates compulsory national service for every citizen and every resident of the United States between the ages of eighteen and twenty-six.[6] The bill's enactment into law is unlikely given military and public opposition.[7] The last action taken on the bill occurred on February 3, 2003 when it was referred to the Subcommittee on Total Force.[8] Nevertheless, the bill is an important symbolic effort to awaken the nation to the current racial composition of America's armed forces and to spur debate over the best way to ensure that our country's military burden is shared equitably.

A law conscripting all young people into national service against their will is not the answer to the racial imbalance in the military. Although the Uniform National Service Act would cure the problem of minority overrepresentation by forcing all young people to enter national service, the solution to a racial imbalance produced by the limited choices available to minority youth should not be to limit the choices of every young person. Instead, the economic prospects of non-minorities that induce them to forego military service should be made available to all young people. At the same time, military and civilian national service should be made more attractive so that more young Americans choose to serve voluntarily. Increased pay and benefits, better service conditions, and greater candor from politicians considering military action may induce more young elites to represent their county. These enhancements will also render pride, rather than a lack of options, as the principal reason for minority service. The men and women of our armed forces who choose to serve should do so under conditions commensurate with the reverence with which we discuss their sacrifice. That sacrifice would be all the more selfless

---

[1] 149 CONG. REC. H57 (daily ed. Jan. 7, 2003) (statement of Rep. Rangel).

[2] The Office of Charles B. Rangel, Committee on Ways & Means, *at* http://rangel.house.gov/ways-means.html.

[3] Universal National Service Act of 2003, H.R. 163, 108th Cong. (2003).

[4] *Id.*

[5] Universal National Service Act of 2003, S. 89, 108th Cong. (2003).

[6] H.R. 163, § 2(a).

[7] *See Washington in Brief*, WASH. POST, Jan. 8, 2003, at A5; CNN, *Poll: Young Not in Step with 30-Plus Crowd, at* http://us.cnn.com/2003/ALLPOLITICS/10/28/elec04.rock.vote.poll/ (stating that 88% of voters under thirty years of age oppose the draft compared with 80% of voters above thirty years of age).

[8] Bill Summary & Status of H.R. 163, *at* http://thomas.loc.gov/cgi-bin/bdquery/z?d108:HR00163:@@@L&summ2=m&.

because no one would be forced to shoulder the responsibility. Therefore, although the Uniform National Service Act seeks a just end, it employs erroneous means. A return to the draft is not the way to redress either the problems faced by minorities that force them into military service or the reluctance of other young Americans to represent their country.

The possibility of war with Iraq brought this debate over military service to the center of the political stage, and Representative Rangel introduced his bill in response to the impending war.[9] He wrote that "if we are going to send our children to war, the governing principle must be that of shared sacrifice . . . . That's why I will ask Congress next week to consider and support legislation I will introduce to resume the military draft."[10] War began two months later.[11] At that time, minorities constituted about 30% of the enlisted force,[12] which is a higher percentage than their representation in the civilian population.[13] The percentage of minority new recruits had reached 37% in 2000, up from 23% in 1973.[14] By contrast, white enlistment has decreased over time and whites serve at percentages much lower than their representation in the civilian population.[15] These figures suggested that the impact of an Iraq war would be felt disproportionately by minorities, and Representative Rangel urged a more equal distribution of the burdens.[16] Representative Rangel admitted

---

[9] Charles B. Rangel, *Bring Back the Draft*, N.Y. TIMES, Dec. 31, 2002, at A19 ("President Bush and his administration have declared a war against terrorism that may soon involve sending thousands of American troops into combat in Iraq.").

[10] *Id.*

[11] *See, e.g.*, Dan Balz & Mike Allen, *'No Outcome But Victory,' Bush Vows; President Pledges Maximum Force and Warns Public of Difficulties*, WASH. POST, Mar. 20, 2003, at A1.

[12] INST. FOR DEF. ANALYSIS, OFFICE OF THE SEC'Y OF DEF., CHANGES IN FORCE COMPOSITION 3 & fig.3, *available at* http://www.economics.osd.mil/force_comp_paper.pdf; *see also* Steven Holmes, *The Nation: For Job and Country; Is This Really an All-Volunteer Army?*, N.Y. TIMES, Apr. 6, 2003, § 4, at 1 (citing statistics showing that whites make up 63% of enlisted military personnel but 70% of the civilian population; blacks make up 22% of enlisted personnel but 12% of the population; and Hispanics make up 9% of enlisted personnel and 13% of the population).

