Arkansas Law Review

Volume 76 | Number 1 Article 1

April 2023

Arkansas Law Review - Volume 76 Issue 1

Journal Editors

ARKANSAS LAW REVIEW

VOLUME 76 NUMBER 1

Articles

Comments

The *Arkansas Law Review* (ISSN 0004-1831) is published quarterly at the University of Arkansas School of Law, Fayetteville, Arkansas 72701-1201, by the Arkansas Law Review, Inc., and the School of Law of the University of Arkansas. Printed at Joe Christensen, Inc., Lincoln, Nebraska 68521. Periodicals postage paid in Fayetteville, Arkansas, and additional mailing offices. The *Arkansas Law Review* is a member of the National Conference of Law Reviews.

Arkansas Law Review
University of Arkansas School of Law
Leflar Law Center, Waterman Hall
Corner of Maple and Garland Avenues
Fayetteville, Arkansas 72701-1201

Telephone: (479) 575-5610

POSTMASTER. Send address changes to the Managing Editor, *Arkansas Law Review*, University of Arkansas School of Law, 1045 W. Maple St., Fayetteville, Arkansas 72701-1201.

ARKANSAS LAW REVIEW

UNIVERSITY OF ARKANSAS
SCHOOL OF LAW
FACULTY AND PROFESSIONAL STAFF

Administration

CYNTHIA E. NANCE, B.S., M.A., J.D., *Dean and Nathan G. Gordon Professor of Law*
TIFFANY R. MURPHY, B.A., J.D., *Associate Dean for Academic Affairs and Professor of Law*
JAMES K. MILLER, B.S., B.A., J.D., *Senior Associate Dean for Students*
JILL WIEBER LENS, B.A., J.D., *Associate Dean for Research and Faculty Development and Robert A. Leflar Professor of Law*
ERIN FELLER, B.B.A., M.B.A., *Interim Director of Development*
TRACY DEFFEBAUGH, B.A., *Director of Admissions*
LYNN STEWART, B.S., B.S., C.P.A., M.B.A., *Director of Budget, Human Resources, and Building Services*
SPENCER BOWLING, B.A., J.D., *Assistant Dean of Admissions*

Current Law Faculty

ALENA ALLEN, B.A., J.D., *Professor of Law*
HOWARD W. BRILL, A.B., J.D., LL.M., *Vincent Foster University Professor of Legal Ethics & Professional Responsibility*
BLAIR D. BULLOCK, B.S., J.D., PH.D., *Assistant Professor of Law*
CARL J. CIRCO, B.A., J.D., *Ben J. Altheimer Professor of Legal Advocacy*
STEPHEN CLOWNEY, A.B., J.D., *Professor of Law*
UCHE U. EWELUKWA, DIP. L., S.J.D., LL.B., B.L., LL.M., LL.M., *E.J. Ball Professor of Law*
SHARON E. FOSTER, B.A., J.D., LL.M., PH.D., *Sidney Parker Davis, Jr. Professor of Law*
WILLIAM E. FOSTER, B.S., J.D., LL.M., *Arkansas Bar Foundation Associate Professor of Law*
CAROL R. GOFORTH, B.A., J.D., *University Professor and Clayton N. Little University Professor of Law*
SARA R. GOSMAN, B.S., J.D., LL.M., *Associate Professor of Law*
CALEB N. GRIFFIN, B.A., J.D., *Assistant Professor of Law*
AMANDA HURST, B.A., J.D., *Assistant Professor of Law*
CHRISTOPHER R. KELLEY, B.A., J.D., LL.M., *Associate Professor of Law*
ANN M. KILLENBECK, B.A., M.A., M.ED., J.D., PH.D., *Professor of Law*
MARK R. KILLENBECK, A.B., J.D., PH.D., *Wylie H. Davis Distinguished Professor of Law*
AMELIA MCGOWAN, B.A., M.A., J.D., *Assistant Clinical Professor and Director of the Immigration Clinic*
ASHLEY U. MENENDEZ, B.A., J.D., *Director of Externships and Assistant Professor of Practice*
MARGARET E. SOVA MCCABE , B.A., J.D., *Senior Advisor for Strategic Projects and Professor of Law*
ALEX NUNN, B.S., J.D., PH.D., *Assistant Professor of Law*
DANIEL RICE, B.A., J.D., *Professor of Law*
SUSAN A. SCHNEIDER, B.A., J.D., LL.M., *Director of the LL.M. Program in Agricultural & Food Law and William H. Enfield Professor of Law*
ANNIE B. SMITH, B.A., J.D., *Director of Civil Litigation & Advoacy, Human Trafficking Clinics and Associate Professor of Law*

The School of Law is a member of the Association of American Law Schools and is accredited by the American Bar Association.

(American Bar Association Council of the Section of Legal Education and Admissions to the Bar, American Bar Association, 321 N. Clark Street, 21st Floor, Chicago, IL 60654, 312.988.6738, legaled@americanbar.org)

ARKANSAS LAW REVIEW

Subscriptions and Claims: Effective volume 56, issue 1, the *Arkansas Law Review* will be published quarterly for $25.00 per year ($30.00 per year for foreign delivery). Domestic claims for nonreceipt of issues should be made within 90 days of the month of publication, overseas claims within 180 days; thereafter, the regular back issue rate will be charged for replacement. Address all subscription and claim correspondence to the Managing Editor of the *Arkansas Law Review*.

Single and Back Issues: Single current issues will be available for $9.00 (plus $1.00 postage and handling) from the Managing Editor of the *Arkansas Law Review*. Back issues may be purchased from William S. Hein & Co., Inc., 1285 Main Street, Buffalo, New York 14209- 1987.

Binding: Binding is available at $31.00 per volume (plus $4.00 postage and handling; decorative gold lines available for an additional $4.00 per volume) from the Managing Editor of the *Arkansas Law Review*.

Manuscripts: The editors of the *Arkansas Law Review* encourage the submission of unsolicited articles, comments, essays, and reviews. Manuscripts should be sent to the attention of the Executive Editor at lawrev@uark.edu. Authors should include a current curriculum vitae.

The *Arkansas Law Review* also accepts submissions through its Scholastica academic publishing accounts.

Exclusive Submissions: The *Arkansas Law Review* will be accepting exclusive submissions which will be considered on an expedited basis. Publication decisions will be made within one week of the date of submission. By submitting exclusively to us, authors agree to accept a publication offer should the *Arkansas Law Review* extend one. Authors should submit their manuscript to the attention of the Managing Editor of the *Arkansas Law Review* at arkansaslawreview@gmail.com with "Exclusive Submission" in the subject line.

For 2023, we will actively consider articles from February 1, 2023 to April 30, 2023.

Form: Please cite this issue of the *Arkansas Law Review* as 76 ARK. L. REV. ___ (2023). Citations in the *Arkansas Law Review* conform to *The Bluebook: A Uniform System of Citation* (21st ed. 2020).

We proudly present Volume 76, Issue 1 of the Arkansas Law Review *for the benefit of all who learn and advance the law, whether judge, advocate, professor, or student. We have carefully developed these materials to elicit informed discussions and provide intellectual and practical assistance to members of the legal community.*

Arkansas Law Review Editorial Board
2022-2023

CONSTITUTIONAL LAW AND TAX EXPENDITURES: A PRELUDE

Johnny Rex Buckles*

INTRODUCTION

"A little learning is a dang'rous thing," admonished Pope.[1] Judges who pen legal opinions drawing on tax expenditure theory should heed the neoclassical bard. Armed with the modest yet obligatory exposure to the concept of tax expenditures presented in the basic federal income tax course in law school, many judges indeed possess enough learning to be dangerous. The thesis of this Article is that tax expenditure theory must be applied with a skillful, critical, and cautious appreciation for nuance in constitutional cases. This conclusion holds even under the assumption that tax expenditure budgeting is a useful tool of fiscal analysis. For several reasons, features of tax expenditure analysis apply uneasily in constitutional adjudication.

This thesis is far from obvious, primarily because tax expenditure theory reigns from a lofty, storied throne in national tax policy.[2] Tax expenditure theory is largely grounded in the influential work of the late Stanley Surrey, an accomplished and prolific Harvard law professor who served as Assistant Secretary of the Treasury for Tax Policy.[3] Surrey famously championed the

---

* Mike and Teresa Baker College Professor of Law, University of Houston Law Center. I thank the University of Houston Law Center for supporting the research of this paper. I thank the participants in the 24th Annual Federalist Society Faculty Conference who offered helpful comments on this paper. I also thank my wife, Tami Buckles, for her constant support.

1. ALEXANDER POPE, *An Essay on Criticism*, *in* ALEXANDAR POPE: THE MAJOR WORKS 17, 24 (Pat Rogers ed., Oxford Univ. Press 2006) (emphasis omitted).

2. *See* Martin J. McMahon Jr., *Taxing Tax Expenditures?*, 2011 TAX NOTES (SPECIAL REPORT) 775, 776 (describing tax expenditure theory as "enshrined into law").

3. *See id.* at 775 ("Stanley Surrey introduced tax expenditure analysis to U.S. tax and budget policy in the late 1960s."); *see also* Stanley S. Surrey, *Tax Incentives as a Device for Implementing Government Policy: A Comparison with Direct Government Expenditures*, 83 HARV. L. REV. 705, 705 (1970) [hereinafter Surrey, *Tax Incentives*].

concept of tax expenditures, catapulted tax expenditure theory to academic prominence, and successfully promoted tax expenditure budgeting as a mooring of fiscal stewardship.[4] Surrey conceptualized tax expenditures as "those special provisions of the federal income tax system which represent government expenditures made through that system to achieve various social and economic objectives."[5] Tax expenditures take many forms—credits, deductions, exclusions, deferrals, and special rates.[6] But whatever their form, their effect is to reduce tax liabilities relative to the tax that would be due in the absence of the provisions in question.[7] Surrey pronounced this reduction in a taxpayer's liability the equivalent of a subsidy to the taxpayer from the government.[8]

The logic of tax expenditure theory is, in Sherlockian tongue, elementary.[9] Had the government collected taxes under a system that omitted the special credit, deduction, or other provision, the government would have collected more revenue. Having thus raised more tax revenue, the government then could transfer to the taxpayer a monetary amount equal to the reduction in tax liability enjoyed by the taxpayer in the system that features the credit, deduction, or other special provision. Under the hypothetically enhanced tax-and-spend model, the subsidy to the taxpayer is explicit. Tax expenditure theory posits that the subsidy is just as real when it is achieved through the mechanism of deduction, exclusion, or other statutory measure.[10] Although

4. Under Surrey's leadership, the United States Treasury Department seriously advanced and applied tax expenditure analysis beginning in 1967-68. *See* STANLEY S. SURREY & PAUL R. MCDANIEL, TAX EXPENDITURES 2 (1985). For a history of the Treasury's Department's implementation of tax expenditure theory, see generally STANLEY S. SURREY, PATHWAYS TO TAX REFORM (1973), and Surrey, *Tax Incentives*, *supra* note 3.

5. Surrey, *Tax Incentives*, *supra* note 3, at 706.

6. *See* SURREY & MCDANIEL, *supra* note 4, at 3. The Congressional Budget and Impoundment Control Act of 1974 adds special exemptions to the list of tax expenditures. *See* Pub. L. No. 93-344, § 3, 88 Stat. 299 (1974) (codified as amended at 2 U.S.C. § 622(3)).

7. *See* SURREY & MCDANIEL, *supra* note 4, at 3.

8. *See id.*

9. *See* J. Clifton Fleming, Jr. & Robert J. Peroni, *Reinvigorating Tax Expenditure Analysis and Its International Dimension*, 27 VA. TAX REV. 437, 446 (2008) (describing tax expenditure analysis as "the essence of simplicity"); *see also* Linda Sugin, *Tax Expenditure Analysis and Constitutional Decisions*, 50 HASTINGS L.J. 407, 410 (1999) ("The basic insight of tax expenditure analysis is very simple . . . .").

10. *See* SURREY & MCDANIEL, *supra* note 4, at 3.

the government has foregone revenue, it has just as surely subsidized taxpayers as in the case of a direct grant. The government has economically expended funds through the mechanism of a tax-reducing provision.[11] Hence, the government has made a "tax expenditure."[12]

Tax expenditure theory has altered how the executive and legislative branches function.[13] The Congressional Budget and Impoundment Control Act of 1974 ("Budget Act")[14] required that the President's budget include a list of tax expenditures.[15] The Department of the Treasury historically has prepared this list of tax expenditures, which is published by the Office of Management and Budget ("OMB").[16] Moreover, the Budget Act created the Congressional Budget Office ("CBO") and requires it to report annually to the congressional budget committees the amount of tax expenditures under current law.[17] In fulfilling its statutory mandate, the CBO relies on the analysis of the congressional Joint Committee on Taxation ("JCT"), which itself began issuing tax expenditure publications even prior to the formation of the CBO.[18] Thus, both the Treasury Department and the JCT estimate revenue losses from tax expenditures.[19] Further,

---

11. *See id.* at 1.

12. *Id.* at 25; U.S. GOV'T ACCOUNTABILITY OFF., GAO-13-167SP, TAX EXPENDITURES: BACKGROUND AND EVALUATION CRITERIA AND QUESTIONS 3 (2012) [hereinafter GAO GUIDE FOR EVALUATING TAX EXPENDITURES].

13. *Cf.* SURREY & MCDANIEL, *supra* note 4, at 2 (observing the "rapidly growing recognition of the role of the tax expenditure concept both in budget policy issues and in tax policy issues").

14. Congressional Budget and Impoundment Control Act of 1974, Pub. L. No. 93-344, 88 Stat. 297 (1974) (codified as amended at 2 U.S.C. §§ 601-13).

15. *See* Congressional Budget and Impoundment Control Act of 1974, Pub. L. No. 93-344, 88 Stat. 297. This requirement appeared in Title VI of the Budget Act, which has been repealed. *See* Pub. L. No. 105-33, § 10118(a), 111 Stat. 695 (repealed 1997). The Treasury Department has continued to publish a tax expenditures budget.

16. *See* BENJAMIN H. HARRIS ET AL., TAX POL'Y CTR., EVALUATING TAX EXPENDITURES: INTRODUCING OVERSIGHT INTO SPENDING THROUGH THE TAX CODE 10 (2018), [https://perma.cc/B9VR-BNYC]; McMahon, *supra* note 2, at 776; *see also, e.g.*, OFF. OF TAX ANALYSIS, U.S. DEP'T OF THE TREASURY, TAX EXPENDITURES (2021) [hereinafter TREASURY REPORT], [https://perma.cc/J5SR-X42Y].

17. Pub. L. No. 93-344, §§ 201-02, 88 Stat. 302-04 (1974) (codified as amended at 2 U.S.C. §§ 601-02).

18. *See* STAFF OF JOINT COMM. ON TAX'N, 116TH CONG., ESTIMATES OF FEDERAL TAX EXPENDITURES FOR FISCAL YEARS 2020-2024, at 2 n.4 (2020) [hereinafter JCT TAX EXPENDITURES REPORT].

19. GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 4.

the Budget Act requires the budget committees of each house of Congress to request, evaluate, and report on tax expenditure studies as they develop congressional budget resolutions.[20] The Congressional Research Service (CRS) regularly prepares a committee print for the Senate Budget Committee to facilitate its compliance with the Budget Act.[21]

Tax expenditure theory also finds expression in case law. However, the United States Supreme Court has demonstrated remarkable ambivalence towards tax expenditure theory when deciding constitutional questions arising from nominal tax exemptions, deductions, and credits.[22] In some decisions, justices write as though they were discipled at the feet of Stanley Surrey.[23] In others, justices distance their analysis from the apparent implications of the theory by distinguishing direct monetary subsidies from the indirect benefits reaped by those who avail themselves of various statutory mechanisms for reducing tax liabilities.[24] Tax expenditure theory has thus left a meandering

---

20. *See* §§ 101-02, 88 Stat. at 299-301.

21. *See* CONG. RSCH. SERV., 116TH CONG., TAX EXPENDITURES: COMPENDIUM OF BACKGROUND MATERIAL ON INDIVIDUAL PROVISIONS, at III [hereinafter TAX EXPENDITURES: COMPENDIUM] (Comm. Print 2020).

22. *See* Edward A. Zelinsky, *Are Tax "Benefits" Constitutionally Equivalent to Direct Expenditures?*, 112 HARV. L. REV. 379, 380-81 (1998) [hereinafter Zelinsky, *Tax "Benefits"*] (stating that the Supreme Court "has equivocated" in decisions presenting the question of whether tax benefits and direct spending are the same).

23. *See, e.g.*, Ariz. Christian Sch. Tuition Org. v. Winn, 563 U.S. 125, 148 (2011) (Kagan, J., dissenting) ("This novel distinction in standing law between appropriations and tax expenditures has as little basis in principle as it has in our precedent. Cash grants and targeted tax breaks are means of accomplishing the same government objective—to provide financial support to select individuals or organizations."); Tex. Monthly, Inc. v. Bullock, 489 U.S. 1, 14 (1989) (Brennan, J., plurality opinion) ("Every tax exemption constitutes a subsidy that affects nonqualifying taxpayers . . . ."); Bob Jones Univ. v. United States, 461 U.S. 574, 587-88, 592 (1983) (finding that Internal Revenue Code sections 501(c)(3) and 170 reflect a congressional desire "to provide tax benefits to charitable organizations, to encourage the development of private institutions that serve a useful public purpose," and holding a private school's admission policy forbidding interracial dating violated "fundamental public policy"); Regan v. Tax'n with Representation of Wash., 461 U.S. 540, 544 (1983) (Rehnquist, J.) ("Both tax exemptions and tax deductibility are a form of subsidy that is administered through the tax system.").

24. *See, e.g.*, *Winn*, 563 U.S. at 144, 146 (holding that taxpayers lack standing to challenge the constitutionality of a state program providing tax credits for transfers to tuition organizations and rejecting the idea "that income should be treated as if it were government property even if it has not come into the tax collector's hands"); Camps Newfound/Owatonna, Inc. v. Town of Harrison, 520 U.S. 564, 588-89 (1997) (holding that the Commerce Clause was violated by a state law that exempted the property of most

trail of scattered footprints in the pages of Supreme Court opinions.

Existing legal scholarship features different approaches to analyzing Supreme Court opinions that discuss tax expenditure theory or rely on its basic ideas.[25] Commonly, analysts discuss tax expenditure theory with a focused commentary on individual cases or distinct doctrinal areas of constitutional law.[26] Some scholars assume or assert the salience of tax expenditure theory to constitutional adjudication.[27] Others suggest a more cautious, even critical, view of employing tax expenditure concepts in constitutional cases. A few commentators argue that tax expenditure theory has limited value in at least some constitutional contexts.[28] Some offer a more systematized approach than others.[29] Each of these critical approaches

---

charities from taxation but denied the general exemption to charities that operated primarily for the benefit of nonresidents, and stating that "tax exemptions and subsidies serve similar ends" but "differ in important and relevant respects, and our cases have recognized these distinctions"); Walz v. Tax Comm'n of N.Y., 397 U.S. 664, 679-80 (1970) (holding that granting a property tax exemption to religious organizations along with other charities did not violate the Establishment Clause).

25. *See, e.g.*, Donna D. Adler, *The Internal Revenue Code, the Constitution, and the Courts: The Use of Tax Expenditure Analysis in Judicial Decision Making*, 28 WAKE FOREST L. REV. 855 (1993); E.C. Lashbrooke, Jr., *An Economic and Constitutional Case for Repeal of the I.R.C. Section 170 Deduction for Charitable Contributions to Religious Organizations*, 27 DUQ. L. REV. 695 (1989); Zelinsky, *Tax "Benefits," supra* note 22.

26. *See, e.g.*, Boris I. Bittker, *Churches, Taxes and the Constitution*, 78 YALE L.J. 1285 (1969) [hereinafter Bittker, *Churches*] (analyzing the constitutionality of tax exemptions for churches under the Establishment Clause); Boris I. Bittker & Kenneth M. Kaufman, *Taxes and Civil Rights: "Constitutionalizing" the Internal Revenue Code*, 82 YALE L.J. 51, 61-74 (1972) (discussing how tax expenditure concepts relate to state action and the Due Process Clause of the Fifth Amendment in the context of private acts of discrimination); Sugin, *supra* note 9, at 413 ("This Article looks at the significance of the similarities and differences between tax benefits and direct spending for purposes of the equal protection and establishment clauses, with a particular focus on the charitable contribution deduction.").

27. *See, e.g.*, Adler, *supra* note 25, at 864-65; Lashbrooke, *supra* note 25, at 717-18.

28. In her thoughtful treatment of tax expenditures and the Constitution, Professor Linda Sugin offers four reasons that tax expenditure analysis "is problematic if applied unreflectively as a basis for constitutional adjudication." *See* Sugin, *supra* note 9, at 415-30.

29. Professor Edward Zelinsky's excellent article critiquing the typical, binary approach of either embracing or rejecting the constitutional equivalence of tax expenditures and government subsidies presents a framework for analyzing the underlying nature of the benefit under scrutiny. *See, e.g.*, Zelinsky, *Tax "Benefits," supra* note 22, at 400-13. His framework considers a benefit's structural features (in terms of permanence, eligibility, and quantity) and how that benefit is perceived under three perspectives (that of the beneficiary as it receives funds, that of the government as it decides the scope and purpose of the provision, and that of the government as it conducts the process of dispensing benefits). *See*

meaningfully advances legal scholarship on tax expenditures and the Constitution.

Nonetheless, much existing scholarship tends to move quickly to how tax expenditure analysis has (or has not) informed, or should (or should not) inform, the resolution of specific cases or doctrinal areas. Only a few scholars have explored whether there may be excellent reasons to apply tax expenditure theory differently, or how tax expenditure theory presents unique challenges, in cases raising constitutional issues.[30] Certainly, a comprehensive treatment of the limits of tax expenditure theory in constitutional contexts at a high level of generality is wanting. Existing literature, as well as future judicial opinions, would benefit from a prelude. Such a prelude would alert the judiciary and the legal academy to a wide assortment of assumptions, lessons, and vagaries of tax expenditure theory relevant to deciding constitutional cases.