[13] INST. FOR DEF. ANALYSIS, *supra* note 12, at 4 & fig.4. *See also* Holmes, *supra* note 12.

[14] David M. Halbfinger & Steven A. Holmes, *A Nation at War: The Troops; Military Mirrors a Working-Class America*, N.Y. TIMES, Mar. 30, 2003, at A1.

[15] INST. FOR DEF. ANALYSIS, *supra* note 12, at 3–4.

[16] Rangel, *supra* note 9. The war in Iraq required a large ground component supplied by the Army. *See* Bradley Graham, *Iraq Stabilization Impinges on Army Rotation, Rebuilding*, WASH. POST, June 6, 2003, at A21. The overrepresentation of minorities in the armed forces is most pronounced in the Army. Halbfinger & Holmes, *supra* note 14 (showing that Army males are 58% white, 26% black, and 9% Hispanic compared with 70%, 11%, and 14% of the civilian population respectively, and that Army females are 38% white, 46% black, and 9% Hispanic compared with 69%, 14%, and 12% of the civilian population respectively; that Navy males are 62% white, 19% black, and 10% Hispanic, and that Navy females are 50% white, 31% black, and 11% Hispanic; Air Force males are 75% white, 16% black, and 5% Hispanic, and that Air Force females are 62% white, 28% black, and 6% Hispanic; that Marine males are 67% white, 16% black, and 13% Hispanic, and that Marine females are 56% white, 23% black, and 16% Hispanic.).

that he could not know which race of troops would suffer disproportionate casualties, but maintained that his principal contention was that all Americans should sacrifice.[17] Official Department of Defense statistics showing the breakdown of casualties from the current Iraqi conflict were not available at the publication of this Article.

The Universal National Service Act would require all United States residents between the ages of eighteen and twenty-six to perform a two-year period of national service.[18] The period of national service would be performed either as a member of an active or reserve component of the uniformed services or in a civilian capacity that promotes the national defense, including national or community service and homeland security, as determined by the President.[19] The President would also be responsible for determining the number of persons whose service would be in the military and for selecting the individuals to be inducted for military service.[20]

The bill provides that the President may extend the period of military service with the consent of the serviceperson, for the purpose of furnishing medical care for maladies suffered in the line of duty, or "for the purpose of requiring the serviceperson to compensate for any time lost to training for any cause."[21] The bill further provides for early termination of the period of national service upon the "voluntary enlistment and active service of the person in an active or reserve component of the uniformed services for a period of at least two years," the "admission and service of the person as a cadet or midshipman" in one of the military academies, or the "enrollment and service of the person in an officer candidate program, if the person has signed an agreement to accept a Reserve commission in the appropriate service with an obligation to serve."[22] Deferments could be received by both high school students and by those who suffer extreme hardship or physical or mental disability.[23] The President may postpone or suspend the induction of persons for military service as necessary to limit the number of persons receiving basic military training and education to the maximum number that can be trained adequately.[24] The bill further provides that people selected for military serv-

---

[17] Charles B. Rangel, *Race of Front-Line Troops Isn't Real Issue*, USA TODAY, Jan. 27, 2003, at 16A.

[18] Universal National Service Act of 2003, H.R. 163, 108th Cong. §§ 2(a), 3(a) (2003).

[19] *Id.* § 2(b).

[20] *Id.* § 2(d). In order to assist implementation of the Act's provisions, the bill empowers the President to "prescribe such regulations as are necessary to carry out the Act." *Id.* § 4(a). The Act specifies that such regulations include the types of civilian service satisfying the national service obligation, standards for performing that civilian service, the manner in which persons shall be inducted and notified of their selection for induction, standards for determining conscientious objection exemptions, and standards for compensation and benefits. *Id.* § 4(b).

[21] *Id.* § 3(b).

[22] *Id.* § 3(c).

[23] *Id.* § 6(a)–(b).

[24] *Id.* § 6(c).

ice who claim conscientious objector status because their religious training and belief should exempt them from the combatant training included as part of military service, and whose claims are sustained, shall participate in military service that does not include any combatant training component.[25] Finally, the bill provides that it shall apply to both women and men.[26]

Representative Rangel believed that America would be less likely to rush into a war if the composition of the armed forces reflected the composition of American society.[27] He opposed the United States entering a war with Iraq,[28] and introduced the Uniform National Service Act to "bring a greater appreciation of the consequences of decisions to go to war."[29] Local response to Representative Rangel's proposal legitimized his concerns. For instance, an African American New Yorker praised Representative Rangel for voicing his constituents' point of view, and asked, "Are our young people expendable? We are not permitted to share the wealth, so why must we carry the burden?"[30] Representative Rangel's proposal resonated beyond his own district. A Massachusetts resident wrote, "If the Bush administration and Congress are determined to start a war, a military draft should be reinstated to ensure that the burden of fighting is equitably distributed throughout society."[31] These Americans echoed Representa-

---

[25] *Id.* § 8(a). Any person whose claim is sustained may be transferred to a national service program to complete his or her national service obligation. *Id.* § 8(b).

[26] *Id.* § 10.

[27] *See* Rangel, *supra* note 9 ("I believe that if those calling for war knew that their children were likely to be required to serve—and to be placed in harm's way—there would be more caution and a greater willingness to work with the international community in dealing with Iraq."); *see also* John N. Rippey, Letter to the Editor, *On the Tactics of War, and the Searing Pain*, N.Y. TIMES, Apr. 1, 2003, at A18 ("[A]n immediate worry is the relative freedom from political considerations that a draft-free Army gives a president intent on a foreign policy of pre-emptive wars. It is hard to believe that an American president would make a highly debatable invasion of a Middle Eastern country if the ranks of the Army were filled out not only with volunteers but also with draftees.").

[28] *See* Rangel, *supra* note 9.

[29] *Id.*

[30] Lucia Jack, Letter to the Editor, *A Military Draft, For Equality's Sake?*, N.Y. TIMES, Jan. 1, 2003, at A14.

[31] Julie E. Dunn, Letter to the Editor, *Share the Burden*, N.Y. TIMES, Jan. 7, 2003, at A18. *See also* Jack Dillon, Letter to the Editor, *Reinstate the Draft*, BOSTON GLOBE, Jan. 16, 2003, at A10 ("US [sic] Representative Charles Rangel is right. Bring back the draft, only this time no student exemptions."); Bruce J. Deegan, Letter to the Editor, *A New Draft Must Be Fair*, HARTFORD COURANT, Jan. 13, 2003, at A8 ("How many students residing in the more affluent towns will be receiving recruiting calls? I suspect not very many."); Editorial, *Bring Back the Draft*, DENVER POST, Jan. 6, 2003, at B07 ("Resuming the draft, preferably with a shorter period of active duty and no deferments, would revive the idea that the country's defense is everybody's responsibility. And the possibility of their sons being in the front lines might make some of the power elite less eager to cavalierly go to war."); Clarence Page, Editorial, *Feeling a Draft Coming On; Maybe We Should Bring Back the Draft So That We'll Start Paying Attention to the Hawks Who Are Having Too Easy of a Time Getting What They Want*, CHI. TRIB., Jan. 1, 2003, at C23 ("It is easy to understand why Rangel would think the return of the draft would make his fellow members of Congress think a little longer and harder before they send our nation's sons and our

tive Rangel's concern that "service in our armed forces is no longer a common experience."[32] Representative Rangel introduced his bill to ensure that all young people experienced equally the burdens of an Iraq war.[33]

Many Americans who oppose the war, however, do not believe that reinstating the draft is the answer. One Virginia resident wrote that Representative Rangel's proposal was "misguided and frightening" and that "a draft of our youth to fight a war that many have questioned is morally reprehensible."[34] Another writer added that conscription was not the proper means to the noble end of an equitable distribution of the military burden.[35] He argued that "the only effective way to share the cost equitably is to pay military personnel those salaries that make military service attractive . . . . [A] draft that forces young people into service at salaries insufficient to attract them voluntarily concentrates the cost of military involvement unfairly on those young people who are drafted."[36] These Americans voiced the traditional anti-war opposition to the draft.[37] They demonstrate that one of the more remarkable facets of Representative Rangel's proposal is that it represents a dramatic shift in attitudes toward the draft. Now, anti-war liberals introduce legislation to reinstate conscription while military officials shun the opportunity to gain new recruits.[38] These opinions also suggest that Representative Rangel's proposal creates as much division among anti-war Americans as it creates

---

daughters into harm's way.").