This Article is such a prelude. It discusses numerous reasons—most of which are interconnected—for applying tax expenditure analysis in constitutional cases with caution, qualification, and a critical eye. In doing so, it also suggests how judges should assess various aspects of tax expenditure theory in resolving constitutional questions.

The reader should understand *ab initio* what this Article is not. It is not a critique of tax expenditure theory as a fiscal tool, let alone a hostile assault on it generally. This Article necessarily probes tax expenditure theory and identifies special challenges that the theory presents for constitutional law. However, to interpret this Article as refuting the gist of tax expenditure theory or as rehashing old debates over it is to misread the Article. The argument assumes, *arguendo*, that tax expenditure theory is a valuable tool of fiscal policy.[31]

Further, this Article is not a detailed, case-specific analysis of the body of Supreme Court decisions accepting or rejecting

---

id. He then applies the framework to areas of constitutional law on a case-by-case basis. *See id.* at 413-32.

30. *See, e.g.*, Sugin, *supra* note 9, at 415-30; Zelinsky, *Tax "Benefits*,*" supra* note 22, at 380-81.

31. Therefore, both the strongest advocates for tax expenditure theory and its most vocal opponents should keep reading.

concepts embraced by tax expenditure theory. The Article refers to Supreme Court opinions when they illustrate a normative or descriptive point; however, the purpose of the paper is not to restate, deconstruct, or synthesize Supreme Court jurisprudence on tax expenditures. Neither is this Article primarily an analysis of specific constitutional cases or doctrinal areas. This Article is a prelude, one that is currently missing in legal scholarship. Building on this prelude, further work examining Supreme Court cases involving tax expenditures in discreet doctrinal areas is plainly warranted.

The structure of this Article is simple. Part I summarizes the basic idea and insights of tax expenditure theory, as well as the major objections to its features that are relevant to constitutional analysis. Part II explains why the application of tax expenditure theory to constitutional questions is uneasy. It also cautions courts against an indiscriminate, reflexive approach to the theory. Part III concludes.

I. FUNDAMENTAL INSIGHTS AND CRITICISMS OF TAX EXPENDITURE THEORY

A. Tax Expenditure Basics

In their classic, co-authored book, Stanley Surrey and Paul McDaniel describe the idea of tax expenditures as envisioning an income tax with "two distinct elements."[32] The first "consists of structural provisions necessary to implement a normal income tax."[33] Examples include provisions that determine accounting rules, identify taxable entities, define net income, fix tax rates, set forth personal exemptions, and specify the scope of taxable international transactions.[34] The second major feature of a tax system consists of the "tax incentives," the "special preferences found in every income tax."[35] These items depart "from the

32. SURREY & MCDANIEL, *supra* note 4, at 3; *see also* Stanley S. Surrey & Paul R. McDaniel, *The Tax Expenditure Concept and the Budget Reform Act of 1974*, 17 B.C. INDUS. & COM. L. REV. 679, 679-80 (1976) (positing the same two elements of an income tax system).

33. SURREY & MCDANIEL, *supra* note 4, at 3.

34. *See id.*

35. *Id.*

normal tax structure"[36] and constitute "government spending" through tax laws for someone or something.[37] The spending takes the form of reductions in tax liabilities that otherwise would be due.[38]

Examples of tax expenditures that appear in the 2021 tax expenditures budget published by the United States Department of the Treasury[39] are the credit for construction of energy efficient homes,[40] the excess of the deduction for percentage depletion[41] over cost depletion for minerals,[42] the exclusion of life insurance death benefits from the income of a beneficiary named in the policy,[43] various tax credits and exclusions for post-secondary education,[44] and the exclusion from an employee's income of employer-paid health insurance premiums.[45] Clearly, the reductions in tax liability recognized as tax expenditures may be direct offsets to tax otherwise due (in the case of a tax credit) or adjustments to various income figures employed in calculating taxable income (for example, a gross income exclusion or a deduction from gross income or adjusted gross income). The form of the tax expenditure is irrelevant to its classification as such.[46]

For each tax expenditure provision, the governmental entities responsible for compiling official lists of tax expenditures (the Treasury Department and the JCT) calculate the tax expenditure associated with the provision as the difference between income tax liability under existing law (i.e., with the provision in place) and the hypothetical tax liability that would exist without the provision.[47] This methodology is simple. The

---

36. *Id.*
37. *Id.*
38. *See* SURREY & MCDANIEL, *supra* note 4, at 3.
39. *See* TREASURY REPORT, *supra* note 16, at tbls. 1-4.
40. *See id.* at 7; I.R.C. § 45L.
41. *See* TREASURY REPORT, *supra* note 16, at 7; I.R.C. § 613.
42. *See* I.R.C. § 611(a).
43. *See* I.R.C. § 101(d).
44. *See, e.g.*, I.R.C. §§ 25A, 117, 127, 529.
45. *See* I.R.C. § 106.
46. *See* SURREY & MCDANIEL, *supra* note 4, at 3.
47. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 4. The tax expenditure lists prepared by the Treasury Department and the JCT differ in six modest respects. *See* JCT TAX EXPENDITURES REPORT, *supra* note 18, at 15-16.

calculations of revenue loss for a single tax expenditure assume that other tax expenditures continue.[48] The calculations further assume that taxpayer behavior does not change when the tax expenditure is repealed.[49] Consequently, the revenue losses that tax expenditure budgets estimate do not likely equate to revenues the government would gain from repealing tax expenditures.[50]

B. Critiques of Tax Expenditure Analysis

Critiques of tax expenditure theory have varied from the thoroughly unpersuasive to the intellectually sophisticated.[51] This Section discusses two critiques that are most relevant to constitutional analysis.

1. The Assumption of a Normative Tax Base

Perhaps the most controversial feature of tax expenditure analysis is its reliance on a normatively correct base.[52] The base of a tax is simply that upon which tax is assessed. For example, in a capitation tax, the tax base is human beings.[53] Under a real property ad valorem tax, the base is the value of land and structures built on it for residential and non-residential use.[54] In

48. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 5.

49. *See id.*

50. *See id.*

51. For a summary, see McMahon, *supra* note 2, at 778-80.

52. *See* J. Clifton Fleming, Jr. & Robert J. Peroni, *Can Tax Expenditure Analysis Be Divorced from a Normative Tax Base?: A Critique of the "New Paradigm" and Its Denouement*, 30 VA. TAX REV. 135, 142 (2010) [hereinafter Fleming & Peroni, *A Critique*] (describing the "strongest attacks" on tax expenditure theory as targeting the "baseline"). For critical assessments of the attempt to establish a normative income tax base, see Douglas A. Kahn & Jeffrey S. Lehman, *Tax Expenditure Budgets: A Critical View*, 1992 TAX NOTES (SPECIAL REPORT) 1661, and David A. Weisbach & Jacob Nussim, *The Integration of Tax and Spending Programs*, 113 YALE L.J. 955, 976 (2004) ("There is no such thing as a normative tax base.").

53. *Poll Tax*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("A fixed tax levied on each person within a jurisdiction. . . . Also termed . . . *capitation* [or] *head tax*."); *see also* Joseph M. Dodge, *What Federal Taxes Are Subject to the Rule of Apportionment Under the Constitution?*, 11 U. PA. J. CONST. L. 839, 841 (2009).

54. *Ad valorem Tax*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("A tax imposed proportionally on the value of something (esp. real property), rather than on its quantity or some other measure."). *See generally* 71 AM. JUR. 2D *State and Local Taxation* § 18 (2022) (stating that *ad valorem* taxes are taxes "levied according to the value of property as determined by an assessment or appraisal" and are "invariably based upon ownership of

a retail sales tax, the tax base is consumption measured at point of sale to the ultimate consumer.[55]

Although tax expenditure analysis applies regardless of the type of tax at issue,[56] its most celebrated and rigorous implementation has occurred in evaluating the federal income tax.[57] Plainly, deciding on the normative income tax base is necessary to tax expenditure analysis for a simple reason: the very idea of an indirect subsidy taking the form of a "special" provision that departs from what the tax would otherwise be assumes the existence of a norm.[58] There can be no deviation unless first there is a standard, or norm, from which to stray.[59] It is thus instructive to consider how Surrey approached the task of identifying a normal income tax base in his tax expenditure analysis. His starting point was a concept of economic income.

The most widely accepted theoretical construct of economic income in this country is the Haig-Simons concept.[60] Henry Simons defined personal income as follows: "the algebraic sum of (1) the market value of rights exercised in consumption and (2)

---

property"; such taxes are "payable regardless of whether the property is used or not although the value may vary in accordance with such a factor").

55. *Sales Tax*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("A tax imposed on the sale of goods and services . . . . Also termed *retail sales tax*."). *See generally* 67B AM. JUR. 2D *Sales and Use Taxes* § 1 (2022).

56. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 3 n.6; SURREY & MCDANIEL, *supra* note 4, at 233. Professor Boris Bittker expressed skepticism that the tax expenditure concept could improve the federal estate tax. *See* Boris I. Bittker, *Accounting for Federal "Tax Subsidies" in the National Budget*, 22 NAT'L TAX J. 244, 260 (1969) [hereinafter Bittker, *"Tax Subsidies"*].

57. *See* Fleming & Peroni, *A Critique*, *supra* note 52, at 138 n.7.

58. *See* SURREY & MCDANIEL, *supra* note 4, at 3-4; TREASURY REPORT, *supra* note 16, at 1; Bittker, *Churches*, *supra* note 26, at 1296; Fleming & Peroni, *supra* note 9, at 450-51.

59. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 3; TREASURY REPORT, *supra* note 16, at 1.

60. *See, e.g.*, Boris I. Bittker, *A "Comprehensive Tax Base" as a Goal of Income Tax Reform*, 80 HARV. L. REV. 925, 932 (1967) (observing that commentators advocating a comprehensive tax base state or imply that Congress should strive to enact the Haig-Simons concept of income to the extent possible); Victor Thuronyi, *The Concept of Income*, 46 TAX L. REV. 45, 46 (1990) (stating that the "income concept that is now widely accepted by analysts" is the Haig-Simons concept). Although Henry Simons's concise articulation of income is the one that is most often cited, it is referred to as the "Haig-Simons" concept to acknowledge the prior work of Robert Haig. *See id. See generally* Robert Murray Haig, *The Concept of Income–Economic and Legal Aspects*, *in* THE FEDERAL INCOME TAX 1, 1-28 (Robert Muray Haig ed., 1921).

the change in the value of the store of property rights between the beginning and end of the period in question."[61] A short-hand expression of this definition of income is accumulation plus consumption for the taxable period. Surrey himself embraced this notion of income and argued that it embodied the basic norm underlying the federal income tax.[62]

But nailing down the breadth of this norm for purposes of applying tax expenditure theory has proven more challenging than a casual reading of the Haig-Simons definition suggests. When pressed by the late Professor Boris Bittker,[63] Surrey admitted that tax expenditure theory did not rely exclusively on the Haig-Simons definition of income.[64] Both practical administrative realities and public acceptance of taxation bear upon the standard, deviations from which are counted as tax expenditures.[65] For example, the Simons definition of income includes unrealized appreciation of assets, but not even Simons thought it practical to insist on trying to tax unrealized appreciation.[66] Neither did Surrey classify this unrealized appreciation as a tax expenditure.[67] Moreover, Surrey did not consider the progressive income tax rate schedules or the deduction for the personal exemption as tax preferences to those who benefit from them, but rather as components of "the structure of an income tax system based on ability to pay."[68] At a minimum, then, the standard by which tax expenditure theory identifies "deviations" may reflect norms—including administrative norms and the ability-to-pay norm—that deviate

---

61. HENRY C. SIMONS, PERSONAL INCOME TAXATION: THE DEFINITION OF INCOME AS A PROBLEM OF FISCAL POLICY 50 (1938).

62. *See* SURREY & MCDANIEL, *supra* note 4, at 4-5, 186-88.

63. *See* Bittker, *"Tax Subsidies," supra* note 56, at 247-51, 260-61.

64. *See* Stanley S. Surrey & William F. Hellmuth, *The Tax Expenditure Budget—Response to Professor Bittker*, 22 NAT'L TAX J. 528, 531 (1969).

65. *See* Victor Thuronyi, *Tax Expenditures: A Reassessment*, 1988 DUKE L.J. 1155, 1165-66.

66. *Id.*

67. Surrey justified the non-inclusion of unrealized appreciation in the tax expenditure budget by appealing to public conceptions of income, the historical treatment of realization as integral to income, and administrative simplicity. *See* SURREY & MCDANIEL, *supra* note 4, at 198-99. However, he encouraged periodic reassessment of this item. *See id.*

68. *See* Surrey & Hellmuth, *supra* note 64, at 529. Similarly, both the Treasury Department and the JCT do not classify the personal exemption as a tax expenditure. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 4.

from economic income as captured by the Haig-Simons concept.[69]

When an income standard is constructed by trying to account for an unspecified and incommensurable aggregate of various legitimate norms of tax policy, the precise boundaries of income are difficult to determine.[70] Surrey acknowledged as much. He wrote that "[t]he precise contours of the dividing line will of course be uncertain."[71] He also approvingly cited an analysis recognizing that exclusions from the tax expenditure listing "are to some extent arbitrary" and that the list should not include "highly complicated or controversial items."[72] Such language bespeaks an awareness that reasonable minds can differ on the finer points of the income standard by which tax expenditures are measured.

Notwithstanding the uncertainties surrounding the analytical borders of income, Surrey argued for "a very large area of tax law which can be considered within the guidelines" of the Treasury's tax expenditure analysis.[73] Many have agreed with him. Still, the lack of precision in the scope of the income standard has long opened space for questioning not just omissions from the tax expenditure budget, but also inclusions in it. For example, the late Professor William Andrews is well known for focusing on the element of income consisting of personal consumption.[74] Andrews pressed the point that not all non-business-related

69. Surrey classified some provisions that deviate from the Haig-Simons concept of income as other than tax expenditures because they comport with the ability-to-pay norm. But not all features of tax law consistent with this norm escape classification as a tax expenditure. *See* Boris I. Bittker, *The Tax Expenditure Budget—A Reply to Professors Surrey & Hellmuth*, 22 NAT'L TAX J. 538, 539 (1969) [hereinafter Bittker, *A Reply*].

70. *See* Charlotte Crane, *The Income Tax and the Burden of Perfection*, 100 NW. U. L. REV. 171, 185 (2006); *cf.* TREASURY REPORT, *supra* note 16, at 2 ("[D]eciding whether provisions are exceptions, therefore, is a matter of judgment.").

71. Surrey & Hellmuth, *supra* note 64, at 531.

72. *Id.* at 529-30 (quoting the *1969 Economic Report of the President: Hearings Before the Joint Econ. Comm.*, 91st Cong. 33 (1969)) (statement of Joseph W. Barr, Secretary of the Treasury).

73. *Id.* at 533.

74. William D. Andrews, *Personal Deductions in an Ideal Income Tax*, 86 HARV. L. REV. 309 (1972). For a summary of Andrews's understanding of income and how personal deductions figure into an income base, see Thomas D. Griffith, *Theories of Personal Deductions in the Income Tax*, 40 HASTINGS L.J. 343, 366-77 (1989).

transfers represent personal consumption.[75] Andrews made a thoughtful case, albeit a controversial one, that charitable contributions are best viewed as other than personal consumption.[76] Under Andrews's logic, if income is accumulation plus consumption, and a charitable contribution is not personal consumption, then a taxpayer's taxable income should not include her charitable contributions.[77] It follows that a deduction for charitable contributions is appropriate in calculating the donor's taxable income.[78] So understood, the charitable contributions deduction is not a tax expenditure; rather, it is a mechanism for arriving at the proper tax base.

This position is, of course, contrary to the tax expenditure lists promulgated by the federal government[79] and to Surrey's assessment of the charitable contributions deduction.[80] But it is nonetheless representative of serious scholarship taking issue with some fairly basic assumptions underlying the income standard on which tax expenditure analysis hangs. Andrews highlighted a foundational conceptual issue in what is meant by income: if the "income" subject to taxation is that which is consumed or saved, then income that a taxpayer does not save or consume (in the sense of appropriating goods and services purchased or exchanged in a market transaction for the taxpayer) is not part of the income standard.[81] A vigorous debate has ensued as to whether this understanding of the Haig-Simons concept of income has any legs.[82] The present point is not that Andrews was

---

75. *See* Andrews, *supra* note 74, at 313-15.

76. *See id.* at 344-75.

77. *See id.* at 346.

78. *See id.*

79. *See, e.g.*, JCT TAX EXPENDITURES REPORT, *supra* note 18, at 9; TREASURY REPORT, *supra* note 16, at 14 (item 104), 17 (item 129).

80. *See* SURREY & MCDANIEL, *supra* note 4, at 79, 169-70; Stanley S. Surrey, *Federal Income Tax Reform: The Varied Approaches Necessary to Replace Tax Expenditures with Direct Governmental Assistance*, 84 HARV. L. REV. 352, 384-85 (1970).

81. *See* Andrews, *supra* note 74, at 325.

82. For critiques of the analysis of Andrews on the role of personal deductions generally or the charitable contributions deduction specifically, see Mark P. Gergen, *The Case for a Charitable Contributions Deduction*, 74 VA. L. REV. 1393, 1414-26 (1988), and Mark G. Kelman, *Personal Deductions Revisited: Why They Fit Poorly in an "Ideal" Income Tax and Why They Fit Worse in a Far from Ideal World*, 31 STAN. L. REV. 831 (1979), and Stanley A. Koppelman, *Personal Deductions Under an Ideal Income Tax*, 43 TAX L. REV. 679 (1988). For more sympathetic assessments of Andrews's arguments, see Johnny Rex

right and Surrey was wrong about the charitable contributions deduction. The point is that there is serious debate about what the Haig-Simons concept of income even means, or should mean when it is employed to craft a legal definition of income, with respect to non-purchased consumption. At a minimum, there is at least some reason to question the vastness of the domain of Surrey's "very large area of tax law which can be considered within the guidelines" of tax expenditure analysis.

2. The Effect on Behavioral Changes and Resulting Revenues

Tax expenditure budgets do not attempt to quantify the real revenue effects of eliminating the targeted tax provisions because they do not account for changes in taxpayer behavior likely to result from a change in the law. As Professor Boris Bittker argued long ago, the informational value of tax expenditure budgets so constructed is limited.[83] The government does not know how much revenue it loses through a tax provision without knowing what taxpayers would do in response to eliminating the provision.[84]

One intuitive response to this observation is that some information is probably better than none, and tax expenditure budgets provide some basis for comparing direct spending alternatives and support through tax provisions.[85] Further, the practical reality, acknowledged by the Treasury Department, is that computing an accurate estimate of revenue loss after considering taxpayer behavioral changes is difficult, if not impossible.[86] For Bittker, this fact alone calls into question the decision to label a tax provision a "tax expenditure."[87] At a minimum, the unrealistic assumption of tax expenditure budgeting that taxpayer behavior remains unchanged means that

---

Buckles, *The Community Income Theory of the Charitable Contributions Deduction*, 80 IND. L.J. 947 (2005), and Jeffrey H. Kahn, *Personal Deductions—A Tax "Ideal" or Just Another "Deal"?*, 2002 L. REV. MICH. STATE U. DETROIT COLL. L. 1 (2002).

83. *See* Bittker, *"Tax Subsidies," supra* note 56, at 247.

84. *See id.*

85. *See* Fleming & Peroni, *supra* note 9, at 521-22.

86. *See* Bittker, *"Tax Subsidies," supra* note 56, at 247.

87. *See id.*

any comparison between tax expenditure budgets and direct appropriations must be qualified.[88]

II. THE UNEASY APPLICATION OF TAX EXPENDITURE THEORY IN CONSTITUTIONAL CONTEXTS

Surrey and McDaniel devote an entire chapter in their co-authored book to discussing how courts have analyzed cases involving tax expenditures.[89] Their position—indeed, nearly the entire analytical depth of their reflection—is encapsulated in the first paragraph of the discussion. They assert that, because "tax expenditures are government assistance programs, it would seem almost axiomatic" that constitutional doctrines governing direct government spending and those who accept it also apply to "tax expenditure benefits and to private entities receiving them."[90]

Surrey and McDaniel frame the issue common to numerous constitutional contexts as "whether tax assistance is equivalent to direct assistance."[91] They conclude the answer "must be" yes "under rational governmental and judicial decisions."[92] Moreover, in an analytical quantum leap, they insist that a court need not independently analyze tax provisions to determine if they are a form of assistance.[93] Courts should simply accept the tax expenditure lists appearing in the federal budgets.[94] Further, equating the Senate Budget Committee's characterization of tax expenditures with the "congressional view" itself,[95] Surrey and McDaniel maintain that "it would seem difficult—and wrong—for courts to apply different rules to direct programs and to tax expenditures."[96]

---

88. *See id.*
89. *See* SURREY & MCDANIEL, *supra* note 4, at 119-55.
90. *See id.* at 118.
91. *See id.* at 119.
92. *Id.*
93. *See id.*
94. *See* SURREY & MCDANIEL, *supra* note 4, at 119.
95. *Id.*
96. *Id.* Such language tends to feed into criticisms of tax expenditure theory. *See, e.g.*, Kahn & Lehman, *supra* note 52, at 1662 ("What is disturbing about the language of tax expenditures is its tone of moral absolutism.").

In this summary articulation of their position, Surrey and McDaniel thus do little more than assert that judges deciding constitutional questions are duty-bound to accept the executive's (or a legislative committee's) characterization of a tax provision as equivalent to a cash subsidy.[97] But it turns out that Surrey's ultimate view is that judges do have a bit of freedom; he permits them to *add* to the tax expenditure list, insofar as the tax expenditure budget does not purport to be exhaustive.[98] Surrey's methodology is therefore a one-way ratchet under which judges "rationally" cannot question the executive's characterization of a tax provision as a cash subsidy equivalent, but apparently they can and should rationally employ the tax expenditure concept to expand the official list.

Surrey's discussion of specific cases adds little to this synopsis. Representative of Surrey's view is his description of *Regan v. Taxation with Representation*,[99] which he praises for its reflection of tax expenditure insights.[100] In *Regan*, the Court upheld the constitutionality of the requirement in section 501(c)(3) of the Internal Revenue Code (the "IRC") that "no substantial part" of a tax-exempt charitable organization's activities consist of "carrying on propaganda, or otherwise attempting to influence legislation."[101] Writing for the majority,

---

97. In his review of Surrey & McDaniel's book, Professor Bernard Wolfman offers a similar, more general, assessment. *See* Bernard Wolfman, *Tax Expenditures: From Idea to Ideology*, 99 HARV. L. REV. 491 (1985) (book review). Wolfman states that "the book is more rigid and dogmatic than it is persuasive and pragmatic." *Id.* at 495. He continues:

> Instead of demonstrating in case upon case that tax expenditure analysis is a useful tool and arguing that it should therefore be used more often than not because of the practical benefits it offers, the authors insist that government must utilize tax expenditure analysis as a matter of logical necessity.

Id. (emphasis omitted).

98. *See* SURREY & MCDANIEL, *supra* note 4, at 144.

99. 461 U.S. 540 (1983).

100. *See* SURREY & MCDANIEL, *supra* note 4, at 120-22.

101. *Regan*, 461 U.S. at 542 n.1 (emphasis omitted). An extensive academic commentary discusses the political speech constraints on charities under section 501(c)(3) of the Internal Revenue Code. *See, e.g.*, PHILIP HAMBURGER, LIBERAL SUPPRESSION: SECTION 501(C)(3) AND THE TAXATION OF SPEECH (2018); Erik J. Ablin, *The Price of Not Rendering to Caesar: Restrictions on Church Participation in Political Campaigns*, 13 NOTRE DAME J.L. ETHICS & PUB. POL'Y 541 (1999); Ellen P. Aprill, *Churches, Politics, and the Charitable Contribution Deduction*, 42 B.C. L. REV. 843 (2001); Samuel D. Brunson, *Dear IRS, It Is Time to Enforce the Campaigning Prohibition. Even Against Churches*, 87 UNIV. COLO. L. REV. 143 (2016); Samuel D. Brunson, *Reigning in Charities:*

Using an Intermediate Penalty to Enforce the Campaigning Prohibition, 8 PITT. TAX REV. 125 (2011); Johnny Rex Buckles, *A Rawlsian Critique of the Political Speech Constraints on Charities*, 16 FIU L. REV. 479 (2022); Johnny Rex Buckles, *The Penalty of Liberty*, 25 TEX. REV. L. & POL. 159 (2020) [hereinafter Buckles, *The Penalty of Liberty*]; Johnny Rex Buckles, *Curbing (or Not) Foreign Influence on U.S. Politics and Policies Through the Federal Taxation of Charities*, 79 MD. L. REV. 590 (2020); Johnny Rex Buckles, *Does the Constitutional Norm of Separation of Church and State Justify the Denial of Tax Exemption to Churches that Engage in Partisan Political Speech?*, 84 IND. L.J. 447 (2009); Johnny Rex Buckles, *Is the Ban on Participation in Political Campaigns by Charities Essential to Their Vitality and Democracy? A Reply to Professor Tobin*, 42 RICH. L. REV. 1057 (2008); Johnny Rex Buckles, *Not Even a Peep? The Regulation of Political Campaign Activity by Charities Through Federal Tax Law*, 75 CIN. L. REV. 1071 (2007); Wilfred R. Caron & Deirdre Dessingue, *I.R.C. § 501(c)(3): Practical and Constitutional Implications of "Political" Activity Restrictions*, 2 J.L. & POL. 169 (1985); Anne Berrill Carroll, *Religion, Politics, and the IRS: Defining the Limits of Tax Law Controls on Political Expression by Churches*, 76 MARQ. L. REV. 217 (1992); Laura Brown Chisolm, *Politics and Charity: A Proposal for Peaceful Coexistence*, 58 GEO. WASH. L. REV. 308 (1990); Roger Colinvaux, *The Political Speech of Charities in the Face of* Citizens United*: A Defense of Prohibition,* 62 CASE W. RSRV. L. REV. 685 (2012); Deirdre Dessingue, *Prohibition in Search of a Rationale: What the Tax Code Prohibits; Why; to What End?*, 42 B.C. L. REV. 903 (2001); Alan L. Feld, *Rendering unto Caesar or Electioneering for Caesar? Loss of Church Tax Exemption for Participation in Electoral Politics*, 42 B.C. L. REV. 931 (2001); Edward McGlynn Gaffney, Jr., *On Not Rendering to Caesar: The Unconstitutionality of Tax Regulation of Activities of Religious Organizations Relating to Politics*, 40 DEPAUL L. REV. 1 (1990); Brian Galle, *Charities in Politics: A Reappraisal*, 54 WM. & MARY L. REV. 1561 (2013); Miriam Galston, *When Statutory Regimes Collide: Will* Citizens United *and* Wisconsin Right to Life *Make Federal Tax Regulation of Campaign Activity Unconstitutional?*, 13 U. PA. J. CONST. L. 867 (2011); Miriam Galston, *Lobbying and the Public Interest: Rethinking the Internal Revenue Code's Treatment of Legislative Activities*, 71 TEX. L. REV. 1269 (1993); Richard W. Garnett, *A Quiet Faith? Taxes, Politics, and the Privatization of Religion*, 42 B.C. L. REV. 771 (2001); Michael Hatfield, *Ignore the Rumors—Campaigning from the Pulpit Is Okay: Thinking Past the Symbolism of Section 501(c)(3)*, 20 NOTRE DAME J.L. ETHICS & PUB. POL'Y 125 (2006); Oliver A. Houck, *On the Limits of Charity: Lobbying, Litigation, and Electoral Politics by Charitable Organizations Under the Internal Revenue Code and Related Laws*, 69 BROOK. L. REV. 1 (2003); Steffen N. Johnson, *Of Politics and Pulpits: A First Amendment Analysis of IRS Restrictions on the Political Activities of Religious Organizations*, 42 B.C. L. REV. 875 (2001); Chris Kemmitt, *RFRA, Churches and the IRS: Reconsidering the Legal Boundaries of Church Activity in the Political Sphere*, 43 HARV. J. ON LEGIS. 145 (2006); Randy Lee, *When a King Speaks of God; When God Speaks to a King: Faith, Politics, Tax Exempt Status, and the Constitution in the Clinton Administration*, 63 LAW & CONTEMP. PROBS. 391 (2000); Benjamin M. Leff, *"Sit Down and Count the Cost": A Framework for Constitutionally Enforcing the 501(c)(3) Campaign Intervention Ban*, 28 VA. TAX REV. 673 (2009); Jill S. Manny, *Nonprofit Legislative Speech: Aligning Policy, Law, and Reality*, 62 CASE W. RSRV. L. REV. 757 (2012); Lloyd Hitoshi Mayer, *Charities and Lobbying: Institutional Rights in the Wake of* Citizens United, 10 ELECTION L.J. 407 (2011); Lloyd Hitoshi Mayer, *Grasping Smoke: Enforcing the Ban on Political Activity by Charities*, 6 FIRST AMEND. L. REV. 1 (2007); Lloyd Hitoshi Mayer, *Politics at the Pulpit: Tax Benefits, Substantial Burdens, and Institutional Free Exercise*, 89 B.U. L. REV. 1137 (2009); Ann M. Murphy, *Campaign Signs and the Collection Plate—Never the Twain Shall Meet?*, 1 PITT. TAX REV. 35 (2003); Patrick L. O'Daniel, *More Honored in the Breach: A Historical Perspective of the Permeable IRS Prohibition on Campaigning by Churches*, 42 then-Associate Justice Rehnquist opined that both federal income tax exemption and the ability to receive donations that are deductible by donors under IRC section 170 constitute forms of governmental subsidy.[102] Echoing Surrey without citing him, Justice Rehnquist explained that, by conditioning the favorable tax benefits on complying with the lobbying limitations, "Congress has merely refused to pay for the lobbying out of public moneys."[103] Surrey expressed hope that the reasoning employed by Justice Rehnquist in *Regan* would influence future cases.[104]

Notwithstanding Surrey's assertions to the contrary, that judges should defer in knee-jerk fashion to the characterization of a tax provision in the executive's tax expenditure budget—unless, of course, a judge seeks to add to the list of tax expenditures—is far from "axiomatic." Further, a judge's exercise of judgment as to the nature of a tax provision when deciding constitutional cases hardly seems "wrong." Rather, exercising judgment sounds exactly like what Article III of the United States Constitution requires judges to do. Surrey simply declined to explore reasons that constitutional questions might require a judge to analyze tax provisions on the tax expenditure list differently from how Treasury and OMB officials sensibly approach the compilation of tax expenditure budgets.

This Section of the Article explores reasons that standard tax expenditure theory fits uneasily in constitutional analysis. Some of these reasons are fairly obvious. Some are not. When numerous nuances are explored, Surrey's declaration of the

B.C. L. REV. 733 (2001); Allan J. Samansky, *Tax Consequences When Churches Participate in Political Campaigns*, 5 GEO. J.L. & PUB. POL'Y 145 (2007); Donald B. Tobin, *Political Campaigning by Churches and Charities: Hazardous for 501(c)(3)s, Dangerous for Democracy*, 95 GEO. L.J. 1313 (2007); Ellis M. West, *The Free Exercise Clause and the Internal Revenue Code's Restrictions on the Political Activity of Tax-Exempt Organizations*, 21 WAKE FOREST L. REV. 395 (1986); Joel E. Davidson, Note, *Religion in Politics and the Income Tax Exemption*, 42 FORDHAM L. REV. 397 (1973); Joseph S. Klapach, Note, *Thou Shall not Politic: A Principled Approach to Section 501(c)(3)'s Prohibition of Political Campaign Activity*, 84 CORNELL L. REV. 504 (1999); Judy Ann Rosenblum, Note, *Religion and Political Campaigns: A Proposal to Revise Section 501(c)(3) of the Internal Revenue Code*, 49 FORDHAM L. REV. 536 (1981).

102. *See Regan*, 461 U.S. at 544.

103. *Id.* at 545.

104. *See* SURREY & MCDANIEL, *supra* note 4, at 122.

"axiomatic" equivalence of tax expenditures and direct cash subsidies for purposes of constitutional law is unsustainable.

A. The Question of Budgetary Purpose

Perhaps the greatest utility of tax expenditure theory, at least if history is an insightful guide, is its use in budgeting. Surrey emphasized that, because tax expenditures had long existed as hidden government subsidies, they should be brought to light and explicitly considered in the budgeting process.[105] Congress and the Treasury responded positively to Surrey's persuasion.[106] Now, tax expenditure budgets are published annually.[107] Both government officials and the general public can readily observe what Surrey argued was just another variant of government spending.

At a minimum, the publication of tax expenditure budgets should equip the federal government to better decide how much to budget for direct appropriations.[108] For example, in deciding the types of green energy projects to subsidize directly and the magnitude of green energy grants, Congress and the OMB could identify what indirect subsidies already exist through various income tax credits by scrutinizing the tax expenditure budgets. Although the tax expenditure budget does not take into account expected behavioral changes by taxpayers were a specific incentive eliminated, the budget does provide at least a reasonable idea of taxpayer activities that government is already incentivizing and the degree to which government is foregoing revenues in order to stimulate those activities.[109] This information is surely instructive to government officials in deciding whether and how much to spend directly to stimulate

105. *See* Surrey & Hellmuth, *supra* note 64, at 528 (recounting that Surrey had identified the need for a "full accounting" of tax expenditures by the late 1960s).

106. For an overview of the Budget Act and its treatment of tax expenditures, see SURREY & MCDANIEL, *supra* note 4, at 45-47.

107. *See supra* notes 13-21 and accompanying text.

108. Surrey argued that the OMB should better coordinate tax expenditures and direct spending programs. *See* SURREY & MCDANIEL, *supra* note 4, at 33.

109. *Cf.* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 20 ("Coordinated reviews of tax expenditures with related federal spending programs . . . could help policymakers reduce overlap and inconsistencies and direct scarce resources to the most effective or least costly methods to deliver federal support.").

these and other taxpayer activities. Even if some provisions of the IRC, currently denominated tax expenditures, are more accurately classified otherwise, the tax expenditure budget still conveys information helpful in establishing budget priorities.[110]

The role of the tax expenditure concept in the budgeting process is apparent.[111] Surrey even once identified this role as the purpose of the tax expenditure budget.[112] However, this budgetary function has little or no relevance to the courts in deciding constitutional issues.[113] Courts obviously do not have the power to tax,[114] but only to interpret tax statutes and regulations and to decide their constitutionality. Relatedly, courts do not bear responsibility for disbursing funds to advance public policies.[115] For example, whether Congress should spend more to defray the cost of higher education tuition, given current tax incentives for the same,[116] is not for judges to decide. Perhaps the tax expenditure concept would be relevant to a court in deciding the proper scope of legislation that imposed spending limits on certain categories of expenditures. But beyond that, it is difficult to link the budgetary value of tax expenditure analysis to the judicial task.

This modest point hardly establishes that the tax expenditure concept is irrelevant in court. But it does suggest a general qualification to consider, and perhaps even a presumption to avoid, in applying tax expenditure theory. That the tax

110. *Cf.* Sugin, *supra* note 9, at 415 (describing tax expenditure theory as "an immensely important policymaking tool"). Indeed, Professors David Weisbach and Jacob Nussim have urged rejecting the "normative consequences" of the tax expenditure label in favor of asking what information is "useful." *See* Weisbach & Nussim, *supra* note 52, at 976.

111. *See* JCT TAX EXPENDITURES REPORT, *supra* note 18, at 2 ("Estimates of tax expenditures are prepared for use in budget analysis.").

112. *See* Surrey & Hellmuth, *supra* note 64, at 530.

113. *Cf.* Sugin, *supra* note 9, at 413 ("[T]ax expenditure analysis is well suited to legislatures but not to courts.").

114. Only Congress has the federal power to tax. *See* U.S. CONST. art I, § 8.

115. *Cf.* Sugin, *supra* note 9, at 417 (stating that courts "do not get to pick and choose from among a variety of policy alternatives" in ruling on legislation).

116. *See, e.g.*, I.R.C. § 25A (granting the American Opportunity Tax Credit and the Lifetime Learning Credit); I.R.C. § 117 (excluding qualified scholarships from gross income); I.R.C. § 127 (excluding reimbursement of employee education expenses from gross income); I.R.C. § 529 (providing numerous tax benefits to qualified tuition programs, their beneficiaries, and their donors).

expenditure idea importantly aids the legislative and executive bodies in performing budgetary functions does not imply that the concept equally, analogously, or even meaningfully informs the judicial function in nonbudgetary matters.[117] To determine the relevance of tax expenditure theory in constitutional cases, one must scrutinize its features, premises, and limits.

B. The Question of the Normative Tax Base

The reliance of tax expenditure theory on a normative baseline has generated enormous commentary.[118] Although the difficulties of establishing the normative tax base surely have budgeting implications, they present unique challenges for courts, especially in constitutional cases. This Section explores these challenges. It first discusses several indeterminacies of the normative tax base. It then discusses how resolving these indeterminacies poses special problems for judges. Finally, this Section identifies a systemic problem that tax expenditure theory foists on the judiciary.

1. Indeterminacies of Surrey's Own Admission

The starting point for appreciating the difficulties that judges face because of the indeterminacy of normative baselines under tax expenditure theory is Surrey and McDaniel's own words. The uncertain boundaries of the normative income tax base, for example, are not merely the imaginary goblins of Surrey's critics. The tax expenditure analysis of Surrey and McDaniel with respect to the federal income tax suggests that identifying tax expenditures is fraught with indeterminacy at the margin.

Surrey and McDaniel argue that the Haig-Simons definition is the "accepted norm" in countries with a modern income tax, but they acknowledge that this concept "covers only basic aspects

117. *Cf.* Sugin, *supra* note 9, at 412 (distinguishing the utility of tax expenditure analysis for policymakers from its more limited relevance to courts).

118. *See, e.g.*, Henry Aaron, *What Is a Comprehensive Tax Base Anyway?*, 22 NAT'L TAX J. 543, 547-48 (1969); Bittker, *supra* note 60, at 925; Bittker, *"Tax Subsidies," supra* note 56, at 251; Bittker, *A Reply*, *supra* note 69, at 538; Crane, *supra* note 70, at 185; Fleming & Peroni, *A Critique*, *supra* note 52, at 142; Surrey & Hellmuth, *supra* note 64, at 529-33; Thuronyi, *supra* note 65, at 1163-70.

and a few details."[119] They further recognize that numerous factors bearing upon the income tax "have produced numerous questions of detail, some of them involving quite difficult classification questions."[120] To construct a complete list of tax expenditures thus requires "an extension" of the Haig-Simons concept to address issues that have arisen since the initial articulation of the concept.[121] But just how, and how far, to extend Haig-Simons is unclear.

Another question is the significance of public consensus as to the ideal tax base. Surrey and McDaniel state that the application of the Haig-Simons definition in federal income tax law is "tempered" by "the generally accepted structure of an income tax."[122] Apparently, establishing the normative tax base requires a determination of just what is "generally accepted." Determining this acceptance, and how much of it is sufficient to rise to the requisite level of generality, remains elusive. As discussed more fully below, "general acceptance" for Surrey and McDaniel ultimately means public consensus, although they appear loathe to say so explicitly.[123] They tend to defer to the judgment of the agency legislatively tasked with compiling the tax expenditure budget, but they never explain the connection between compiling that list and discerning public acceptance.

---

119. SURREY & MCDANIEL, *supra* note 4, at 5.

120. *Id.*

121. *Id.*

122. *Id.* at 4 (quoting U.S. DEP'T OF THE TREASURY, ANNUAL REPORT OF THE SECRETARY OF THE TREASURY ON THE STATE OF THE FINANCES FOR FISCAL YEAR ENDED JUNE 30, 1968, at 327 (1969)).

123. It is not abundantly clear under Surrey's explication of tax expenditure theory just whose notions of the public's perception of the normative tax base should control. The first entity that comes to mind is Congress, for its members are elected by the public. At times, Surrey seems to charge Congress with implementing public perception. *See, e.g.*, Surrey & Hellmuth, *supra* note 64, at 537. But to rely on Congress to implement the norm or norms controlling the definition of income and the structure of the federal income tax system is potentially problematic for Surrey; doing so lends credence to the positivism of Bittker, who suggests that the normative tax base is what Congress subjects to tax. Another candidate for discerning and advancing the public's perception of the normative tax base is the executive branch. Surrey did urge great deference to the Treasury and OMB. *See* SURREY & MCDANIEL, *supra* note 4, at 113-15. If the Treasury Department and OMB must announce the public's idea of the normative income tax, the determination is made by appointees not directly accountable to the voters whose acceptance is key. One may reasonably question administrative agencies' ability to discern and implement popular conceptions of the income tax base.

Regardless of who has responsibility for gleaning public acceptance of the normative income tax, how to harvest the knowledge of general acceptance is also a mystery.

Thus, Surrey and McDaniel themselves identify several respects in which tax expenditure theory is indeterminate: (1) some items are difficult to classify under the Haig-Simons ideal; (2) the Haig-Simons ideal is not itself an absolute baseline, for it requires extension; and (3) the Haig-Simons ideal must be modified to some degree to better comport with public acceptance of the tax base.[124] The third point of indeterminacy is further complicated by the absence of any clear method for discerning public acceptance of the base.

The point is not that these indeterminacies render tax expenditure theory useless. The point is that these indeterminacies are real. In constitutional cases that turn in part on whether tax provisions are properly viewed as indirect subsidies, judges inevitably must consider the implications of these indeterminacies.

2. Indeterminacies of Textual Structure in Resolving Indeterminacies

A judge who must decide the constitutional implications of a tax provision that may or may not qualify as a tax expenditure faces a formidable task. Courts commonly analyze the text and structure of a statute to determine its meaning. If deciding whether a statutory provision is a tax expenditure were simply a matter of interpreting a statute's text and structure, the task would readily fall within the competence of judges. Unfortunately, identifying the proper classification of a tax provision is not so simple.

The structure of legislative text—what it sets forth as the general rule, followed by exceptions—is of limited use in identifying tax expenditures. Consider a general rule that taxes compensation income, followed by exceptions that exclude the portion of compensation income that the taxpayer invests in special accounts. Taken together, for a taxpayer whose sole

---

124. *See* SURREY & MCDANIEL, *supra* note 4, at 187-88.

source of income is compensation, the provisions have the effect of taxing the taxpayer's consumption, not income.[125] If provisions governing other forms of income are drafted similarly, the legislature's normative tax base is apparently consumption, not economic income. That the legislation structures the exclusions as "exceptions" to general provisions does not mean that the "special exceptions" should be conceptualized as deviations from the ideal base.[126]

As argued by Boris Bittker, a discomfiting circularity exists in announcing that special deductions, exclusions, etc., are deviations from a base only after first assuming a broad base from which deviations tautologically follow.[127] To illustrate, assume the legislature of a newly admitted state decides to raise revenue by imposing a state tax on at least some property. The structure of the legislation sets forth a general rule that all property located in the state or owned by a resident natural or legal person is subject to the state tax. But there are exceptions to the general rule. One exception is for personal property. Another exception is for any property owned by individuals and trustees of trusts. A third exception is for any property owned by nonprofit entities. If the tax base is conceptualized as "all property located in the state or owned by a resident natural or legal person," then each of the nominal "exceptions" is a tax expenditure. However, if the tax base is conceptualized as real property owned by business entities, no exception described above is a tax expenditure. The nominal exceptions are really just structural mechanisms for specifying the tax base.[128]

As Surrey recognized, which concept of tax expenditures is best is a question of the normative "image" one has in mind for the tax.[129] But more is required than simply announcing that the exceptions are objectionable because they are inconsistent with one possible "image"—all property located in the state or owned

---

125. If a taxpayer can deduct savings from income, the resulting tax base is consumption. *See* Fleming & Peroni, *supra* note 9, at 508-09.