[32] Rangel, *supra* note 9 ("A disproportionate number of the poor and members of minority groups make up the enlisted ranks of the military, while the most privileged Americans are underrepresented or absent.").

[33] *See id.* ("Those who would lead us into war have the obligation to support an all-out mobilization of Americans for the war effort, including mandatory national service that asks something of us all.").

[34] Greg McCracken, Letter to the Editor, *A Military Draft, For Equality's Sake?*, N.Y. TIMES, Jan. 1, 2003, at A14.

[35] Donald J. Boudreaux, Letter to the Editor, *A Military Draft, For Equality's Sake?*, N.Y. TIMES, Jan. 1, 2003, at A14.

[36] *Id.*

[37] *See* Bernice B. Rosenberg, *How the Young See the Military*, N.Y. TIMES, May 28, 2003, at A22. ("As a high school teacher during the Vietnam War and the mother of a high school student and a college student during that period, I find it interesting that today's students are so gung-ho about the military. Could it be that since there is no longer a military draft that could upset their lives, today's students do not feel threatened? Perhaps they should have a serious conversation with their parents, who may have a different perspective"); Clyde Haberman, *Draft Talk, But Source is Antiwar*, N.Y. TIMES, Jan. 3, 2003, at B1 ("Time was when being antiwar meant that you probably also hated compulsory military service.").

[38] *See, e.g.*, Holmes, *supra* note 12 ("[S]upporters of the all volunteer force say, the military is . . . more professional, better motivated and more stable when soldiers, sailors, pilots and others stay in for longer stints. They point to performances in the Persian Gulf war, the Afghanistan campaign and now Iraq. And they shudder at returning to the often-troubled conscripted military of the Vietnam era, just to make a point about equity that not everyone feels could even be remedied."); *Washington in Brief*, *supra* note 7 ("Rumsfeld said troops from Vietnam War conscription added 'no value, no advantage, really, to the United States armed services . . . because the churning that took place, it took an enormous amount of effort in terms of training, and then they were gone.'").

between anti-war and pro-war Americans. Representative Rangel is correct to seek greater consideration by Congress of decisions authorizing the use of force, and ensuring that middle- and upper-class youth serve in the military may well accomplish this goal. However, increasing the attractiveness of military service so that a greater cross-section of America's youth serve voluntarily may accomplish this same goal while maintaining a united anti-war front.

Representative Rangel's chances of seeing his bill become law are dim in light of the lack of support from either the military or other members of Congress.[39] The only action taken on the bill by the House Committee on Armed Services was to request executive comment from the Department of Defense and to refer the bill to the Subcommittee on Total Force.[40] The only congressional commentary took place the day after it was introduced when Representative Pete Stark (D-Cal.) announced his support for the bill.[41] Representative Stark emphasized the fact that few legislators had children in active military service and that a return to the draft was necessary to make a vote for war "more real" for Congress.[42] Experts, however, cite the success of recent politicians despite their lack of any significant military connection as one of the reasons for the bill's dismal chance of passage.[43] The bill is currently stalled, regardless of whether the reason is antiwar protestors, military officials, or leading politicians.[44] "There is no serious discussion of [bringing back the draft]," said Representative John A. Boehner (R-Ohio), Chairman of the Committee on Education and the Workforce.[45]

Supporters of conscription argue that the all-volunteer force (AVF) places the burden of war disproportionately on minorities. They claim that minorities bearing a burden of defense greater than their proportion in the population signifies unfairness.[46] In the event of war, blacks will proportionally suffer the greatest casualties because they will comprise such a large percentage of front-line soldiers.[47] Critics of the AVF sug-

---

[39] *See* Haberman, *supra* note 37.

[40] Bill Summary & Status of H.R. 163, *at* http://thomas.loc.gov/cgi-bin/bdquery/z?d108:HR00163:@@@L&summ2=m&.

[41] *See* 149 CONG. REC. E40 (daily ed. Jan. 8, 2003) (statement of Rep. Stark).

[42] *See id.*

[43] *See* Haberman, *supra* note 30.

[44] *See* Carl Hulse, *Threats and Responses: The Draft; A New Tactic Against War: Renew Talk about Draft*, N.Y. TIMES, Feb. 9, 2003, § 1, at 17.