126. *Cf.* Edward A. Zelinsky, *Are Tax "Benefits" for Religious Institutions Constitutionally Dependent on Benefits for Secular Entities?*, 42 B.C. L. REV. 805, 836 (2001) ("[M]uch tax exemption is best understood as base defining.").

127. *See* Bittker, *Churches*, *supra* note 26, at 1304.

128. *See id.* at 1291.

129. *See* Surrey & Hellmuth, *supra* note 64, at 537.

by a resident. To do so is to engage in circular reasoning, for the selection of the "image" dictates whether the "exceptions" are (or at least might be) tax expenditures. If the alternative image—real property owned by business entities—better captures the thoughts of all who play a role in adopting the tax, the "exceptions" are normatively compelled, not tax expenditures.[130]

Thus, an analysis of the structure of a text cannot establish the normative tax base. Exceptions in form could represent deviations from the ideal base. However, they just as likely could be the drafter's mechanisms for arriving at the normative base.

3. Indeterminacy in How to Resolve Indeterminacies

The previous discussion identifies several indeterminacies in establishing or recognizing a normative tax base. However, a still more vexing problem remains, one that has not received adequate attention by courts or commentators: precisely whose judgment should control the resolution of whether a tax provision is properly conceptualized as a tax expenditure in a constitutional case? This question is basic, yet it is trickier than meets the eye.

Conducting an independent assessment of what should be the normative tax base is hardly the proper role of the judiciary, so that approach can be dismissed summarily. Another option, well within the competency of the judiciary, is to analyze the text and structure of tax legislation in hopes of determining whether a provision constitutes a tax expenditure in the judgment of Congress. However, for reasons already explained, analyzing the structure of a statute is doomed to produce indeterminacy, for a provision enacted as a statutory exception may simply be the mode of defining the normative tax base.[131] Moreover, the text of a statute never begins, "Congress hereby pronounces the following a tax expenditure."[132]

130. Professor Bittker argued similarly with respect to the taxation of real property. *See* Bittker, *A Reply*, *supra* note 69, at 540.

131. *See supra* Section II.B.2.

132. The Budget Act, which requires the compilation of a tax expenditure budget, does not define the normative base. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 3.

A more promising idea is that a judge should decide what a legislature intends as the normative tax base in view of all available evidence. However, even that inquiry is problematic. Because tax statutes do not explicitly announce the precise normative base, courts must look elsewhere, such as committee reports and congressional hearings, for clues. Courts frequently consult legislative history in deciding tax cases;[133] doing the same to discern the presence of a tax expenditure is within the realm of judicial competence. But, commonly, these legislative sources do not reveal Congress's view of the normative tax base. Further, even if legislative history does speak to the normative tax base, the familiar critiques of trying to mine the intent of the full Congress by pulling nuggets from committee reports and statements of select lawmakers cast at least some doubt on how determinative such a judicial approach really is. Moreover, a subsequent Congress may let a previously enacted tax provision stand not because they agree with the prior Congress's rationale for enacting it as a subsidy, but because they view the provision as reflecting the normative tax base. No legislative history would necessarily document that determination.

As another option, one favored by Surrey,[134] courts could just defer to the judgment embodied in governmental tax expenditure budgets.[135] For example, in the case of the federal income tax, the tax expenditure budget prepared by the JCT or by the Treasury Department could receive authoritative status. At least this approach provides objective answers. Further, career legislative staffers and Treasury officials (and their counterparts at the state level) surely have a better idea of the normative tax base than most judges. But this approach presents its own problems. It essentially confers on either legislative committee staffers or executive agencies authority to resolve one of the crucial questions that must be analyzed by a court in deciding constitutional cases involving tax provisions that may or may not be indirect subsidies. Absolutely deferring to the opinions of nonjudicial public servants on matters requiring constitutional

133. *See, e.g.*, Harrison v. N. Tr. Co., 317 U.S. 476, 479 (1943).

134. *See* SURREY & MCDANIEL, *supra* note 4, at 119.

135. This approach is possible only when a tax expenditure budget exists, of course.

judgment should give any judge pause, even though the Senate Budget Committee has stated that tax expenditures "may, in effect" be viewed as a government outlay.[136] After all, courts ultimately decide constitutional questions, not Congress, and not the executive branch.[137] The prospect that an executive agency may politicize the compilation of tax expenditure lists further counsels against absolute judicial deference to their determinations.[138]

Another alternative is for a court independently to assess an open host of factors to discern the "image" of the normative tax base that appears to underlie the statutorily enacted tax. That sounds more like metaphysics than constitutional law. A court might try it, but it is hardly likely to generate the ideal tax base envisioned by the public—if it exists at all.

Thus, one returns to the observation that began this Section. Deciding whose substantive judgment should receive controlling weight in determining whether a tax provision is properly classified as a tax expenditure for purposes of constitutional law is complex. As a formal matter, courts have the final say. But how they decide cases ultimately is affected by their willingness to defer to executive agencies and legislative committees, as well as by their interest in exploring legislative intent, statutory design, and other potentially relevant factors.

---

136. *See id.* at 119, 262. Surrey & McDaniel place great weight on governmental tax expenditure budgets. They go so far as to say that courts would be "wrong" to treat direct grants differently from tax expenditures in view of how the latter are reflected in the budget and are conceptualized by the Senate Budget Committee. *See id.* at 119. But the argument is unconvincing. That the tax expenditures budget describes tax expenditures as an "alternative" to other forms of aid, including direct assistance, in no way establishes that they are legally equivalent. A fine may be an alternative to incarceration, but the two are not the same. The statement of the budget committee that tax expenditures "may, in effect, be viewed" as the equivalent of taxation and spending is a far cry from insisting on their legal equivalence. The statement is tentative, non-descript as to who is doing the viewing, and focused only on a first-order economic effect (not design, and not real-world consequences). Further, the OMB, the Treasury Department, and the Senate Budget Committee do not speak for the full Congress on the nature of tax expenditures.

137. *See* Marbury v. Madison, 5 U.S. (1 Cranch) 137, 178-80 (1803).

138. *See* Sugin, *supra* note 9, at 426.

4. Why Indeterminacy Matters Greatly in Constitutional Cases

The foregoing discussion highlights the indeterminacy of the normative base and list of tax expenditures, as well as the difficulties courts face in resolving those indeterminacies.[139] For Congress and the executive branch, this indeterminacy is hardly a fatal problem.[140] Members of each of these two branches of government can decide whether and to what extent the appearance of an item in a tax expenditure budget is relevant to their spending priorities and proceed accordingly. That is their business.

But the business of courts is different, and the indeterminacy of the normative tax base is more troublesome when a court must decide the constitutional implications of a tax expenditure.[141] If a provision is truly best conceptualized as a deviation from the normative tax, it is plausible to view it as a form of government subsidy. Once a court has determined that an indirect subsidy exists, the analysis can then proceed to an evaluation of the constitutional implications of this indirect subsidy. But if the provision is best conceptualized as a structural mechanism for arriving at the normatively correct tax base, treating it as a subsidy is error. Erring when constitutional rights and duties are at stake is more disturbing than merely mischaracterizing an item when compiling a budget.

This issue is not one that a court will rarely encounter. Surrey recognized the existence of many details of a tax for which there is no easy test of normativity.[142] Courts decide cases involving such details. Indeed, hard cases routinely come before a court. Courts must therefore be prepared to closely examine a provision to determine whether it is plausibly conceptualized as a

---

139. *Cf. id.* at 419 (referring to the "theoretical impossibility" of identifying tax expenditures); GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 3 ("Determining whether a tax code provision meets the definition of a tax expenditure requires judgment.").

140. *See* Sugin, *supra* note 9, at 416-17.

141. *See id.* at 413 ("The definitional difficulties inherent in the tax expenditure concept . . . make tax expenditure analysis too unreliable for constitutional adjudication . . . ."); *id.* at 417 (stating that the difficulties of defining the tax base "make tax expenditure analysis an inadequate guide for deciding individual lawsuits").

142. SURREY & MCDANIEL, *supra* note 4, at 5.

mechanism for implementing a normatively correct tax or instead as an indirect form of subsidy.

For example, if the charity income tax exemption is properly conceptualized as (in part) a federal grant to churches to carry out their general religious purposes, the exemption raises questions under the Establishment Clause.[143] But this conceptualization of the exemption is problematic. A plausible argument exists that the income tax by design primarily reaches natural persons, directly or indirectly (i.e., through ownership of business entities),[144] and that taxing charitable entities is inappropriate.[145] Moreover, although Surrey eventually concluded otherwise,[146] the charity income tax exemption is not classified as a tax expenditure by the Department of the Treasury or the JCT,[147] a fact that, under the JCT's methodology, implies its conclusion that no reasonable basis exists for a contrary classification.[148]

---

143. *See id.* at 132; Adler, *supra* note 25, at 912-14. For an excellent study of the Founding era understanding of church taxes and the Establishment Clause, see generally Mark Storslee, *Church Taxes and the Original Understanding of the Establishment Clause*, 169 U. PA. L. REV. 111 (2020) (arguing that the history of publicly funding religious schools while objecting to the payment of church taxes means only that the Founding generation understood disestablishment to preclude funding that was specifically aimed at advancing religion, not funding intended to supply broader public goods provided by religious persons or entities). For an argument that tax exemptions for religious entities are constitutional, see Zelinsky, *supra* note 126, at 807 ("[T]ax exemption does not subsidize churches, but leaves them alone.").

144. Professor Steven Bank has argued that policymakers historically viewed the tax on corporate income as an indirect tax on shareholders. *See* Steven A. Bank, *Entity Theory as Myth in the Origins of the Corporate Income Tax*, 43 WM. & MARY L. REV. 447, 527-33 (2001).

145. *See* Bittker, *Churches*, *supra* note 26, at 1289-91; Bittker, *"Tax Subsidies," supra* note 56, at 256. Professor Bittker later argued for the federal income tax exemption of most nonprofit entities under the theory that their income is not measurable under conventional concepts. *See, e.g.*, Boris I. Bittker & George K. Rahdert, *The Exemption of Nonprofit Organizations from Federal Income Taxation*, 85 YALE L.J. 299, 307 (1976). In response, Professor Henry Hansmann has argued that the conceptual difficulties that Bittker and Rahdert raise are not insurmountable. *See* Henry Hansmann, *The Rationale for Exempting Nonprofit Organizations from the Corporate Income Tax*, 91 YALE L.J. 54, 58-62 (1981).

146. *See, e.g.*, SURREY & MCDANIEL, *supra* note 4, at 219-20 ("[T]he U.S. tax treatment of nonprofit organizations should be classified as a tax expenditure. . . . Because it is likely that the revenue cost of the exemption for nonprofit organizations is substantial, the omission from the U.S. tax expenditure lists is a serious one and should be rectified.").

147. *See* JCT TAX EXPENDITURES REPORT, *supra* note 18, at 9.

148. *See id.* at 2 ("A provision traditionally has been listed as a tax expenditure by the Joint Committee staff if there is a reasonable basis for such classification and the provision results in more than a *de minimis* revenue loss . . . .").

Further, the charity income tax exemption has a long history in this country, suggesting—to borrow from Surrey's test of public acceptance—that both the Congress and the public think of the charity income tax exemption as a standard feature of the federal income tax system.[149] In the view of this Author, the better view is that the income of charities—at least certain categories of them—is not in the first instance properly included in the federal income tax base.[150] But the question is debatable, and thoughtful analysts reach opposite conclusions.[151] A court that fails to wrestle with the proper characterization of the exemption risks reaching an erroneous decision. And the consequences of reaching a wrong decision are hard to rectify once they become constitutionally embedded by the Supreme Court.

When congressional staffers or Treasury officials mistakenly characterize a tax provision as a tax expenditure, it tends to cause more open scrutiny, and hence public debate, over the provision. When courts make the same mistake in a constitutional case, it often will likely remove the provision from legislative debate because of mootness; such mischaracterization may render the provision unconstitutional. When congressional staffers and Treasury officials err in their judgment of what

149. Bittker and Rahdert explain that charity income tax exemptions date from the Revenue Act of 1894, and they were reenacted in subsequent legislation (including the Revenue Act of 1913, which imposed a federal income tax after the adoption of the Sixteenth Amendment). *See* Bittker & Rahdert, *supra* note 145, at 301-04.

150. *See* Buckles, *supra* note 82, at 947, 979 (explaining that the charity income tax exemption and the charitable contributions deduction may reflect a decision not to tax income attributed to the community in general); *see also* Evelyn Brody, *Of Sovereignty and Subsidy: Conceptualizing the Charity Tax Exemption*, 23 J. CORP. L. 585, 586 (1998) (arguing that § 501(c)(3) reflects a governmental recognition of a charity's sovereign prerogative to operate free from governmental intrusion).

151. A number of scholars embrace some form of subsidy theory of the charity income tax exemption and the charitable contributions deduction. *See, e.g.*, Miranda Perry Fleischer, *Libertarianism and the Charitable Tax Subsidies*, 56 B.C. L. REV. 1345, 1351 (2015) ("[M]ost tax theorists consider sections 170(c) and 501(c)(3) subsidies for the charitable sector."); *see also* Rob Atkinson, *Altruism in Nonprofit Organizations*, 31 B.C. L. REV. 501, 601, 605, 610 (1990); Nina J. Crimm, *An Explanation of the Federal Income Tax Exemption for Charitable Organizations: A Theory of Risk Compensation*, 50 FLA. L. REV. 419, 430-32 (1998); Mark A. Hall & John D. Colombo, *The Donative Theory of the Charitable Tax Exemption*, 52 OHIO STATE L.J. 1379, 1383-88 (1991); Hansmann, *supra* note 145, at 54-55. Professor Rob Atkinson has mostly renounced his altruism theory. *See* Rob Atkinson, *Tax Favors for Philanthropy: Should Our Republic Underwrite de Tocqueville's Democracy?*, 6 WM. & MARY POL'Y REV. 1, 48-49 (2014).

belongs in the normative tax base, the result may simply mean less direct funding for the "favored" activity. When courts err similarly, they may wrongly subject income to taxation, and hence penalize the activity generating the income—even when the item involves constitutionally protected activity. When congressional staffers or Treasury officials make mistakes, their counterparts who serve a few years later can correct those mistakes. But the stability that is generally desirable in constitutional case law also tends to fossilize a court's mistakes in characterizing tax provisions as indirect subsidies.

This discussion of the normative tax base demonstrates that tax expenditure theory poses special challenges for courts, including those that are generally underemphasized by judges. Determining the normative base is complex, especially at the margin,[152] and reaching the wrong result in constitutional cases is uniquely troubling.[153]

5. The Judicial Problem of Assuming an Extra-Statutory Normative Tax Base

One final problem concerning the normative tax base looms large in constitutional cases. The preceding discussion highlights the challenges of determining tax expenditures at the margin. The gravest problem, however, is more fundamental. Tax expenditures exist only when tax provisions deviate from an assumed normative base. Just as St. Augustine conceived of evil as a privation of the good, Surrey conceived of tax expenditures as a privation of the normative tax base.[154] If federal income tax law reflected the normative income tax base, no tax expenditures would exist. To assert the existence of tax expenditures under the federal income tax is to assert non-normativity in the law as it has

152. Professor Sugin concludes that courts have often "wisely resisted" tax expenditure analysis because of its ambiguities. *See* Sugin, *supra* note 9, at 418.

153. *See id.* at 419.

154. *See* AUGUSTIN, *Book Third*, *in* A SELECT LIBRARY OF THE NICENE AND POST-NICENE FATHERS OF THE CHRISTIAN CHURCH 60, 64 (Philip Schaff ed., Buffalo, The Christian Literature Co. 1886). The analogy to Augustine's view of sin does not mean that Surrey judged all tax expenditures as "bad" policy. Surrey did not do so. *See* SURREY & MCDANIEL, *supra* note 4, at 5-6. The analogy to Augustine compares analytical interrelationships, not moral or policy judgments as to the advisability of a tax expenditure.

been enacted by Congress and the President. Similarly, to characterize various provisions of a law implementing another tax system as tax expenditures is to maintain that the enacted tax base is non-normative.

Thus, to insist that courts identify certain tax provisions as tax expenditures and then declare them as equivalent to cash subsidies for purposes of constitutional analysis is to demand that courts constitutionalize an image of a normative tax base that lawmakers have decided not to enact. Speaking of the "normative tax base" is a fancy way of saying "the tax base that the legislature should adopt," or at least the tax base that they "should adopt under tax policy norms apart from other policy goals."[155] Reduced to its essence, the assertion that courts deciding constitutional cases must equate tax expenditures with cash subsidies is tantamount to the claim that courts must decide constitutional cases according to what lawmakers should tax but decided not to tax.[156] This methodology smacks of requiring courts to think like super-legislatures.[157]

---

155. Some commentators have tried to escape this conclusion, but in the opinion of this Author, it is inescapable. To illustrate, as a thought experiment, let us replace the phrase "should adopt" with "could adopt" and test its coherence. To say that a tax expenditure is equivalent to taxing and spending merely because the legislature "could adopt" the broader base would subvert the entire tax expenditure enterprise. If "could adopt" means what a legislature could actually accomplish in the real world, one would be forced to ponder whether enacting the broad base is politically viable. Tax expenditure theory has never tried to make that showing. On the other hand, if "could adopt" means only that a legislature hypothetically could enact the normative base, tax expenditure theory becomes nondirective. A legislature "could adopt" a base so broad that it taxes the imputed income from leisure and household services, or even that attributable to the taxpayer's decision to stay in a more satisfying job than one in which the taxpayer could command a higher salary. Any decision not to tax income under this base that "could be" would then be a tax expenditure. But few tax expenditure champions would accept this approach as a credible way of identifying tax expenditures. Thus, classical tax expenditure theory does not contemplate merely the base that a legislature "could adopt." It assumes a base that conforms to the normative ideal, the base that "should be" under the proper implementation of tax policy norms.

156. Some would prefer a less poignant phrasing of the essential inquiry, perhaps akin to "the tax base that lawmakers should adopt strictly under tax policy norms." But that base is also hypothetical, one that lawmakers did not actually enact. So the more polite inquiry still asks courts to find a subsidy based on an ideal that the legislature refused to implement under the actual tax.

157. Surrey's recommendation—that judges (at a minimum) accept whatever the OMB and Senate Budget Committee declare to be tax expenditures—is not a real solution. *See* SURREY & MCDANIEL, *supra* note 4, at 119. That approach just shifts super-legislature status to the executive branch or a single congressional committee (or both).

It is reasonable for the Senate Budget Committee to publish a tax expenditures budget to inform congressional colleagues of what it believes is equivalent to indirect subsidies. It is also sensible for the Department of the Treasury to inform the Congress and the public of what it considers to be the normative ideal, and thus the indirect subsidies that result from deviations from that ideal. But it does not follow that the courts should be making constitutional law according to what someone thinks the tax base should be when no lawmaking body has adopted it.

This point is not the same as the reductionist claim that a specific credit or deduction is merely a refusal to tax. The greater problem is systemic. The intrinsic problem of equating all tax expenditures with direct cash subsidies in constitutional cases is that it forces courts to ground their analysis on an idea of a tax base that "should be" rather than encouraging them to scrutinize the tax base that "is." How a legislative body has crafted the tax base may well be constitutionally suspect. But requiring a court to resolve a constitutional question according to an unenacted tax ideal is suspect.

C. The Question of Shifting Consensus

The previous Section introduced Surrey's reliance on public acceptance as a legitimate constraint on implementing the Haig-Simons concept of income under the federal income tax.[158] No good reason exists to confine his argument to the income tax; presumably, Surrey would maintain that tax expenditure analysis of any tax base should also account for public acceptance. As explained previously, this reliance on public acceptance in tax expenditure theory poses challenges to constitutional analysis.[159] This Section expounds on this problem at a deeper level.

That Surrey relied on public acceptance is manifest in his writings, and serious students of tax expenditure theory must not brush the point aside merely because it is inconvenient. Surrey distinguished between (1) an economic concept of income, and (2) "widely accepted definitions of income" and the "generally

---

158. *See supra* Section II.B.1.
159. *See supra* Section II.B.4.

accepted structure of an income tax."[160] He illustrates the distinction between the two with imputed income. Although economic income includes imputed income from a taxpayer's property and self-provided services, in the United States, these are "not yet within the general understanding of the proper structure of an income tax."[161] Excluding imputed income from the tax base is considered "normative" in Surrey's tax expenditure analysis because it has "not been commonly regarded as income for tax purposes."[162] Stated another way, imputed income is "not yet within the general understanding of the proper structure of an income tax."[163] Surrey similarly writes that people "would in general be puzzled by the inclusion of the exemption of gifts and bequests as a tax expenditure."[164] It is difficult to perceive why an economist, or even a member of Congress, would find the inclusion of gifts in the normative income tax base "puzzling." He also refers to the initial stage of developing the margins of an income tax structure, when "most people" would reject a proposal based on some economic concept of income.[165] Thus, when Surrey speaks of "common regard," "wide acceptance," "most people," and "general acceptance," he is not likely speaking merely of the views of legislators, legal experts, or economists. He apparently means the consensus of the general public.[166] For Surrey, tax expenditure theory does not stop with the Haig-Simons ideal. Tax expenditure theory "tempers"[167] the ideal by treating as normative those forms of income, and presumably deductions, that at least a majority of the general public in the United States accept as proper for an income tax.

Surrey's distinction between economic income and the public's view of a proper income tax base amplifies the concerns

---

160. Surrey & Hellmuth, *supra* note 64, at 532; *see also id.* at 528 (quoting a speech by Surrey in which he described tax expenditures in terms of "deliberate departures from accepted concepts of net income").

161. *Id.* at 532.

162. SURREY & MCDANIEL, *supra* note 4, at 4.

163. Surrey & Hellmuth, *supra* note 64, at 532.

164. *Id.*

165. *Id.*

166. Apparently, Surrey thought legislators should consider the views of the general public as they enact tax legislation. He placed responsibility for developing the proper "image" of the tax structure on legislators. *See id.* at 537.

167. *Id.* at 531-32.

of the previous Section regarding judicial deference to another governmental branch's taxonomy of tax expenditures.[168] But Surrey's appreciation for the importance of public acceptance raises another issue. Public perceptions are not static, a fact that Surrey recognizes and even seems to welcome.[169] He writes that this "standard of general acceptance of course results in changes over time" as the public warms to concepts of economic income.[170] Accordingly, the scope of the income tax structure is "an evolutionary matter."[171] The import of this analysis is that, as the general public increasingly embraces economic concepts of income but the law does not yet reflect this understanding with a broader tax base, the list of tax expenditures correspondingly increases.[172]

Thus, at least for Surrey, the list of tax expenditures (1) depends in part on public consensus, and (2) changes with public opinion of the proper income tax base and structure. These finer points of crafting the tax expenditure budget present no major hurdles for Congress or the executive branch. They can revise the budget as they see fit, according to their perceptions of public opinion.[173]

But these finer points pose some difficulties for constitutional adjudication. First is the institutional position and role of the courts. Unlike Congress, federal courts are largely insulated from popular consensus on matters of public concern. If Congress aggressively legislates in a direction contrary to public opinion, the public tends to vote their representatives out of office. This reality is constitutionally designed.[174] But the federal judiciary is constituted independently of the ballot box—except in the indirect sense of initially being nominated and

168. *See* discussion *supra* Section II.B.3.

169. *See* SURREY & MCDANIEL, *supra* note 4, at 198-99.

170. Surrey & Hellmuth, *supra* note 64, at 532.

171. *Id.*

172. *Cf. id.* (discussing the "second stage" of a change in the taxation of an income item, in which the economic concept of the item is recognized as normatively proper, but its exclusion is retained in the tax law and classified as a tax expenditure).

173. *See* SURREY & MCDANIEL, *supra* note 4, at 198-99.

174. *See* U.S. CONST. art. I, § 2, cl. 1 ("The House of Representatives shall be composed of Members chosen every second Year by the People of the several States . . . ."); U.S. CONST. amend. XVII ("The Senate of the United States shall be composed of two Senators from each State, elected by the people thereof, for six years . . . .").

confirmed by elected officials. The independence of the federal judiciary is also constitutionally designed, presumably to encourage judicial decision-making that transcends popular opinion.[175] The idea that courts would resolve constitutional questions by deferring to public consensus on the normative tax base (or anything else) at any moment is problematic.[176]

Relatedly, deferring to public consensus on the boundaries of income—and hence the classification of a federal tax provision as a tax expenditure—risks a lack of stability not anticipated by Surrey. Surrey's writings reveal that he envisioned the evolution of tax expenditure theory in a linear fashion, as the public gradually embraces an economic concept of income.[177] Surrey did not explore other possible forms of evolutionary development of public opinion. One is that, over time, public opinion might vacillate in its acceptance of economic income. What becomes widely accepted in one span of two or three decades might change course over the next twenty to thirty years.[178] Another possibility is that public opinion might swell in support of other norms important to the design of the income tax, such as the ability-to-pay norm or administrative norms. As public consensus changes, these norms might move the income tax system away from one that broadly reflects economic income. In either case, the linear pathway of tax expenditure analysis imagined by Surrey would not materialize.[179]

Congress and the executive branch can respond to a lack of stability in the public's understanding of a normative tax base

175. *See* U.S. CONST. art. II, § 2, cl. 2 ("[The President] shall nominate, and by and with the Advice and Consent of the Senate, shall appoint . . . Judges of the supreme Court, and all other Officers of the United States, whose Appointments are not herein otherwise provided for, and which shall be established by Law . . . ."); U.S. CONST. art. III, § 1 ("The Judges, both of the supreme and inferior Courts, shall hold their Offices during good Behaviour . . . .").

176. Although judges and legal scholars hold different perspectives on the scope and value of originalism, Gallup polls—as an extreme example—should not control constitutional analysis.

177. *See, e.g.*, Surrey & Hellmuth, *supra* note 64, at 532.

178. In fact, since the zenith of Surrey's influence, federal income tax statutory law has moved away from a comprehensive income tax base in many respects. *See* Sugin, *supra* note 9, at 428-30.

179. Relatedly, elected public officials and their appointees may, for political reasons, alter tax expenditure lists, thereby promoting instability in the normal tax base. *See id.* at 424-27.

easily enough. They can modify federal tax laws, revise tax expenditure lists, and adjust spending priorities as needed. But if the courts were beholden to a public consensus that did not develop as Surrey imagined, but instead evolved unpredictably, perhaps even cyclically,[180] constitutional law would suffer. A lack of stability in the public's concept of income, and hence the list of tax expenditures, could promote the resolution of important constitutional issues in an erratic fashion over time if judges calibrated constitutional analysis to popular thought. The disparate outcomes would not result from changing judicial philosophies, but simply from whatever "general acceptance" happens to be, or is perceived to be, when a case makes its way to a court.[181]

In short, Surrey's willingness to account for the general public's acceptance of income tax norms in formulating the list of tax expenditures could compromise judicial independence and stability in constitutional doctrine if courts simply embraced at face value the tax expenditure budget of the day in deciding constitutional cases.

D. The Question of Legislative Purpose

Some constitutional questions require a court to examine governmental purpose in deciding whether governmental action violates the Constitution.[182] Caution is in order when a court must

---

180. Tax legislation tends to behave in a cyclical pattern—towards, and then away from, a norm. *See* Daniel Shaviro, *Beyond Public Choice and Public Interest: A Study of the Legislative Process as Illustrated by Tax Legislation in the 1980s*, 139 U. PA. L. REV. 1, 4 (1990).

181. Since Surrey first championed its use, the tax expenditure budget has not changed in a way that suggests a widespread public endorsement of economic income as the federal income tax base. Congress has continued to implement policies through provisions identified as tax expenditures, and the public tolerates—some might say even encourages—the practice. This chronic resort to enacting tax expenditures is some evidence that the public behind the Congress has little appetite for subjecting all economic income to taxation. To the contrary, the enactment of tax provisions that are more consistent with a consumption tax base (e.g., providing for the expensing and accelerated depreciation of business assets, and expanding methods for deferring income reserved for future retirement) suggests resistance to a comprehensive income tax base.

182. *See, e.g.*, Town of Greece v. Galloway, 572 U.S. 565, 591-92 (2014) (holding that a New York town did not violate the Establishment Clause by opening its board meetings with the prayers of local clergy on a rotating basis). In this case, the Court commented on

decide whether an unconstitutional legislative purpose to subsidize exists merely because an item has appeared on a tax expenditure list.

First, an item listed as a tax expenditure does not necessarily reflect legislative purpose to favor an activity by implementing a deviation from the normative tax base.[183] Tax expenditure budgets are published by the OMB (using the analysis of the Treasury Department) and the JCT.[184] However "legislative purpose" is deduced, a Congress that enacts a provision pronounced to be a tax expenditure by one or more of these governmental bodies may not have a purpose to enact a deviation from the normative tax base. As explained previously, the normative income tax base and structure are not perfectly defined.[185] Even Surrey backed away from the Haig-Simons model as absolutely controlling.[186] Certain provisions enacted into law simply may reflect a legislative judgment that some income items should be excluded from income, or some transfers should be deducted from income, to arrive at the proper income tax base.

For example, a Congress that expands the charitable contributions deduction may do so with the understanding that amounts dedicated to charitable purposes are not properly included in the normative tax base. Perhaps Congress is persuaded by the logic that personal consumption means purchased consumption enjoyed or controlled by a taxpayer, and thus it does not include charitable contributions.[187] Or perhaps the Congress supports the charitable contributions deduction as

governmental purpose as follows: "Absent a pattern of prayers that over time denigrate, proselytize, or betray an impermissible government purpose, a challenge based solely on the content of a prayer will not likely establish a constitutional violation." *Id.* at 585. For discussions of governmental purpose in constitutional cases, see Ashutosh Bhagwat, *Purpose Scrutiny in Constitutional Analysis*, 85 CALIF. L. REV. 297 (1997), and Calvin Massey, *The Role of Governmental Purpose in Constitutional Judicial Review*, 59 S.C. L. REV. 1 (2007).

183. *See* Zelinsky, *Tax "Benefits," supra* note 22, at 411 (distinguishing tax base determinations from purposive subsidies).

184. *See supra* notes 13-21 and accompanying text.

185. *See* discussion *supra* Section I.B.1.

186. *See* discussion *supra* Section I.B.1.; *see also* discussion *supra* Section II.B.1.

187. *See, e.g.*, Andrews, *supra* note 74, at 346.

normative on ability-to-pay grounds.[188] Or maybe, relying on Surrey's notion of general acceptance,[189] Congress perceives that the general public holds one or both of these views. Whether Congress is correct in its assessment is beside the point. Congress might be wrong. But when legislative purpose is what a Court must scrutinize, that Congress had no purpose to confer an indirect subsidy on charitable donors, but only to properly derive the income tax base, is surely relevant.

The force of this argument is not undermined by the response that, whatever congressional purpose may be, Congress knows that some taxpayers will benefit from the provision in question. An awareness that a provision benefits a taxpayer is not tantamount to an intent to subsidize the taxpayer through the income tax system. The point is easily illustrated by the deduction for trade or business expenses.[190] Obviously, taxpayers benefit from deducting the expenses of producing business income.[191] Congress surely recognizes this reality. But the deduction for trade or business expenses is not properly viewed as a "subsidy" administered through the income tax system. As Surrey rightly explained, the deduction is a necessary mechanism in implementing a normatively correct income tax.[192] When a court examines legislative purpose, the question is not simply whether Congress knows that a taxpayer benefits from a statutory provision in question.

Even when lawmakers adopt a mechanism with the aim of benefitting taxpayers through a deviation from what some legislators consider the normative tax structure, tax expenditure analysis does not resolve all relevant questions about legislative purpose. Consider again the federal income tax. Numerous tax

---

188. A taxpayer who has transferred wealth to a charitable organization during the year obviously has less wealth at the end of the year with which to pay taxes. That the taxpayer, prior to the transfer, had the power to consume or save the amount transferred does not alter this end-of-year financial reality. The real question for policymakers concerned with ability-to-pay is how to treat a taxpayer's voluntary reduction in wealth.

189. *See* discussion *supra* Section II.C.

190. *See* I.R.C. § 162.

191. *See* Fleming & Peroni, *A Critique*, *supra* note 52, at 142 (recognizing that § 162 is not a tax expenditure although it provides a benefit to taxpayers claiming the § 162 deduction).

192. *See* SURREY & MCDANIEL, *supra* note 4, at 222.

expenditures cannot credibly be justified as consistent with an income tax base informed by tax policy norms. One example is the allowance of a deduction for the full cost of a business machine in the year of purchase.[193] Although this provision is surely properly characterized as a tax expenditure in an income tax system,[194] it may well be that a majority of Congress in any given session support this provision because they favor a consumption tax over an income tax for broad policy reasons. Excluding the costs of business machines from the tax base is consistent with a consumption tax base.[195] For these legislators, the purpose of maintaining the expensing provision may be to move the tax system towards a consumption base for an increasing number of business taxpayers, not to subsidize business purchasers through a deviation from the normative income tax base.[196]

Thus far, the examples offered in this Article to establish the complexity of the legislative purpose inquiry have reflected an ambiguity in the tax base or a desire to move the system away from a recognized tax base towards another. However, the legislative purpose inquiry remains complex even when tax expenditure theory plainly points to an intent to subsidize. Consider the deduction for the excess of percentage depletion over cost depletion for certain mineral deposits.[197] Percentage depletion is theoretically incorrect under either a normative

---

193. *See* I.R.C. § 179.

194. *See, e.g.*, TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 435-41 (discussing the history of IRC § 179 and its possible policy justifications).

195. Savings and investments, whether business-related or not, are properly excluded from a consumption tax base. *See* William D. Andrews, *A Consumption-Type or Cash Flow Personal Income Tax*, 87 HARV. L. REV. 1113, 1149 (1974); Fleming & Peroni, *supra* note 9, at 508-09.

196. Many provisions characterized as tax expenditures are consistent with a consumption tax base. *See* Sugin, *supra* note 9, at 429. Our current system is a hybrid income-consumption tax in a number of respects. *See* Andrews, *supra* note 195, at 1120, 1128. *See generally* Edward J. McCaffery, *Tax Policy Under a Hybrid Income-Consumption Tax*, 70 TEX. L. REV. 1145 (1992) (arguing that a hybrid income-consumption tax system is likely superior to either an income tax or a consumption tax because the hybrid system can differentially treat life-cycle, precautionary, and bequest savings).

197. This difference is often a tax expenditure. *See* TREASURY REPORT, *supra* note 16, at 5.

income tax based on economic income or a consumption tax.[198] The provision surely reflects legislative intent to subsidize mineral extraction.[199] But discerning a legislative intent to subsidize does not end the purpose inquiry. Although a cynic might say the provision is designed merely to reward a powerful political lobby, perhaps a majority of Congress believe that this mechanism is necessary to ensure a degree of national energy security.[200] More broadly, even when tax expenditure analysis leads one to discern a purpose to subsidize an activity, the inquiry must proceed to explore the ultimate goal of the subsidy.[201] Properly placing an item on a tax expenditure list provides only modest insight as to the ultimate legislative purpose for enacting the provision. Similarly, that a good argument exists for identifying a provision as a tax expenditure, notwithstanding its current omission from the list, is not enough to glean the legislative purpose for its omission.

Thus, when a legislative purpose of a tax expenditure is indeed to subsidize, the design of the provision for purposes of constitutional law should be analyzed in its broader legislative context. All kinds of contextual factors might justify a subsidy.[202] To name just a few, perhaps the benefits reaped by taxpayers engaging in an indirectly subsidized activity are necessary to accomplish critical national goals that benefit a much broader segment of the public. Another possibility is that Congress has determined to subsidize an activity indirectly through tax law to

---

198. For a discussion of percentage depletion, see TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 113-18. Cost depletion is a form of capital recovery. *See id.* at 113. In contrast, percentage depletion allows a deduction for a fixed percentage of the gross revenues from sales of the mineral. Consequently, aggregate depreciation deductions over time typically exceed the actual capital investment. *See id.* at 113-14.

199. *Cf. id.* at 114 ("The difference between percentage depletion and cost depletion is considered a subsidy.").

200. The Congressional Research Service identifies this rationale and then critiques it: Percentage depletion has been justified on national security grounds and the volatile nature of oil and gas prices. In either case, it is likely the concerns could be more adequately addressed through other means. For example, to address national security concerns, one alternative is an oil stockpile program such as the Strategic Petroleum Reserve. TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 118.

201. *Cf.* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 8 (observing that the purpose of a tax expenditure is not always clear).

202. *See, e.g.*, *id.* at 9 (discussing multiple reasons for allowing a deduction of interest on home mortgage indebtedness).

offset the cost of burdens that government has imposed on taxpayers through non-tax law (e.g., through environmental regulation, consumer protection statutes, etc.). Or perhaps one special provision afforded one type of taxpayer is intended to offer a benefit similar to that of another provision aiding a different type of taxpayer; but for various reasons, perhaps even those grounded in constitutional norms, Congress believes it appropriate to employ both provisions.

A plausible example of this last scenario is the enactment of both the exclusion of the rental allowance for parsonages for members of the clergy (and the exclusion of the rental value of a parsonage provided in kind)[203] and the exclusion of the value of lodging furnished to employees for the convenience of the employer on its business premises.[204] The parsonage allowance under IRC § 107 has recently faced judicial scrutiny.[205] Considered in its broader statutory context, § 107 likely reflects a congressional desire to extend to church ministers a benefit similar to that enjoyed by employees of secular employers,[206] but in a way that minimizes church-state squabbles over the location of a church's "business premises" and that maintains neutrality over the church's decision to compensate in kind or with money for rent.[207] This broader legislative design is relevant, even if these provisions are properly classified as tax expenditures.

In summary, that a provision benefits a taxpayer does not necessarily establish a legislative purpose to subsidize. Moreover, classifying a provision as a tax expenditure does little to establish legislative purpose.[208] A legislative body might disagree with an executive agency's determination of the precise elements of a recognized normative tax base or might reject the normativity of a commonly accepted tax base. Further, even

---

203. *See* I.R.C. § 107.

204. *See* I.R.C. § 119.

205. *See, e.g.*, Gaylor v. Mnuchin, 919 F.3d 420, 436-37 (7th Cir. 2019) (upholding the constitutionality of IRC § 107(2)).

206. *See id.* at 428-32.

207. *See* Edward A. Zelinsky, *Do Religious Tax Exemptions Entangle in Violation of the Establishment Clause? The Constitutionality of the Parsonage Allowance Exclusion and the Religious Exemptions of the Individual Health Care Mandate and the FICA and Self-Employment Taxes*, 33 CARDOZO L. REV. 1633, 1659-1661 (2012).

208. *See* Sugin, *supra* note 9, at 424.

when the intent to subsidize exists, the inquiry is not complete. A court must probe much more deeply than any tax expenditure budget when deciding constitutional questions.

E. The Question of Taxpayer Behavioral Adjustments and Tax Incidence

The tax expenditure budget attempts to quantify revenue losses from tax expenditures.[209] But it does not attempt to account for behavioral changes of taxpayers that would result were the provision repealed, nor does it attempt to account for inter-relationships among the various tax preferences were all tax expenditures repealed.[210] In other words, the tax expenditure budget does not measure the actual revenue effects that would likely result from the repeal of any one tax expenditure or all tax expenditures.[211] Further, the tax expenditure budget does not illumine the economic incidence of taxation and tax expenditures.

That the tax expenditure budget fails to account for behavioral changes and the incidence of taxation is relevant to constitutional analysis.[212] When a particular tax expenditure is presented in the tax expenditure budget as "costing" the government, say, $10 billion, it does not mean that the activity being "subsidized" necessarily would suffer a corresponding loss of taxpayer participation in the absence of the tax expenditure. The effect on taxpayer behavior depends on how responsive taxpayers are to the provision in question. Furthermore, who ultimately benefits from a provision may be different from the taxpayer who is immediately affected by the provision.[213]

---

209. *See* SURREY & MCDANIEL, *supra* note 4, at 6.

210. TREASURY REPORT, *supra* note 16, at 1-2.

211. *See id.*; Bittker, *A Reply*, *supra* note 69, at 541. For a broader discussion, see *supra* Section I.B.2.

212. Identifying tax incidence is not easy. Nominally exempt entities may indirectly incur burdens of taxation. For example, Professor Bittker suggests that churches may indirectly assume burdens of taxation because of their inability to shift costs to others. *See* Bittker, *Churches*, *supra* note 26, at 1306-07.

213. *See* GAO GUIDE FOR EVALUATING TAX EXPENDITURES, *supra* note 12, at 16 ("A tax expenditure intended to benefit a particular activity, industry, or class of people may wind up benefiting others not targeted by the tax expenditure by changing prices and incomes.").

A classic example is the exclusion of interest income on bonds issued by state and local governments under IRC § 103.[214] The immediate "tax beneficiary" of § 103 is the bondholder, but the intended ultimate beneficiary of § 103 is the governmental issuer, who theoretically can obtain financing at below-market interest rates because bondholders are willing to receive a lower stated rate of interest (i.e., because it is not subject to taxation).[215] One might think that the repeal of § 103 "costs" bondholders economically, but if their non-taxed, below-market interest received on state and local bonds is no greater than the after-tax returns on taxable bonds, the real loser from repeal of the tax expenditure would be state and local government. Alas, even this analysis is too simplistic. In fact, to secure adequate financing, state and local governments issue bonds bearing interest rates that are high enough to attract not just taxable investors in the highest marginal income tax bracket, but also those in lower tax brackets.[216] The return on state and local bonds is actually higher than the after-tax interest income that upper-income taxpayers receive on taxable securities, and therefore, they have captured some of the benefit of the tax expenditure intended for government issuers.[217] Repeal of § 103 therefore might not just harm state and local issuers, but also upper-income bondholders.[218] The point is that the tax expenditure budget does not convey a great deal of information about the real effects of tax expenditures or their repeal.

---

214. For an extended discussion, see Michael J. Graetz, *Assessing the Distributional Effects of Income Tax Revision: Some Lessons from Incidence Analysis*, 4 J. LEGAL STUD. 351 (1975); *see also* Fleming & Peroni, *supra* note 9, at 446-48.

215. For a critique of IRC § 103 because governmental issuers are economically forced to share far too much of the benefits of exempting interest with bondholders, see Calvin H. Johnson, *Repeal Tax Exemption for Municipal Bonds*, 117 TAX NOTES 1259, 1259 (2007).

216. The evidence indicates that high-income bondholders have indeed captured most of the benefit of § 103, for tax-exempt bonds pay interest at a discount from taxable interest rates well below the highest marginal income tax rate. *See id.* at 1260.

217. *See id.*

218. If, on the other hand, state and local governments did not need to raise interest rates to attract taxable investors who are not in the highest marginal income tax bracket, it is likely that the economic burden of repeal of the § 103 exclusion would fall primarily on governments.

When the constitutionality of a tax provision depends on its effects,[219] classification of an item as a tax expenditure is often non-dispositive. That classification does not answer the question of what effects are produced by the tax expenditure. A court must analyze the real economic and other effects of the provision in question. Surrey's largely dismissive treatment of the relevance of tax incidence in tax expenditure analysis[220] will not do in constitutional adjudication.