[45] *Id.*

[46] *See* Robert K. Fullinwider, *The All-Volunteer Force and Racial Balance*, *in* CONSCRIPTS AND VOLUNTEERS: MILITARY REQUIREMENTS, SOCIAL JUSTICE, AND THE ALL-VOLUNTEER FORCE 179 (Robert K. Fullinwider ed., 1983) [hereinafter CONSCRIPTS AND VOLUNTEERS].

[47] *See* David R. Segal, *Military Organization and Personnel Accession: What Changed with the AVF . . . and What Didn't*, *in* CONSCRIPTS AND VOLUNTEERS, *supra* note 46, at 11 ("In our combat divisions, it is not uncommon to find units that are 50 percent black. (These units, in the event of combat, would sustain the greatest number of casualties and fatalities.)").

gest that black Americans should not die in wars that white America chooses to fight.[48] Furthermore, black and minority casualty rates of 30% to 40% might generate opposition to the war movement by precipitating a backlash in the black community.[49] Supporters of Representative Rangel's bill revive the argument that overrepresentation of blacks in the armed services is a latent problem that war will graphically bring to the forefront of the mind of the American public.[50]

Proponents of conscription find a war yielding high black and minority casualties especially troubling given the economic plight of minorities. Blacks and other minorities are historically over-represented in the armed forces because the military represents an avenue of social advancement that they are otherwise denied in American society.[51] Scholar Robert K. Fullinwider notes that "for black teenagers, facing the highest unemployment rates in our economy, the Army (and other branches of service) provides opportunities for job training and social and economic mobility."[52] These unemployment rates suggest to many that poor blacks and other underprivileged members of our society "choose" military service because they lack the choices of white America.[53] Therefore, high

---

[48] Fullinwider, *supra* note 46, at 181 ("If there were a war, large numbers of blacks might die, and in any case they would die out of proportion to their share of the general population. This prospect deeply troubles critics of the AVF .... The image of blacks pulling the dirty load for white America is compelling for many and surely must evoke at least a question from all of us about the racial imbalance in the Army.").

[49] *See id.* at 185 ("In the emotional climate aroused by combat deaths, attitudes and feelings are subject to volatile shifts and may suddenly crystallize into adamant opposition to the military action, especially under the stimulation of concentrated and graphic television coverage of a highly telegenic issue.").

[50] *See, e.g.*, Charles C. Moskos, Jr., *Serving in the Ranks: Citizenship and the All-Volunteer Force (Summary)*, *in* REGISTRATION AND THE DRAFT: PROCEEDINGS OF THE HOOVER-ROCHESTER CONFERENCE ON THE ALL-VOLUNTEER FORCE 129, 134 (Martin Anderson ed., 1982) ("[A] U.S. involvement in a Middle East war would involve a disproportionate impact on black military personnel and that the resulting overproportion of casualties they would suffer would be 'outrageous.' ... To hide our heads in the sand and think such an eventuality is not going to have political consequences is, I think, deceitful. This *is* going to become an issue.").

[51] *See id.* at 182 ("It is true that the rise in the number of blacks reflects not only the large increase in the number of blacks eligible for military service, but also the unprecedently high unemployment rates that characterize our society's minority youth community.").

[52] ROBERT K. FULLINWIDER, THE AVF AND RACIAL IMBALANCE 3 (Ctr. for Philosophy and Pub. Policy, Univ. of Md., Working Paper, 1981). The rate of unemployment for young blacks today is typically more than double the rate for young whites (19.8% versus 8.9% in 2001; 22.4% versus 10.7% in 2002). *See* U.S. DEP'T OF LABOR, BUREAU OF LABOR STATISTICS, EMPLOYMENT AND EARNINGS, TABLE 24: UNEMPLOYED PERSONS BY MARITAL STATUS, RACE, AGE, AND SEX, *available at* http://www.bls.gov/news.release/youth.t02.htm.

[53] *See, e.g.*, Fullinwider, *supra* note 46, at 179 ("The fatal defect in the All-Volunteer Force is that time has proven the current system is not a 'volunteer' system at all .... [The AVF is a] system in which those who have the least in our society are offered the opportunity to be trained to risk all in exchange for the very thing they have been denied by the society they are asked to defend, a job.") (quoting Congressman John Cavanaugh, SYNERGIST, Winter 1980, at 14).

black casualty rates suffered by the "all-volunteer" force may be unacceptable to black communities.