For example, consider the possible effects of repealing the charitable contributions deduction. The benefit that charitable donees, such as churches, receive from the ability of their donors to deduct contributions to them depends on various factors. Charitable donations by taxpayers who do not itemize their deductions, typically lower-income taxpayers, presumably would not in the first instance be affected directly by repeal of the deduction, for their donations are not generally deductible in any event.[221] But donations by upper-income taxpayers, who more commonly claim itemized deductions, would become more expensive on an after-tax basis were the deduction repealed. Thus, one would expect charities funded primarily by the wealthy to suffer more than charities funded largely by lower-income taxpayers as result of repealing the charitable contributions

219. For many years, in Establishment Clause cases, the Court often applied the *Lemon* test, one prong of which scrutinized the main effect of governmental action. *See, e.g.*, Lemon v. Kurtzman, 403 U.S. 602, 612 (1971) (citing Bd. of Educ. v. Allen, 392 U.S. 236, 243 (1968)) (holding that the Establishment Clause requires that the "principal or primary effect [of state action] must be one that neither advances nor inhibits religion"). The Supreme Court has recently announced that the *Lemon* test was "long ago abandoned." *See* Kennedy v. Bremerton Sch. Dist., 142 S. Ct. 2407, 2427 (2022). But even *Kennedy v. Bremerton School District* does not categorically dismiss the relevance of the effects of governmental action, for the opinion extensively analyzes the non-coercive nature of the high school football coach's prayers at issue. *See id.* at 2428-32.

220. *See, e.g.*, SURREY & MCDANIEL, *supra* note 4, at 88 (pointing out the difficulty of quantifying "third level" effects of tax expenditures and characterizing the inquiry as distracting).

221. The charitable contributions deduction is authorized by IRC § 170, and it is not identified as a deduction in computing adjusted gross income. *See* I.R.C. § 62. Hence, it is a so-called "itemized deduction." I.R.C. § 63(d). Individuals who do not elect to itemize deductions generally can claim only the standard deduction. *See* I.R.C. § 63(b)(1). Lower-income taxpayers generally are better off claiming the standard deduction; their limited incomes typically cannot generate larger itemized deductions. A small charitable contributions deduction is available to non-itemizers for a limited time under special COVID-related relief legislation. *See* I.R.C. §§ 63(b)(4), 170(p).

deduction, at least when donations are sensitive to the tax cost of giving.[222]

Perhaps, then, the charitable contributions deduction benefits secular, elite charities more than churches. Churches and broadly supported redistributive charities (e.g., the American Red Cross) tend to receive a greater portion of support from less affluent donors than do charities advancing higher education and the arts.[223] Thus, in part because churches are supported by non-itemizers to a greater degree than are charities operating in the arts and humanities, the charitable contributions deduction probably disproportionately benefits charities promoting the arts and secular education.[224]

Repeal of the charitable contributions deduction could foreseeably harm elite, secular charities more than churches for other reasons. The religious ethic of giving to religious bodies, grounded in the biblical tithe[225] but extending well beyond it under New Testament theology,[226] provides a compelling reason for many donors to meet the financial needs of their churches. Naturally, the ethic of the tithe also motivates giving to synagogues, and an ethic of giving is taught in religious faiths besides Judaism and Christianity, as well.[227] This ethic may prompt donors to meet the needs of their religious bodies even when giving becomes more costly. Granted, repeal of the

222. *Cf.* Paul R. McDaniel, *Federal Matching Grants for Charitable Contributions: A Substitute for Income Tax Deduction*, 27 TAX L. REV. 377, 391 (1972) ("[T]he data show that high income individuals and low income individuals do not give to the same charities.").

223. *See* Aprill, *supra* note 101, at 845-46.

224. *See id.* at 868 (stating that the charitable contributions deduction "favors the charitable activities favored by the wealthy"); Charles T. Clotfelter, *Tax-Induced Distortions in the Voluntary Sector*, 39 CASE W. RSRV. L. REV. 663, 685 (1989) ("Charities favored by the rich simply receive more favorable rates of subsidy through the itemized deduction than those favored by the poor."); Todd Izzo, Comment, *A Full Spectrum of Light: Rethinking the Charitable Contribution Deduction*, 141 U. PA. L. REV. 2371, 2373-75 (1993) (arguing that the charitable contributions deduction enables wealthy taxpayers to dictate the charities that the federal government subsidizes most); Edward H. Rabin, *Charitable Trusts and Charitable Deductions*, 41 N.Y.U. L. REV. 912, 922 (1966) (summarizing arguments that the law favors wealthy taxpayers' charities of choice).

225. *See, e.g.*, *Leviticus* 27:30-33.

226. *See, e.g.*, *Mark* 12:41-44; *Acts* 2:42-47; 2 *Corinthians* 9:1-15; 1 *Timothy* 6:17-19.

227. For example, the third pillar of Islam is Zakat, which requires people of a specified means to give a percentage of their wealth to others. *See* Imam Mufti, *The Third Pillar of Islam: Compulsory Charity*, THE RELIGION OF ISLAM (June 25, 2019), [https://perma.cc/AC28-FJYS].

charitable contributions deduction would likely encourage some itemizers to reduce their donations to churches. But one must not dismiss another possibility. Perhaps upper-income donors would mostly maintain their current giving to churches and reduce only (or mainly) their donations to secular charities were Congress to repeal the charitable contributions deduction. Modifying their giving pattern in this manner would enable these donors to maintain the after-tax cost of their total giving at the same levels. Further, even if wealthier itemizers reduced their charitable giving proportionately across the board, lower-income taxpayers with a strong religious ethic might simply give more to their churches to compensate for any budget shortfalls caused by the reduced giving of their wealthier counterparts. In either case, the primary effect of repealing the charitable contributions deduction could be to diminish support for secular charities rather than churches.

Thus, the charitable contributions deduction, which on its face seems to benefit churches in the same way that it advantages elite universities and opera houses,[228] might on an after-tax basis have the effect of much more significantly propping up the donative status of high-brow, secular charities. Correspondingly, the repeal of this provision, even if properly labeled a tax expenditure, could seriously impair the fundraising of non-religious charities while leaving religious charities largely unscathed.

In summary, to characterize a feature of tax law as a tax expenditure reveals neither the ultimate beneficiary of the provision nor the consequences that would result from its repeal.[229] Those who appear to benefit from the provision may economically do so less than others (or not at all), and the repeal of a provision may have a far greater impact on certain nominal beneficiaries of a tax expenditure than others. Indeed, repealing a tax expenditure may economically harm the class of nominal

228. *See* I.R.C. § 170(c)(2)(B) (describing eligible donees, in relevant part, as entities "organized and operated exclusively for religious, charitable, scientific, literary, or educational purposes").

229. Similarly, Professor Sugin has observed that characterizing an item as a tax expenditure says nothing about its distributional effects. *See* Sugin, *supra* note 9, at 424.

beneficiaries less than others who have captured the economic benefit of the tax provision.

When the effects of a law are relevant to constitutional analysis, a court must account for these realities. A judge should not simply assume that the primary effect of a tax expenditure is to benefit the "nominally subsidized" class, or to benefit members of that class equally.

F. The Question of Appropriative Power

Tax expenditure theory tends to skim over, or at least superficially treat, a distinction between direct expenditures and tax expenditures that is important in constitutional doctrine: the locus of appropriative power. Tax expenditure theory posits how Congress "should or could have" acted, not how Congress "would have" acted.[230] The thinking is that a statutory provision conferring a tax benefit places the taxpayer in the same position in which she would have been had Congress taxed her without the preference and then transferred the added tax revenue to her in the form of a direct subsidy equal to the tax savings in the world featuring the tax preference (i.e., the tax world that is).[231] But tax expenditure analysis does not assume that Congress really would appropriate funds for this purpose in the absence of the tax provision at issue. Perhaps Congress would not so appropriate funds that the government truly collects.

For Surrey, that Congress would not directly appropriate funds correspondingly to tax expenditures is a reason to question their wisdom on policy grounds.[232] But for purposes of constitutional analysis, that a legislature would not likely directly appropriate in the manner corresponding to the tax expenditure highlights that the primary appropriative decisionmaker is not the government, but instead a private decisionmaker (the taxpayer). The absence of legislative appropriation reveals a reason to

230. *See* SURREY & MCDANIEL, *supra* note 4, at 1.

231. *Id.* at 82.

232. Surrey discusses the upside-down subsidy effect of tax expenditures, apparently in part to contrast them with the way direct grants typically operate. *Id.* at 71-72, 80-82. The plain implication is that upside-down direct grants would be highly objectionable.

question the equivalence of tax expenditures and direct funding in constitutional cases.

In constitutional law, the identity of the one holding the power to decide upon an appropriation is important. When this person with the power of appropriation—the appropriative decisionmaker—is a private person, rather than a governmental entity or official, indirect governmental support is often upheld as constitutional. Under the logic of a growing body of constitutional law, the presence of a private appropriative decisionmaker who determines to what degree someone benefits from a tax expenditure strengthens the case for the constitutionality of the indirect support (if any) provided through the tax expenditure.

An illustrative case is *Mueller v. Allen*.[233] In *Mueller v. Allen*, the Court upheld the constitutionality of a state income tax law permitting a deduction for payments of tuition, books, and transportation enabling the taxpayers' children to attend school.[234] On the whole, parents of children enrolled in private religious schools likely received much of the benefit of the deduction. One of the reasons that the Court found the deduction permissible under the Establishment Clause is that the indirect benefit that private schools reaped from the law arose "only as a result of decisions of individual parents."[235] The presence of these private appropriative decisionmakers meant that the state government had not signified approval of any one religion, or of religion in general.[236]

Mueller v. Allen is not an isolated case. The presence of a private appropriative decisionmaker has been a significant factor in several Supreme Court opinions.[237]

---

233. 463 U.S. 388 (1983).

234. *See id.* at 395-97, 400, 402-03.

235. *Id.* at 399, 400 (stating that the "historic purposes" of the Establishment Clause "simply do not encompass the sort of attenuated financial benefit, ultimately controlled by the private choices of individual parents," that inured to parochial schools from a neutrally available state income tax deduction).

236. *See id.* at 399.

237. *See, e.g.*, Ariz. Christian Sch. Tuition Org. v. Winn, 563 U.S. 125, 142-43 (2011) (holding that taxpayers lack standing to challenge the constitutionality of a state program providing tax credits for transfers by private persons to tuition organizations); Zelman v. Simmons-Harris, 536 U.S. 639, 653, 662-63 (2002) (upholding the constitutionality of a governmentally funded school-voucher program enabling students to attend private schools

Tax expenditure theory's reliance on what Congress "should or could have" done fares poorly in the jurisprudence controlling at least some constitutional provisions (e.g., the Establishment Clause). Indeed, what is perceived as a negative attribute in tax expenditure theory—the absence of direct legislative control of appropriations when the law features the mechanism of a tax expenditure—is in certain contexts a positive factor in analyzing a tax expenditure's constitutionality. The tax provision marks the boundaries, but taxpayers' choices dictate ultimate recipients of funds within those boundaries. Constitutional doctrine has assigned significance to these private choices.

G. The Question of Tax Penalties

Tax expenditure theory posits not only "positive" spending, but also "negative" spending—the imposition of penalties—through special statutory tax mechanisms.[238] The idea follows from the assumption of a normative tax base and structure.[239] If a deduction or exclusion is normatively correct because it is necessary to measure income, then to deny the deduction or

---

of their choice); Zobrest v. Catalina Foothills Sch. Dist., 509 U.S. 1, 10, 13-14 (1993) (finding that a governmental program requiring a school district to provide sign-language interpreters to help deaf students did not violate the Establishment Clause even when a deaf student was enrolled in a private Catholic school, relying in part on the fact that the choice of school was made by the student's parents rather than the government); Witters v. Wash. Dep't of Servs. for the Blind, 474 U.S. 481, 487-89 (1986) (finding no Establishment Clause violation by a state scholarship program that aided a student studying for the ministry at a religious institution where any benefit realized by the religious entity resulted from the student's private, independent choice). The Court's analysis in *Zelman v. Simmons-Harris* is representative:

> [W]here a government aid program is neutral with respect to religion, and provides assistance directly to a broad class of citizens who, in turn, direct government aid to religious [institutions] wholly as a result of their own genuine and independent private choice, the program is not readily subject to challenge under the Establishment Clause. A program that shares these features permits government aid to reach religious institutions only by way of the deliberate choices of numerous individual recipients. The incidental advancement of a religious mission, or the perceived endorsement of a religious message, is reasonably attributable to the individual recipient, not to the government, whose role ends with the disbursement of benefits.

Zelman, 536 U.S. at 652.

238. *See* SURREY & MCDANIEL, *supra* note 4, at 222-24.

239. *See id.* at 222.

exclusion is effectively to impose a tax penalty.[240] Surrey went so far as to analyze tax penalties as "functional equivalents of direct government regulatory or financial penalty rules."[241] Surrey identified a number of tax penalties, including the denial of a deduction under IRC § 162 for fines, lobbying expenditures, and contributions to political campaigns, when all of these are business-related.[242]

The Supreme Court has appeared slow to recognize the punitive nature of negative tax expenditures when considering their constitutionality. For example, in *Cammarano v. United States*,[243] decided prior to the prominence of the tax expenditure idea, the Court upheld the constitutionality of a Treasury regulation that denied a deduction for lobbying expenses, including grass roots lobbying, paid in the course of carrying on a trade or business[244]—a denial now statutorily implemented.[245] The Court justified the denial of a deduction by reasoning that the taxpayers "are simply being required to pay for those activities entirely out of their own pockets, as everyone else engaging in similar activities is required to do."[246] According to the Court, the provision just reflects the view that "everyone in the community should stand on the same footing as regards its purchase [of publicity over legislation] so far as the Treasury of the United States is concerned."[247]

The problem under tax expenditure theory is that the decision of *Cammarano* does not leave all community members standing "on the same footing . . . so far as the Treasury of the United States is concerned."[248] Expenditures to influence legislation that a taxpayer makes for reasons unrelated to the objective of producing income are non-deductible personal

240. *See id.* (discussing the propriety of deducting costs of producing business income).
241. *Id.*
242. *See id.* at 222-24.
243. 358 U.S. 498 (1959).
244. *See id.* at 512-13.
245. *See* I.R.C. § 162(e)(1)(A), (C).
246. *Cammarano*, 358 U.S. at 513.
247. *Id.*
248. *Id.*

consumption.[249] Non-business taxpayers properly pay for attempts to influence legislation with after-tax dollars.[250] But a taxpayer's business-related lobbying expenditures are expenses of producing income, and hence deductible (or at least capitalizable) under the normative tax structure based on economic income, or indeed any common understanding of business profit.[251] To deny a deduction (or at least basis) for business taxpayers is to penalize them. What appears as a system placing everyone on the "same footing" to the eye untrained by Surrey is really a system that penalizes business taxpayers—to those who peer through the lens of tax expenditure theory. Even many years after the decision in *Cammarano*, the Supreme Court still failed to grasp the punitive nature of denying the business deduction for lobbying expenses.[252]

One may attempt to justify the result in *Cammarano* on various policy grounds, but none is highly satisfying under tax expenditure theory or existing constitutional law. One option is to admit that the denial of a lobbying expense deduction made for business reasons is a penalty but then to argue that it is nonetheless constitutional when balanced against the potential harm to democracy if only business lobbying is untaxed.[253] However, if the denial of the deduction is intended to penalize business-motivated political speech, the denial raises serious constitutional concerns. Advocating before the general public, the type of lobbying at issue in the statewide referenda described in *Cammarano,* implicates core First Amendment values.[254] It is

249. *See* I.R.C. § 262.

250. *See* I.R.C. § 262.

251. *See* SURREY & MCDANIEL, *supra* note 4, at 222.

252. *See* Regan v. Tax'n with Representation of Wash., 461 U.S. 540, 546 (1983) ("Congress has not infringed any First Amendment rights or regulated any First Amendment activity. Congress has simply chosen not to pay for . . . lobbying.").

253. Any such argument is in tension with recent Supreme Court precedent, which rejects anti-distortion rationales for limiting corporate speech in the form of independent expenditures for electioneering. *See, e.g.*, Citizens United v. FEC, 558 U.S. 310, 349-53 (2010).

254. *See* U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech . . . or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances."); *see also, e.g.*, Meese v. Keene, 481 U.S. 465, 465-66, 480, 484-85 (1987) (holding the Foreign Agents Registration Act of 1938 constitutional and reasoning that "Congress did not prohibit, edit, or restrain the distribution

hardly a serious suggestion that penalizing political speech is constitutional simply because it is motivated by the desire to make money.[255]

The analyst is thus left with justifying the result in *Cammarano* on some grounds other than that it constitutionally tolerates a penalty on speech. But understanding *Cammarano* as properly characterizing the denial of the business deduction for lobbying as other than a penalty does not bode well for the role of tax expenditure analysis in constitutional law. The position implies that denying a deduction for expenses of income-producing lobbying (while also denying their capitalization) is part of the normatively correct tax base.[256] That explanation undermines the force of tax expenditure theory itself, and Surrey knew it.[257]

In considering whether various tax penalties can be conceptualized as features of the normative income tax structure, Surrey ultimately concluded that going down this intellectual road would have no end.[258] To begin rationalizing them under sundry policy norms invites the same exercise for all negative tax expenditures, and it would be impossible to decide which public policies produce "normal" tax provisions and which create tax penalties.[259] For Surrey, then, all denials of an income tax

---

of advocacy materials in an ostensible effort to protect the public from conversion, confusion, or deceit").

255. *Cf. Citizens United*, 558 U.S. at 340, 365-66 (holding that laws that burden political speech are subject to strict scrutiny; holding unconstitutional a law that limited corporate independent expenditures for electioneering). According to the Court in *Citizens United*, "If the First Amendment has any force, it prohibits Congress from fining or jailing citizens, or associations of citizens, for simply engaging in political speech." *Id.* at 349. First Amendment protection extends to the speech of corporations. *See id.* at 342-43.

256. Perhaps administrative expediency justifies this notion of the tax base. The argument would be that ascertaining whether a taxpayer's motive to influence legislation is personal or business-related is too difficult.

257. Critics of tax expenditure theory understand well the point that disallowing a deduction for an expense to produce income reflects a normative judgment. *See, e.g.*, Kahn & Lehman, *supra* note 52, at 1661-62 ("The disallowance of a deduction for illegal bribes confirms that we think they are naughty."). For critics, these tax provisions just affirm that the enacted tax law is itself the normative base chosen by the public. *See id.*

258. *See* SURREY & MCDANIEL, *supra* note 4, at 223.

259. *See id.* More broadly, and under essentially the same logic, to rationalize denials of deductions for expenses incurred to produce income on a normative basis invites the same exercise for all tax expenditures. If one norm justifies classifying one of Surrey's tax penalties as part and parcel of the normative tax base, another norm just as surely justifies

deduction for expenses incurred to generate income must be characterized as tax penalties.[260]

A still deeper problem exists. Tax penalties sometimes present more troublesome state action than tax expenditures. In the case of tax expenditures, even if it is assumed that the provisions are deviations from the normative tax structure, the most that often can be said is that government is supporting the exercise of private choices. Some, perhaps many, of these choices are exercises of rights enjoying enhanced constitutional protection. So a tax expenditure often represents, at most, government support of the exercise of constitutionally protected rights.[261] But some tax penalties target constitutionally protected behavior. Phrased more pointedly, through some tax provisions, government penalizes the exercise of constitutionally protected rights.[262] The nature of the state action thus is at times much more suspect in the case of tax penalties because they may suppress constitutionally protected activity.

This discussion shows that one feature of tax expenditure theory—the concept of tax penalties—fits uncomfortably with existing Supreme Court precedent. Yet the concept of tax penalties is part and parcel of the overall logic of tax expenditure analysis. Further, although existing Supreme Court precedent does not reflect this observation, tax penalties probably present greater constitutional concerns than positive tax expenditures in a number of contexts.

H. The Question of Tax History and Historical Inquiry in General

Tax expenditure theory, at least as explained by Surrey, is underdeveloped in assessing the relevance of tax history in

---

classifying one of Surrey's tax expenditures as an element of the proper base. There is hardly a terminus to the exercise. If there is a conceptual end, it might simply be the one that Bittker perceived—that tax law "is made, not discovered." Bittker, *A Reply*, *supra* note 69, at 541.

260. *See* SURREY & MCDANIEL, *supra* note 4, at 223.

261. *See id.* at 118.

262. *See, e.g.*, HAMBURGER, *supra* note 101, at 190-213 (arguing that IRC § 501(c)(3) penalizes the speech, petition, and religious exercise rights of charitable and religious organizations). For a detailed (and critically supportive) review of Professor Hamburger's arguments, see Buckles, *The Penalty of Liberty*, *supra* note 101.

determining the normative tax structure. On the one hand, as noted previously, Surrey envisioned an expanding list of tax expenditures, and then eventual repeal of newly identified tax expenditures, as the public generally comes to accept economic concepts of income.[263] This aspect of Surrey's analysis positively assesses historical developments. On the other hand, Surrey's recognition that the normative structure of the income tax *at any given point in time* also reflects general acceptance and normative factors beyond economic concepts of income[264] suggests that historical views of income and tax policy also serve some role in determining the normative tax. These two aspects of Surrey's thinking raise interesting questions in evaluating the role of tax expenditure theory in constitutional law.

One question is whether the duration of a provision nominally structured as an exclusion, deduction, or exemption should influence the classification of the provision as a tax expenditure. For example, in the arc of federal income tax history, the charitable contributions deduction and the charity income tax exemptions are ancient.[265] That these provisions have existed for so long is surely some evidence that Congress, and the general public that elects its members, view them as part of the normative income tax structure.[266] In other words, the longstanding presence of these provisions may suggest the "public acceptance" of their normativity—a factor Surrey claimed was relevant.[267]

A related inquiry is even more important for purposes of this Article: beyond tax expenditure theory proper, how is the constitutional analysis of a tax provision affected by its duration? History matters mightily in some—and some would say most or

---

263. *See* Surrey & Hellmuth, *supra* note 64, at 532.

264. *See id.* at 531-32.

265. The charitable contributions deduction was enacted in 1917. *See* War Revenue Act, ch. 63, § 1201(2), 40 Stat. 300, 330 (1917). The charity income tax exemption appeared not only in the first internal revenue statute enacted after the adoption of the Sixteenth Amendment but also in previous revenue acts. *See* Bittker & Rahdert, *supra* note 145, at 301-03.

266. Of course, as a logical matter, the evidence is not conclusive. Perhaps the public both recognizes the non-normativity of a tax provision as a matter of tax policy and yet supports the tax provision on non-tax-related policy grounds.

267. *See supra* Section II.C.

all—constitutional contexts. For example, in a growing number of cases, the Court has relied on the history of a practice in assessing whether it survives Establishment Clause scrutiny.[268] The long history of the charity income tax exemption and the charitable contributions deduction bode well for these provisions under the Court's growing reliance on historical acceptance. Charity property tax exemptions likewise have a long history, a point that figured prominently in *Walz v. Tax Commission of New York*.[269]

A similar question is to what degree historical inquiry in general should guide constitutional analysis of tax expenditures. As to the federal income tax, Surrey plainly preferred to account for evolving notions of income in conducting tax expenditure analysis.[270] Applied in constitutional contexts, this preference suggests dynamic constitutional interpretation. But the current Court is heavily influenced by originalism.[271] Thus, in resolving constitutional questions involving tax expenditures under the federal income tax, the Court would likely consider both the history of the tax provision at issue and the historical context of the adoption of the constitutional provision bearing upon the tax provision.

---

268. *See, e.g.*, Kennedy v. Bremerton Sch. Dist., 142 S. Ct. 2407, 2415-16, 2426 (2022) (holding that a school district violated the First Amendment rights of a high school football coach by forbidding him from praying at midfield after games had ended); *id.* at 2428 ("An analysis focused on original meaning and history, this Court has stressed, has long represented the rule rather than some 'exception' within the 'Court's Establishment Clause jurisprudence.'"); Town of Greece v. Galloway, 572 U.S. 565, 591-92 (2014) (holding constitutional a township's practice of opening official meetings with prayer by invited clergy members). According to the Court, "to define the precise boundary of the Establishment Clause" is unnecessary when history supports the permissibility of the specific practice under examination. *Id.* at 577.

269. 397 U.S. 664, 675-80 (1970) (citing the long history of tax exemptions for property owned by religious entities as supporting the constitutionality of such exemptions under the Establishment Clause).

270. *See* SURREY & MCDANIEL, *supra* note 4, at 197-209.

271. *See, e.g.*, Dobbs v. Jackson Women's Health Org., 142 S. Ct. 2228, 2242-43 (2022) (holding that the Due Process Clause of the Fourteenth Amendment does not confer a right to abortion because it fails the test of being "deeply rooted in this Nation's history and tradition" and "implicit in the concept of ordered liberty"); N.Y. State Rifle & Pistol Ass'n v. Bruen, 142 S. Ct. 2111, 2122, 2126 (2022) (holding that New York's licensing regime, which required an applicant to demonstrate a special need for self-defense to carry a hand gun outside of the home, violates the Second and Fourteenth Amendments, stating that a regulation burdening an individual's right to bear arms must be "consistent with this Nation's historical tradition" to survive constitutional scrutiny).

In brief, the history of a tax provision is often important to constitutional analysis. But in tax expenditure theory, the significance of tax history is not well developed. A study of the history of a provision of tax law is probably warranted in determining the normative tax base, and it may be important for other reasons in applying relevant constitutional doctrines. Further, the history of the constitutional text under which a tax provision must be analyzed by a court is also relevant. To the extent that other branches of government have not accounted for legal history in compiling tax expenditure budgets, judicial deference to tax expenditure lists in constitutional cases is all the more problematic.

I. The Question of Mechanical Interchangeability

Tax expenditures can take various forms—exclusions, deductions, deferrals, credits, and special nominal rates.[272] One of these forms can often be expressed in another form to produce an identical result. In some contexts, understanding this mechanical interchangeability among tax expenditure forms is necessary. This Section first illustrates mechanical interchangeability and then explains why mechanical interchangeability is sometimes important.

Consider the equivalence between partial exclusions of gains and reduced rates of tax. One way the federal income tax system encourages investments in capital is by imposing a special rate of tax on net capital gain.[273] Assume a taxpayer has $100,000 of net capital gain taxed at the rate of 20% when the taxpayer is in a 50% marginal income tax bracket. The preferential rate gives rise to a tax expenditure of $30,000—the product of the gain and the difference between the marginal tax rate and the preferential tax rate.[274] Instead of taxing net capital gain at the rate of 20%, Congress could confer an equivalent benefit by excluding 60% of the taxpayer's net capital gain from income and taxing the

272. The Congressional Budget and Impoundment Control Act of 1974 identifies these forms, as well as special exemptions, as tax expenditures. *See* Pub. L. No. 93-344, § 3(a)(3), 88 Stat. 297, 299 (codified as amended at 2 U.S.C. § 622(3)).

273. *See* I.R.C. § 1(h).

274. The tax expenditure is *(0.5-0.2) x $100,000 = $30,000*.

remaining gain at the normal rate of 50%.[275] The two formal mechanisms produce the same benefit to the taxpayer.

Congress also has a choice of providing for a credit or a deduction. Assume a deduction is allowed for up to $5,000 of college tuition paid by a taxpayer on behalf of the taxpayer's children. Further assume that the highest marginal income tax rate is 50%. To the taxpayer in the 50% marginal income tax bracket, this deduction produces a tax benefit of $2,500.[276] Instead of permitting a deduction for tuition, Congress could enact a tax credit equal to 50% of the total cost of tuition paid, subject to a $5,000 tuition ceiling. Thus, when this taxpayer pays tuition equal to or exceeding $5,000, her tax liability otherwise due is offset by a credit of $2,500.[277] For our taxpayer in the 50% marginal income tax bracket, the deduction and the credit produce equivalent after-tax benefits.

The mechanical interchangeability of these various forms of tax expenditures is important to recognize. What might at first glance seem like a remarkable "giveaway" might not be such a bonanza upon further scrutiny. For example, consider a state tax credit for private school tuition equal to 100% of the first $1,000 of tuition paid on behalf of the taxpayer's children. One might react to the credit as a windfall for parents sending their children to private schools, insofar as a credit is a dollar-for-dollar reduction in tax liability. But the mechanical interchangeability of tax expenditures should temper this reaction. For a taxpayer in the 50% income tax bracket, a 100% tax credit for up to $1,000 of tuition paid is equivalent to a deduction for private school tuition paid up to a $2,000 ceiling.[278] For a taxpayer in the 50% marginal tax bracket who incurs $10,000 of private school tuition, a tax credit of $1,000 for tuition paid equaling or exceeding $1,000 is a much smaller tax expenditure than a full deduction. This taxpayer would prefer a full deduction for all tuition paid,

---

275. The tax expenditure is the amount of tax foregone on the excluded portion of the gain, or *0.5 x (0.6 x $100,000) = $30,000*.

276. The taxpayer's income subject to tax is reduced by $5,000, for a tax benefit of *0.5 x $5,000 = $2,500*.

277. The tax credit is *50% x $5,000 = $2,500*.

278. Deducting tuition of $2,000 produces a tax savings of *0.5 x $2,000 = $1,000*.

for the full deduction would produce a tax benefit five times the benefit of the $1,000 tax credit.[279]

Mechanical interchangeability among various tax expenditures is also important to recognize when a base-defining provision is expressed in a less familiar form. To illustrate, assume that, contrary to the limitations on the deductibility of state and local taxes under current law,[280] and contrary to the assumptions underlying current tax expenditure budgets,[281] Congress determines that the normative federal income tax base does not include income transferred to pay state and local governments. One obvious way to implement this conclusion is to provide for a full deduction for the payment of state and local taxes.[282] But perhaps Congress is concerned that it will be unjustly accused of favoring the wealthy, whose high marginal tax rates render a deduction more valuable to them than to lower-income taxpayers.[283] So Congress replaces the deduction with a tax credit for state and local taxes paid in an amount equal to the product of such taxes and the highest marginal income tax rate. If state and local taxes are properly excluded from the normative income tax base, and hence a deduction for state and local taxes is normative, this credit mechanism does not produce a tax expenditure for taxpayers in the highest marginal income tax bracket. It produces a tax result equal to a full deduction. The credit mechanism described would, however, produce a tax

---

279. For the taxpayer in the 50% marginal income tax bracket, a full income tax deduction produces a tax benefit of *$10,000 x 50% = $5,000*.

280. *See* I.R.C. §164(b)(6)(B) (limiting, through 2025, an individual taxpayer's deduction for most state and local taxes to $10,000).

281. *See* JCT TAX EXPENDITURES REPORT, *supra* note 18, at tbl. 1; TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 1071-78; TREASURY REPORT, *supra* note 16, at 9-10, 21.

282. The deduction for state and local taxes has received thoughtful academic commentary. *See, e.g.*, Brookes D. Billman, Jr. & Noel B. Cunningham, *Nonbusiness State and Local Taxes: The Case for Deductibility*, 28 TAX NOTES 1107 (1985); Louis Kaplow, *Fiscal Federalism and the Deductibility of State and Local Taxes Under the Federal Income Tax*, 82 VA. L. REV. 413 (1996); Kirk J. Stark, *Fiscal Federalism and Tax Progressivity: Should the Federal Income Tax Encourage State and Local Redistribution?*, 51 UCLA L. REV. 1389 (2004); Edward A. Zelinsky, *The Deductibility of State and Local Taxes: Income Measurement, Tax Expenditures and Partial, Functional Deductibility*, 6 AM. J. TAX POL'Y 9 (1987).

283. The greater value of deductions to high-income taxpayers explains the familiar argument that tax expenditures confer upside-down subsidies. *See, e.g.*, SURREY & MCDANIEL, *supra* note 4, at 72-82.

expenditure for those in lower-income tax brackets, in proportion to the difference between their marginal rate and the highest marginal rate. This conclusion follows because the normative tax system postulated would simply grant these taxpayers a full deduction, but the credit mechanism—based as it is on the top marginal tax rate—grants them a greater tax benefit than would a full deduction for taxes paid.

The implications for constitutional analysis are apparent. First, a "dollar-for-dollar" credit based on a taxpayer's expenditures subject to a ceiling does not necessarily imply as much governmental support as one might initially suspect. Secondly, a credit or special tax rate that approximates the same results as a tax base-defining deduction or exclusion should be analyzed accordingly. The mere presence of a credit does not necessarily imply the existence of an indirect government subsidy or a legislative intent to subsidize. Indeed, and perhaps counterintuitively, a credit mechanism can have the effect of a negative subsidy (i.e., a penalty). If (i) a legislature enacts a partial tax credit in lieu of a deduction, but (ii) the deduction would generate a larger tax benefit than a credit to the taxpayer in question, and (iii) the full deduction is consistent with the normative tax base, then it follows that (iv) the credit mechanism functions in part as a tax penalty.[284]

The form of a tax provision may reflect legislative goals, compromises, and values that have little to do with an intent to confer an indirect government subsidy on some or all of those who rely on the provision to compute their tax liabilities. Constitutional analysis must not stop with a cursory look at the mechanical form of a tax provision, including one that appears on a tax expenditure list.

J. The Question of Changes in Tax Rates and Taxability Thresholds

Tax expenditure theory does not assume a certain tax rate or rate structure as normative independently of the rate structure

284. Surrey & McDaniel acknowledged a similar point. *See id.* at 80 (observing that the denial of a normatively required deduction functions as an upside-down penalty).

congressionally established.[285] To the contrary, Surrey emphatically maintained that it was folly to criticize tax expenditure theory on the grounds that it assumes all income belongs to the government—an assumption equivalent to insisting that any tax rate below 100% must be a tax expenditure.[286] Similarly, standard allowances that shield a minimum of income from taxation—the standard deduction[287] and (when effective) the personal exemption[288]—are not informed by tax expenditure theory.[289] Tax expenditure theory just accepts these subsistence allowances as extraneous realities, determined from time to time legislatively.[290]

One anomaly that follows from this perspective is that the size of the tax expenditure budget is a function of tax rates and the subsistence allowances, even though they are regarded as independent of the process of identifying tax expenditures. Surrey himself declared tax expenditures "hostage to the regular rate structure."[291] He recognized that reductions in tax rates, the allowance of the standard deduction, and the allowance of personal exemptions diminish the "scope and cost of tax expenditures."[292] The reason is simple math.

When Congress reduces tax rates on a grand scale, the tax expenditure budget correspondingly diminishes. The reduction

---

285. *See id.* at 220 (stating that, in a general-rate structure, neither a zero-rate bracket nor any other rate of tax below the highest marginal rate is a tax expenditure, explaining that "changes in the general positive rates do not involve tax expenditures").

286. *See id.* at 60-61.

287. *See* I.R.C. § 63(b)(1), (c).

288. *See* I.R.C. §§ 63(b)(2), 151. For taxable years 2018-2025, the deduction for personal exemptions is zero. *See* I.R.C. § 151(d)(5)(A).

289. *See* SURREY & MCDANIEL, *supra* note 4, at 220. Surrey & McDaniel discuss the propriety of both the personal exemption and a zero-rate tax bracket. *See id.* The standard deduction is the functional equivalent of the latter. *Cf.* JCT TAX EXPENDITURES REPORT, *supra* note 18, at 4 (stating that the JCT "views the standard deduction and the personal exemptions as defining the zero-rate bracket that is a part of normal tax law").

290. *Cf.* McMahon, *supra* note 2, at 776 (stating that, because the Haig-Simons concept of income "does not address issues such as accounting methods, taxable units, exemptions levels, or inflation adjustments, those issues must also be addressed in designing a normal income tax"). Professor Bittker was quick to observe that the Haig-Simons concept "provides no guidance to many structural issues that must be decided in any income tax law." Bittker, *"Tax Subsidies," supra* note 56, at 251. Bittker counted "tax exemptions" among such structural provisions. *See id.* at 260.

291. SURREY & MCDANIEL, *supra* note 4, at 103.

292. *See id.* at 104.

follows mathematically from lowering rates. The hypothetical transfers from taxpayers to the government, and then back from the government to taxpayers, are lesser than they would otherwise be because the tax rates—and hence hypothetical tax revenues—are lower than they were before the rate reduction.[293] This reduction in tax expenditures attributable to an overall reduction in tax rates does not necessarily result from any conscious choice to reduce indirect subsidies.[294] Similarly, when Congress raises tax rates, the primary reason it does so is likely to increase actual collections, not necessarily to enhance indirect support across the board for activities benefitting from tax expenditures. But the mathematical result of raising rates, all else held constant, is to expand the tax expenditure budget.

Similarly, raising subsistence allowances reduces the size of the tax expenditure budget. To take an extreme example, imagine that Congress raised the standard deduction to $100,000. Nobody with adjusted gross income of $100,000 or less would have any reason to itemize personal deductions.[295] Thus, these taxpayers would claim no deductions for most state and local income or property taxes,[296] mortgage interest expense,[297] charitable contributions,[298] or medical expenses.[299] As a result, the tax expenditure budget would shrink. Simply raising the subsistence allowance reduces indirect government spending through the tax system under tax expenditure theory.

At any moment, then, how much "subsidy" flows indirectly through the federal tax system to support various activities is in part a function of legislative priorities that likely have little if anything to do with a conscious choice to enhance or cut support for such items as charity, medical care, housing, and state taxes.

---

293. *See id.* at 60. This analysis assumes all else is constant. Clearly, if lowering tax rates stimulates economic activity sufficiently, tax revenues may well increase.

294. Surrey & McDaniel speculate that, when tax rates were dramatically reduced in the early 1980s, Congress was "probably unaware" that the change reduced tax expenditures. *Id.* at 104.

295. Taxpayers may either claim the standard deduction or elect to itemize personal deductions. *See* I.R.C. § 63(b).

296. *See* I.R.C. § 164.

297. *See* I.R.C. § 163(h)(3).

298. *See* I.R.C. § 170.

299. *See* I.R.C. § 213.

Major changes from year to year in the tax expenditure budget reveal little about conscious legislative subsidy choices when they are attributable to alterations to general tax rates and subsistence allowances.[300]

Constitutional analysis should account for these nuances. When intent matters, a court should recognize that the main legislative objective of a change in tax rates or subsistence allowances is likely not to influence taxpayer activities that tax expenditure theory views as publicly subsidized. Concededly, some legislators, at least those on tax-writing committees, should have some sense for how changes in rates and allowances might affect all kinds of taxpayer behavior. But if Congress intends to alter specific aspects of taxpayer activity, the more precise way to do so through taxation is to modify the conditions of tax expenditure provisions most likely to influence such behavior. To assume that Congress appreciates this reality is reasonable.

K. The Question of Economic Equivalence

Another question to consider is the relevance of economic equivalence. That a tax expenditure is economically equivalent to full taxation coupled with the granting of a subsidy equal to the tax benefit is now platitudinal. Under the simplistic assumptions of tax expenditure theory, the point is obvious. To illustrate, if the normal marginal rate of tax on income is 50% but dividend

---

300. Taxpayer behavior in response to changes in tax rates and subsistence allowances may compound unintended effects on tax expenditures. For example, as measured by the tax expenditure budget, charitable giving receives a greater subsidy as tax rates rise. Such is the mathematical result of allowing a charitable contributions deduction to taxpayers now paying a higher rate of tax. But if charitable giving to churches is more inelastic than giving to secular charities, the effect of the government's raising income tax rates or lowering the standard deduction may be to stimulate contributions to secular charities more than to spur giving to religious entities. In a period of gradually rising income tax rates, and all else remaining constant, the tax expenditure budget would show increased subsidies to charitable donees. But when taxpayer behavior is considered, it might be that the rise in tax rates causes secular charitable giving to increase more than religiously motivated support. Similarly, lowering income tax rates, and expanding the standard deduction, may on balance hurt secular charities more than religious groups. When tax rates decline over time, perhaps religious giving stays more constant than secular charitable giving, which wanes as the value of the deduction declines with tax rates. But it is not likely that the primary legislative purpose in modifying tax rates or the standard deduction is to rebalance public support of religious and secular charities.

income is taxed at 25%, a taxpayer who receives a dividend of $1,000 during the taxable year and thus pays tax of $250 at the preferential 25% rate is in the same economic position as a taxpayer in a world without tax expenditures who pays tax of $500 on the $1,000 dividend at the normal 50% rate but receives a $250 direct subsidy from the government. This observation of economic equivalence is mathematically correct. But it is also largely irrelevant to constitutional analysis.

All kinds of economic equivalencies exist, but they establish little or nothing. Consider a taxpayer subject to a normal marginal rate of income tax of 30%. Assume the taxpayer earns $100,000 subject to the highest marginal tax rate, and therefore she pays tax of $30,000 and is left with $70,000 on this band of income after taxes. She is in the same position that she would have been in had the government taxed this band of income at the normal marginal rate of 50% and then transferred a subsidy to her of $20,000. The two systems produce the same economic result to her—she winds up with $70,000 when the dust settles. The same would be true were the highest normal marginal tax rate 75% and the hypothetical subsidy $45,000, or were the highest normal marginal tax rate 90% and the hypothetical subsidy $60,000.

For purposes of constitutional law, to insist that a court find existential significance in one or more of these economic equivalencies is silly. That one can illustrate economic equivalence among these various options means only that one has mastered grammar school-level math, not that economic equivalence is itself important.

Tax expenditure theory relies on economic equivalencies, but it is not established by them. Surrey castigated the objection to tax expenditure theory that it assumed the government's entitlement to all income—a claim which is tantamount to asserting that any tax rate below 100% would be a tax expenditure.[301] Surrey understood the selection of an ordinary, typically progressive, tax rate schedule as distinct from the tax expenditure inquiry.[302] If economic equivalence established the presence of a tax expenditure, Surrey's rebuttal would lack merit.

301. *See* SURREY & MCDANIEL, *supra* note 4, at 60-61.
302. *See id.* at 190-92.

The payment of tax on income at any rate below 100% is mathematically equivalent to a payment at the rate of 100% coupled with a subsidy. But Surrey would have none of that, for the objection rests on the faulty premise that tax expenditure theory assumes that the government is entitled to all taxpayer income in the first instance. Economic equivalence in this example proves nothing.

The deductibility of expenses that are necessary to arrive at the normative tax base further illustrates that economic equivalence does not establish the presence of a tax expenditure. The classic example is the deduction for ordinary and necessary trade or business expenses.[303] The deduction for these expenses is economically equivalent to a system that provides for no deduction but confers a subsidy equal to the product of the tax rate and the amount of such expenses (capped at business gross income). But this economic equivalence does not render the IRC § 162 deduction for business expenses a tax expenditure.[304]

In addition, economic equivalence does not imply other forms of equivalence,[305] such as legislative appropriative process equivalence,[306] lobbying equivalence,[307] financial liquidity and

303. *See* I.R.C. § 162.

304. *See* SURREY & MCDANIEL, *supra* note 4, at 187 (explaining that the authorized deductions for expenses of producing income are consistent with the Haig-Simons concept of income).

305. Even opinions of the Court that are friendly to tax expenditure analysis have recognized this point. *See, e.g.*, Regan v. Tax'n with Representation of Wash., 461 U.S. 540, 544 n.5 (1983) ("In stating that exemptions and deductions, on the one hand, are like cash subsidies, on the other, we of course do not mean to assert that they are in all respects identical.").

306. *See* Zelinsky, *Tax "Benefits," supra* note 22, at 401. Removing a benefit from the appropriations process may reduce governmental transaction costs, enhance stability in the law, and reduce political conflict that would ensue from annually revisiting the subsidy issue. *See id.* at 401-03.

307. *See id.* at 401-02. Citizens receiving tax benefits under provisions that are mainstays of the IRC likely need not expend as much time and money lobbying legislators as they would in a world in which they are fully taxed but must seek direct legislative grants annually.

security equivalence,[308] fiduciary risk equivalence,[309] and symbolic sponsorship equivalence.