Conscription's proponents also argue that, in addition to the inequity of minorities who lack economic opportunities shouldering the burden of war, bringing more privileged youth into the armed forces would improve relations between the military and the public.[54] Professor Charles Moskos, a longtime supporter of national service, argues that the military must represent every stratum of society.[55] For example, Chelsea Clinton's or Jenna Bush's enlisting in the Army would signal that the American public was fully behind the military's war effort.[56] Moskos contends that support for the Vietnam War ended when elites began obtaining draft deferments because "nobody'll accept casualties unless the elite are willing to put their own children's lives on the line."[57] Furthermore, the declining percentage of privileged youth serving in the armed forces means that a warrior caste is developing.[58] Some scholars worry that "a cultural and political gap could open up between civilian and martial societies,"[59] resulting in less effective civilian control over the military.[60] Professor John Allen Williams, a retired navy captain, worries that a military that identifies itself as different from the civilian society it serves could conceivably act on its own values in opposition to civilian leaders.[61] Moskos argues that a draft will reduce this disconnect between the military and American society.[62] Of course, conscription is not the sole mechanism by which to expand the service of elites. Young elites with political ambitions may become more inclined to serve simply by making military service more attractive.

Proponents consider Uniform National Service a feasible way to ensure that the burdens of military service are equally shared without compromising the military's effectiveness or destroying the choices of privileged youth. Moskos defends the Universal National Service Act, arguing that drafting graduates of leading universities into the armed forces would enhance conscription's legitimacy and military service's fairness.[63] Moreover, he believes that a draft would make the military "much more effective."[64] Many military jobs could be filled by short-term draftees who serve eighteen months of active-duty followed by a reserve obliga-

---

[54] *See* Segal, *supra* note 47, at 8.
[55] *See* Halbfinger & Holmes, *supra* note 14.
[56] *See id.*
[57] *Id.*
[58] *See id.*
[59] *Id.* (quoting Dr. Richard H. Kohn, professor of military history at the University of North Carolina).
[60] *See id.*
[61] *See id.*
[62] *See id.*
[63] *See* Charles Moskos, *A New Kind of Draft for the 21st Century*, BOSTON GLOBE, Feb. 9, 2003, at D12.
[64] *Id.*

tion.[65] These jobs range from logistics to administration to security work.[66] Short-term soldiers could fulfill peacekeeping tours of duty in deployments such as in Bosnia or Kosovo that have a tremendous need for low-skill security manpower and typically last only six months.[67] Additionally, conscription would free up money to give pay raises to career soldiers by allowing the military to decrease the pay levels of the lowest ranking enlisted personnel because there would be no need to entice volunteers.[68] The pay ratio between a master sergeant and a private in the draft era was seven to one; today it is less than three to one.[69] "Restoring something like the old balance is the best way to upgrade retention in hard-to-fill skills and leadership positions."[70] Therefore, Moskos asserts that restoring conscription is both equitably and militarily desirable.

Moskos also recognizes the need for choice in a non-voluntary national service system.[71] He argues that a new draft must be a three-tiered system.[72] The first tier would comprise both the active-duty and reserve components of the military, the second tier would consist of homeland security personnel, and the third tier would be civilian programs such as President Clinton's AmeriCorps or President Bush's proposed Freedom Corps.[73] Draftees would receive higher college aid awards and shorter terms of duty for accepting more dangerous assignments.[74] Drafting college graduates and sending soldiers to college after their service would help equalize both the burden of military service and the benefits of education in our society.[75] Moreover, "[i]f serving one's country becomes common among privileged youth, future leaders in civilian life will have had a formative citizenship experience. This can only be to the advantage of the armed forces and the nation."[76]

Opponents of conscription have argued, however, that a draft is not the solution to the racial and economic divide in our nation's military,

---

[65] *See id.*

[66] *See id.*

[67] *See id.*

[68] *See id.*

[69] *See id.*

[70] *Id.*

[71] *See id.*

[72] *See id.*

[73] *See id.* Moskos adds that homeland security conscripts could "serve as airport security, border patrols, guards for nuclear power plants and public facilities, customs and immigration agents, more Coast Guard personnel to inspect ships, and so on." *Id.*

[74] *See id.*

[75] *See id.*

[76] *Id.* For alternative draft proposals, see CHARLES C. MOSKOS, A CALL TO CIVIC SERVICE: NATIONAL SERVICE FOR COUNTRY AND COMMUNITY 136–37 (1988) (suggesting that alternative drafts could be either a "minimal active force draft" that conscripted only the number of men necessary to make up the difference between manpower requirements and the number of volunteers, or a version of universal military training (UMT) under which virtually all eligible men would receive six months of military training followed by some kind of reserve assignment).

and that universal national service would contribute neither to equity nor military efficiency. Scholar Richard Cooper writes that the all-volunteer force surpasses a conscripted army in terms of national security, fairness, and efficiency.[77] These assertions mirror the Pentagon's contention that "the all-volunteer force has served the nation for more than a quarter-century, providing a military that is experienced, smart, disciplined, and representative of America."[78] Furthermore, Cooper believes that universal national service would create new problems of equity and efficiency.[79] Allocating manpower between military and nonmilitary service spawns equity concerns, universal national service involves enormous expense, and draftees fulfilling national service tasks that require few skills may displace currently employed, less educated workers.[80] Therefore, universal national service would undermine its desired goals.

More specifically, detractors contend that a draft would not necessarily distribute the burdens of military service equitably. Fullinwider suggests that only by high induction calls and raising volunteering standards to unreasonably high levels will a draft reduce the proportion of black accessions.[81] Past studies indicate that eliminating minority over-representation would require drafting more than 100,000 people.[82] Moreover, conscription did not produce equal burden sharing in the past,[83] and universal national service could prove equally susceptible to the benefits of privilege.[84] Scholar David Segal writes that "[t]he end of the draft was

---

[77] *See* Richard V. L. Cooper, *Military Manpower Procurement: Equity, Efficiency, and National Security (Summary)*, *in* REGISTRATION AND THE DRAFT: PROCEEDINGS OF THE HOOVER-ROCHESTER CONFERENCE ON THE ALL-VOLUNTEER FORCE 126–27 (Martin Anderson ed., 1979).

[78] Hulse, *supra* note 44 (quoting a recently circulated Pentagon position paper).

[79] *See* Cooper, *supra* note 77, at 123.

[80] *See id.*

[81] *See* Fullinwider, *supra* note 46, at 183; Kenneth J. Coffey, *If The Draft Is Restored: Uncertainties, Not Solutions*, *in* CONSCRIPTS AND VOLUNTEERS, *supra* note 46, at 68. Coffey also points out that drafting large numbers of privileged youth and suppressing black volunteering by raising recruiting standards would likely reduce the level of public support for any such draft. *Id.* at 65–66.

[82] *See* Segal, *supra* note 47, at 13; *see also* Moskos, *supra* note 63 ("Even with larger active duty and reserve components, even with homeland security, even with expanded forms of civilian service, probably not much more than half of the eligible youth cohort would be needed.").

[83] *See id.* at 11–12 ("Nor is there a guarantee of racial representation among those drafted . . . . Conscription has never produced a truly socially representative armed forces in the United States."); *see also* Cooper, *supra* note 77, at 126 ("Many of the inequities that led to the poor bearing a disproportionate share of the burden can be attributed to the specifics of the post-war, pre-lottery draft. Although not as blatant as the draft during World War I, when individuals were classified and inducted according to their 'value to society,' there were still numerous ways of avoiding induction during the postwar draft, ways that served to benefit largely middle- and upper-class white youth.").

[84] *See* MOSKOS, *supra* note 76, at 136 (suggesting that many will attempt to avoid induction under a compulsory system, and that decisions about exemptions would inevitably give rise to charges of inequities and favoritism); Cooper, *supra* note 77, at 126 ("As long as there is an incentive to avoid induction, people will avoid it, and the ones who will be successful are the ones who have the most to gain.").

followed by an increase in female and racial and ethnic minority personnel in the armed forces. It is unlikely that a return to the draft alone would greatly alter this composition."[85] Universal national service would fail, therefore, as a practical proposal to address racial divisions in the military.

Conscription's opponents also argue that altering the racial composition of the military would not be desirable even if universal national service decreased minority overrepresentation. They assert that minorities view the military as an avenue of social advancement and that a draft reducing the number of military positions open to minorities forecloses that avenue.[86] Quite simply, "a return to the draft means taking unwilling whites instead of willing blacks."[87] A proposal designed to improve the position of minorities in society would have the unintended effect of harming those minorities:

> The young blacks whom we worried were being "victimized" by the all-volunteer policy because they were "forced" to choose between service and unemployment are now reduced to one choice: unemployment. Under those circumstances, they might be unable to appreciate how they had been relieved of victimization! If our concern about the all-volunteer policy is injustice to individual blacks who are "coerced" into service, then a more reasonable solution may be one which improves the conditions of their service rather than one which worsens further their limited set of opportunities.[88]

A proposal improving the conditions of military service would have the added benefit of attracting a broader segment of society without forcing anyone to serve. Such a proposal would preserve minority opportunities in the military while at the same time facilitating equal burden sharing of the national defense.

Critics favoring a solution to the racial divide in the military that focuses on the quality rather than the quantity of service also debate the relevance of the divide. For instance, Fullinwider questions whether minority overrepresentation in the military is necessarily unjust.[89] Fullinwider acknowledges the undesirable aspects of military service such as subordination and regimentation, but asserts that many types of civilian jobs share these qualities.[90] He concludes that "[t]he critics must make very much plainer why there is *no* level of compensation which makes

---

[85] Segal, *supra* note 47, at 16.
[86] *See id.* at 11.
[87] FULLINWIDER, *supra* note 52, at 3.
[88] Fullinwider, *supra* note 46, at 183–84.
[89] *See id.* at 179.
[90] *See id.* at 180–81

acceptable the risks that blacks assume in military service, or why, if there is a level, the all-volunteer policy fails to achieve it."[91] Indeed, the overrepresentation of minorities in the military during the post-World War II draft[92] led Richard Cooper to assert that "the main difference between a volunteer force and a conscripted one today is not in who serves, but rather the fact that, in paying a 'fair' wage to youth, the All-Volunteer Force has not discriminated against the poor the way that the draft did."[93]

These detractors also contend that viewing the racial divide as a problem portrays minority youth as so alienated from American society that they choose service solely as an economic choice and that they neither possess nor ought to possess any commitment to American institutions and values.[94] Critics concede that many proponents of universal national service undoubtedly do not share this conception of minority youth, but then ask that "if . . . we view the current military as composed of American youth with American values, what further interest could we have in 'representativeness'?"[95] Some advocates of the all-volunteer force believe that these considerations necessitate a high threshold of proof before deeming overrepresentation of minorities in the military problematic.[96] Equitable distribution of military service is surely a problem requiring analysis, but the important point is that policymakers should consider carefully the solutions necessary to correct numerical discrepancies.

Representative Rangel's proposal is a bad law with good intentions. He correctly seeks to redress the disproportionate burden of our national defense currently borne by blacks and other minorities. The overrepresentation of minorities in the armed forces needs to be rectified because minorities and non-minorities should face the same options when deciding whether to represent our country. Minorities serve with honor, but they should be able to take advantage of the opportunities offered by the military free from economic constraints. America should not countenance a system that forces underprivileged youths to make the ultimate sacrifice in order to achieve social advancement and that allows the more privileged to enjoy all the benefits of our democracy without an equal sacrifice. A different system that forces everyone to make that sacrifice is, however, no better. Our country should strive for a system that provides better opportunities to those currently underprivileged so that their entrance into military service is a genuine choice. America should seek more responsible political leadership so that youths joining the military

---

[91] *Id.* at 181.

[92] *See* Cooper, *supra* note 77, at 126.

[93] *Id.*

[94] *See* Fullinwider, *supra* note 46, at 184.

[95] *Id.* Fullinwider continues, "Why should we be concerned that the military *mirror* the social, educational, racial, economic, religious, regional, and other demographic patterns found in society as a whole? Simple demographic representation seems to have no value in itself." *Id.*

[96] *See id.* at 185–86.

will know that they will not be asked to give their lives for unpopular and unnecessary wars. Our society should enhance our treatment of both military servicepeople and veterans so that their status is commensurate with both the sacrifice that they are asked to make and with the rhetoric with which our citizens are so fond of embracing them. Increased pay and benefits, improved service conditions, and greater candor from politicians will begin to accomplish this objective. These reforms, and not a return to the draft, are better ways of ensuring that the burdens of war are shared more equally because they will induce non-minorities to serve voluntarily and will guarantee that minority service is truly voluntary. Voluntary sacrifice on the part of all Americans will accomplish Representative Rangel's goal of spreading the risks of war across all of American society, the principal goal sought when he introduced the Universal National Service Act of 2003.

—*Ben Schiffrin*