The differences between direct legislative appropriations and tax expenditures are especially stark. Tax expenditures quite clearly do not bear the marks of a special legislative appropriation. As to the latter, a government body decides in advance both the identity of recipients and how much they receive of limited funds.[310] But taxpayers ultimately determine whether they benefit, and to what extent, under a tax expenditure—within the confines of the statutory definitions, of course. This distinction has not escaped the attention of the Supreme Court.[311] Moreover, the annual appropriations process leaves funding "up for grabs," whereas provisions conveying tax benefits tend to endure with less congressional wrangling.[312] This stability itself may reflect, as well as reinforce, the normativity of longstanding tax expenditures; rather than prolonging a subsidy that should be subjected to annual spending discipline, perhaps the choice of structuring some tax benefits as deductions, exclusions, credits, etc., points to the consensus that they are normative and should be beyond the pale of annual appropriative tinkering.

Moreover, Congress by its own actions testifies against the equivalence of appropriations and tax expenditures. Consider the battle over whether to include the Hyde Amendment in the general appropriations bill during the Biden administration.[313]

---

308. *Id.* at 402. When a taxpayer computes a lower-income tax liability because of a tax expenditure for year *X*, and the taxpayer then pays the tax liability in April (or October, with a filing extension) in year *X + 1*, the taxpayer can continuously retain cash equal to the tax savings from the tax benefit. But if the taxpayer is instead fully taxed for year *X* and then forced to seek a direct grant from the government when the taxpayer files a return in Year *X + 1*, the taxpayer is without the use of the cash benefit from the date of filing the return to the date of receiving the grant from the government. Thus, the taxpayer has less cash on hand (for a time) and assumes the risk of not receiving the grant on a timely basis. Moreover, a taxpayer cannot secure money due under a grant until the taxpayer receives it.

309. *Id.* at 403. If a taxpayer retains the cash generated by a tax expenditure, there is no risk that the cash benefit will be mismanaged, wrongly recorded, or misappropriated by a government agent, as in the case of money due under a grant.

310. *See supra* Section II.F; *see also* Zelinsky, *Tax "Benefits," supra* note 22, at 403-13.

311. *See* Ariz. Christian Sch. Tuition Org. v. Winn, 563 U.S. 125, 142-43 (2011).

312. *See* Zelinsky, *Tax "Benefits," supra* note 22, at 401-03.

313. *See* Chandelis Duster, *Top Democrats Disagree on Including Hyde Amendment in Economic Bill*, CNN: POL. (Oct. 3, 2021, 6:09 PM), [https://perma.cc/F7F5-7DG9];

Under the Hyde Amendment, congressional appropriations for Health and Human Services may not fund most abortions.[314] Congress has enacted some version of the Hyde Amendment since the country's bicentennial.[315] Thus, Congress has for nearly half a century prohibited direct federal subsidies for abortion under Medicaid.[316] But during that same time, Congress has consistently permitted Planned Parenthood Federation of America, Inc., the nation's leviathan of abortion services,[317] to maintain its federal income tax exemption as a § 501(c)(3) entity and to receive donations deductible by donors under IRC § 170.[318] Nothing in the law requires Planned Parenthood to pay for its abortions from sources distinct from revenues enhanced by these tax provisions. Congress manifestly does not equate these IRC sections with direct appropriations for the abortion procedures executed by Planned Parenthood. Surely Congress, better than any other governmental body, knows whether tax provisions are tantamount to direct appropriations of federal funds. After all, Congress creates them both.

Economic equivalence simply does not control the proper characterization of an item—as a tax expenditure or not—or the consequences of such characterization.[319] That fact alone limits its usefulness in tax expenditure analysis, constitutional or otherwise. But in constitutional cases, courts must qualify tax

---

Darragh Roche, *Joe Biden Heading Toward Clash with Progressives over Hyde Amendment*, NEWSWEEK (Oct. 6, 2021, 4:49 AM), [https://perma.cc/MLD9-AJ4Z].

314. Roche, *supra* note 313.

315. *See* Kelsey Snell, *Ban on Abortion Funding Stays in House Bill as 2020 Democrats Promise Repeal*, NPR (June 13, 2019, 5:01 AM), [https://perma.cc/2FJW-MF68].

316. For a brief history of the Hyde amendment, see Harris v. McRae, 448 U.S. 297, 300-03 (1980).

317. Planned Parenthood is reported to provide more abortions than any other abortion-services entity. *See* Julie Rovner, *Planned Parenthood: A Thorn in Abortion Foes' Sides*, NPR (Apr. 13, 2011, 12:02 AM), [https://perma.cc/TZY7-FQRW].

318. *Freqeuently Asked Questions: Is My Gift Tax Deductible?*, PLANNED PARENTHOOD, [https://perma.cc/RH9Z-ZAN8] (last visited Feb. 20, 2023).

319. *Cf.* Zelinsky, *Tax "Benefits*,*" supra* note 22, at 412 ("Any such equivalence is a conclusion of the inquiry, not a means to finding a solution.").

expenditure analysis for reasons, and in ways, not recognized by Surrey. Economic equivalence does not alter this imperative.[320]

L. The Question of Constitutional Norms

Tax expenditure theory and constitutional law have distinct normative foundations. One should not expect the same normative analysis to direct constitutional law and the selection of ideal tax bases for budgetary purposes. At least two major reasons explain why.

First, different tax systems themselves rely on and prioritize different norms. While some broad norms apply to all kinds of taxes (e.g., administrability, equity, and efficiency), how these norms apply differs across types of taxes.[321] For example, in an income tax, the norm of vertical equity[322] is much more likely to serve a prominent role than in a retail sales tax.[323] In the federal income tax, administrability is often invoked to justify the realization requirement because without it, annual valuations (or an economically similar proxy) would be necessary, and performing annual valuation is traditionally thought impracticable.[324] But under the typical ad valorem tax on real

320. *See* Sugin, *supra* note 9, at 472 (stating that "the economic equivalence of tax benefits and direct spending is not the most important factor" in interpreting the Establishment Clause).

321. For a classic exposition of income tax policy norms, see Joseph T. Sneed, *The Criteria of Federal Income Tax Policy*, 17 STAN. L. REV. 567 (1965).

322. For a discussion of vertical equity, see *id.* at 581-86, and Alvin Warren, *Would a Consumption Tax Be Fairer Than an Income Tax?*, 89 YALE L.J. 1081 (1980) (arguing that the norm of distributional equity does not compel the conclusion that a consumption base is superior to an income base).

323. Lower-income taxpayers generally pay a larger portion of their incomes in retail sales taxes than do higher-income taxpayers because the former simply cannot save as much of their income as can the latter. Further, except for special taxes on luxury goods, retail sales tax rates imposed on consumer goods are generally uniform. There is no progressive rate structure for purchases of Chicken McNuggets.

324. *See, e.g.*, JCT TAX EXPENDITURES REPORT, *supra* note 18, at 5. Numerous options that would at least partially address concerns based on the administrability of annual valuations are available, however. To name a few, the system could easily value financial assets that are publicly traded. Real estate could be presumptively valued consistently with local ad valorem taxation in many states. Other, non-wasting assets (e.g., art and collectibles) could be valued at cost or in accordance with valuation under the transfer tax system, annually modified to reflect a market-based return reflected in an agreed index. Alternatively, the proceeds of assets that are eventually sold could be adjusted upward by an interest factor to reflect the time value of money and the revenue loss to the government from

property, the administrability norm does not preclude annual valuations; they are routine.[325] The neutrality norm is often listed as a norm underlying income taxation,[326] but some taxes—for example, those imposed on sales of cigarettes—are intentionally non-neutral. These three examples—and countless, unnamed others—show that tax bases rely on different normative priorities. When tax bases themselves build on different normative frameworks, it is logically impossible for constitutional law to possess the same normative framework as every tax base.

Moreover, and just as importantly, constitutional law is governed by norms that often are not the same as those that undergird tax bases. Depending on the constitutional provision under consideration, key constitutional norms include government neutrality, free expression, liberty of movement, religious liberty, associational choice, autonomy, privacy, equal protection, separation of government powers, representative government, fairness of governmental process, and self-preservation. While an ideal tax base may reflect one or more of these norms, other norms serve important, even controlling, roles in selecting the base.

For example, the normative income tax base is customarily thought to begin with the Haig-Simons concept of income, modified to reflect certain norms of tax policy such as administrability and ability-to-pay and further qualified by what the public is willing to accept.[327] However, outside the context of the Sixteenth Amendment, the Constitution rarely requires courts to ponder the scope of economic income, let alone how Henry Simons thought of it.[328] The Constitution simply does not rise

---

the deferral of unrealized gains. For a thorough discussion of ways to implement accrual taxation under the federal income tax, see David J. Shakow, *Taxation Without Realization: A Proposal for Accrual Taxation*, 134 U. PA. L. REV. 1111 (1986).

325. *See, e.g.*, TEX. TAX CODE ANN. § 23.01(a) (West 2020) (requiring annual appraisals of taxable property).

326. *See, e.g.*, Fleming & Peroni, *supra* note 9, at 460-61.

327. *See* SURREY & MCDANIEL, *supra* note 4, at 186-88.

328. *See, e.g.*, Bittker & Kaufman, *supra* note 26, at 64 ("But if the Fourteenth Amendment did not enact Herbert Spencer's principles of political economy, it is equally unlikely that it was intended to enact the Haig-Simons definition of income.").

and fall with concepts of economic income. Other norms are foundational to constitutional law.[329]

A court thus should not expect or demand the norms that converge to identify an ideal tax base, and consequently a list of tax expenditures, to control the meaning and scope of numerous constitutional provisions. That a tax provision may be viewed as a form of indirect subsidy for budgeting purposes does not mean that it should be analyzed as a similar cash grant would be analyzed under norms of constitutional law.[330] To illustrate, in *Goldberg v. Kelly*,[331] the Supreme Court held that the Due Process Clause of the Fourteenth Amendment prohibits termination of cash welfare benefits without affording the welfare recipient the opportunity for an evidentiary hearing before the relevant government official has decided to cut off aid.[332] The Court's rationale was that "termination of aid pending resolution of a controversy over eligibility may deprive an eligible recipient of the very means by which to live while he waits."[333] That certain tax provisions, such as the earned income tax credit ("EITC"),[334] function as aid to the working poor [335] does not necessarily mean that an IRS official processing a tax return must grant the EITC claimant a hearing before deciding that she improperly claimed the credit. The EITC regime providing an indirect tax subsidy differs in meaningful ways from the strictly need-based welfare system subject to due process norms.[336]

---

329. *See, e.g.*, Michael W. McConnell, *Free Exercise Revisionism and the* Smith *Decision*, 57 U. CHI. L. REV. 1109, 1130-32, 1147 (1990) (arguing that courts should robustly interpret the Free Exercise Clause to ensure that religious minorities are truly treated neutrally along with religious majorities).

330. *Cf.* Zelinsky, *Tax "Benefits*,*" supra* note 22, at 383 (urging that courts analyze the equivalence (if any) between tax benefits and direct spending in part by considering "the perspective appropriate for the particular constitutional norm at issue").

331. 397 U.S. 254 (1970).

332. *See id.* at 264 ("[W]hen welfare is discontinued, only a pre-termination evidentiary hearing provides the recipient with procedural due process.").

333. *Id.* (emphasis omitted). For a thoughtful analysis of the intersection of equal protection and due process in contexts involving wealth inequality, see Brandon L. Garrett, *Wealth, Equal Protection, and Due Process*, 61 WM. & MARY L. REV. 397 (2019).

334. *See* I.R.C. § 32.

335. The earned income tax credit has been described as "the largest cash-transfer program for low-income workers with children." Anne L. Alstott, *Why the EITC Doesn't Make Work Pay*, 73 LAW & CONTEMP. PROBS. 285, 285 (2010).

336. Most obviously, the EITC is a delayed benefit, whereas welfare support is necessary for daily survival. *See Goldberg*, 397 U.S. at 264.

The EITC is a tax expenditure.[337] It functions even more closely to a cash grant than many other tax expenditures because it is refundable in cash—even if a taxpayer has no income tax liability for the year.[338] Thus, there is no credible argument that the EITC is necessary to reflect the normative income tax base. It is not just "like" a cash grant; it effectuates a literal transfer of cash in many cases. Further, it benefits the working poor, many of whom no doubt rely on it to pay for subsistence items. But the mere timing of the benefit provided by the EITC is likely enough to distinguish it from direct cash assistance for purposes of analyzing a recipient's due process rights when a government agent is determining eligibility. The constitutional norm is tied to the immediacy of the need of the recipient, not the mere fact that some type of subsidy exists.

The Court has sometimes appreciated the need for normative analysis when considering tax provisions. In *Walz v. Tax Commission of New York*,[339] the Court held that granting property tax exemption to religious organizations, among other charitable organizations, did not violate the Establishment Clause.[340] *Walz* observed the "indirect economic benefit" of tax exemption,[341] but distinguished it from direct subsidies.[342] An important portion of the Court's analysis was its assessment that the purpose of the property tax exemption was "neither the advancement nor the inhibition of religion"[343] and that it "creates only a minimal and remote involvement between church and state and far less than taxation of churches."[344] Thus, the Court invoked two constitutional norms that historically have guided interpretation of the Establishment Clause—neutrality and separation of church and state. The point is not that the Court correctly applied these

---

337. *See* JCT TAX EXPENDITURES REPORT, *supra* note 18, at tbl.1; TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 947-60; TREASURY REPORT, *supra* note 16, at 20.

338. *See* TAX EXPENDITURES: COMPENDIUM, *supra* note 21, at 947.

339. 397 U.S. 664 (1970).

340. *See id.* at 679-80.

341. *See id.* at 674.

342. *Id.* at 675-76.

343. *Id.* at 672.

344. *Walz*, 397 U.S. at 676.

constitutional norms,[345] but that the Court in fact applied them. The Court avoided the superficial reasoning that tax exemption equals cash subsidy, equals aid to religion, equals violation of the Establishment Clause.[346] Instead, the Court scrutinized the constitutional permissibility of religious property tax exemptions through the prism of norms underlying the Establishment Clause. In this respect, the Court's methodological approach was sensible.[347]

A final principle follows from the conclusion that a tax expenditure presenting constitutional questions should be analyzed under the norms that underlie the relevant constitutional texts. Just as the norms that inform selection of an ideal tax base may differ from the norms upon which a specific constitutional provision rests, different norms inform different constitutional clauses. Not all constitutional provisions rest on the same normative foundations.[348] Although some have argued for consistently characterizing tax expenditures in constitutional analysis,[349] a purported tax expenditure that raises a constitutional issue in one context may demand a different normative analysis when that same tax provision presents a constitutional issue in another context.[350] In any case, the appearance of the item on the tax expenditure list is not alone determinative. Courts should not expect a tool of fiscal policy developed primarily to assist the President and congressional budget committees to unify constitutional jurisprudence.

---

345. The better view is that the Court did correctly decide the case, but I reserve a comprehensive analysis of the Court's reasoning for another day. The argument I am advancing here is methodological, not doctrinal.

346. *Cf.* Bittker, *Churches*, *supra* note 26, at 1288 (arguing that the constitutionality of church tax exemptions depends on how they fare under the relevant Establishment Clause test and rejecting the idea that the mere existence of a church tax exemption implies a constitutional violation).

347. *See* Zelinsky, *Tax "Benefits*,*" supra* note 22, at 412 (arguing that the issue in *Walz* was properly analyzed under the meaning of the Establishment Clause, not by equating direct aid and tax expenditures).

348. *Cf.* McConnell, *supra* note 329, at 1137 ("Different clauses of the Constitution perform different functions and have different logical structures.").

349. *See, e.g.*, Adler, *supra* note 25, at 869, 886.

350. Professor Edward Zelinsky has argued this point well. *See* Zelinsky, *Tax "Benefits*,*" supra* note 22, at 413.

CONCLUSION

The greatest champion of tax expenditure theory, Stanley Surrey, powerfully influenced the budgetary operations of the executive and legislative branches of the federal government. He triumphantly made the case for employing tax expenditure theory in the federal budgeting process. His work merits study and respect.

However, Surrey unwisely proclaimed that tax expenditure theory also establishes the equivalence of tax expenditures and direct spending in constitutional analysis. He did so without even so much as exploring reasons to question the application of tax expenditure theory in constitutional adjudication, and even as he denied judges the right to exercise their own assessment of the nature of tax provisions appearing on executive agencies' tax expenditure lists. The mighty Surrey thereby erred, and he erred mightily.[351]

For numerous reasons, tax expenditure theory applies uneasily in constitutional contexts. Judges should not assign presumptive constitutional significance to tax expenditure designations by the JCT, the Treasury Department, and the OMB. The primary purpose of compiling tax expenditure lists—to serve as a tool in creating governmental budgets and exercising spending choices—has little or no relevance in adjudication. Problems of determining the normative tax base are magnified when judges must analyze the constitutional significance of a provision that an executive agency has conceptualized as indirect spending because of its notion of an unenacted ideal. These problems go to the heart of how judges decide hard cases, whether and to what degree they defer to the opinions of other branches, and even how they define their institutional role under the Constitution.

---

351. That Surrey erred in this respect does not detract from the significance of his body of work. His constitutional missteps do, however, reveal that he tried to extend tax expenditure theory too far. Alas, to return to the poet whose proverb introduces this article, "[t]o err is human, to forgive, divine." POPE, *supra* note 1, at 33; *cf. Isaiah* 53:6 ("We all, like sheep, have gone astray, each of us has turned to our own way; and the LORD has laid on him the iniquity of us all."); 1 *John* 1:8-9 ("If we claim to be without sin, we deceive ourselves and the truth is not in us. If we confess our sins, he is faithful and just and will forgive us our sins and purify us from all unrighteousness.").

Adding to these challenges, tax expenditure theory relies on a nebulous connection to public acceptance of a normative tax base that is in tension with the need for stability in constitutional law. Further, deferentially applying tax expenditure theory would often contravene existing constitutional doctrines that require an analysis of legislative purpose, the probable effects of laws, and the locus of appropriative power. The concept of tax penalties poses unique challenges to courts, who have thus far largely not understood it. Moreover, tax expenditure theory is underdeveloped in accounting for legal history, and this deficiency is especially problematic in constitutional cases that must be resolved in part by historical inquiries.

A court that must decide the constitutional implications of a tax provision that may be a tax expenditure must also guard against misanalysis on three fronts. One is the mistake of assuming that the form of the tax provision necessarily controls its nature. The concept of mechanical interchangeability aids in this regard. Another mistake to avoid is assigning significance to changes in the magnitude of tax expenditures that are attributable to adjustments to general tax rates and taxability thresholds. A third mistake to avoid is assuming that the economic equivalence of a tax provision and a direct grant means that the former is best conceptualized as a subsidy for purposes of constitutional law. A close analysis proves that this economic equivalence means little or nothing.

Finally, the normative foundations of tax expenditure theory and those of constitutional law are distinct. Courts should not expect the same normative analysis to direct constitutional law and the selection of ideal tax bases for budgetary purposes. To apply tax expenditure theory as though its normative principles correspond to those underlying one or more constitutional clauses is both naïve and illogical. Constitutional norms, history, and constitutional text must inform constitutional analysis—not a vision of economic income as modified by incommensurable tax policy norms and the willingness of the public to embrace some lesser variant of the vision.

RESCALING CITY PROPERTY

Amnon Lehavi*

INTRODUCTION

The city of Venice, Italy, is taking unprecedented measures in its efforts to address the flow of tourists coming into the city. While the number of permanent residents in the city's historic center is steadily declining—with less than 50,000 persons as of the end of 2022, down more than 120,000 residents from the 1950s—the city has been grappling with millions of visitors each year.[1] One key measure, which was anticipated to take effect in January 2023, is charging visitors for access and setting entrance quotas.[2] Under the scheme, locals, relatives of residents, and tourists who book accommodations in a Venice hotel are exempt from the fee.[3] This means that the entrance quotas and the fees—which are to be set between three and ten Euros depending on the season and how many tourists are expected on that day—are placed mostly on the masses of day-trippers, who are said to have crowded out local residents while causing various types of environmental costs and changing the city's ambiance.[4]

---

*Atara Kaufman Professor of Law, Harry Radzyner Law School, and Academic Director, G City Real Estate Institute, Reichman University, Israel. For helpful comments, I thank Manuel Aalbers, Klaas Heller, and Joanna Kusiak.

1. *See* Angela Giuffrida, *Venetians Fear 'Museum Relic' Status as Population Drops Below 50,000*, GUARDIAN (Aug. 10, 2022, 9:16 AM), [https://perma.cc/QQ82-7NLT]; Anna Momigliano, *Venice Tourism May Never Be the Same. It Could Be Better.*, N.Y. TIMES (Oct. 4, 2021), [https://perma.cc/ZJ6X-P6GU].

2. *See* Emma Featherstone, *Venice's New Tourist Rules Explained: How the Entry Fee of Up to €10 Will Be Charged and Who Will Be Exempt*, iNews (July 4, 2022, 4:50 PM), [https://perma.cc/C2LW-ELRA]; Alessandro Speciale, *Venice Set to Charge Tourists for Entry from Next Summer*, BLOOMBERG (Aug. 21, 2021, 3:50 AM), [https://perma.cc/YKE8-U2KF].

3. *See* Rebecca Ann Hughes, *Venice Sets Date for Introduction of Ticketing and Entry Fees - Here's What You Need to Know*, EURONEWS (Apr. 7, 2022), [https://perma.cc/SW55-CXEC].

4. Speciale, *supra* note 2. Another measure introduced in 2021 banned large cruise ships from the Venice lagoon in an effort to both prevent damaging waves in the sinking city and limit the actual number of day-trippers. *See id.*; Emma Bubola, *Venice, Overwhelmed*

Going beyond such physical gatekeeping, which is intended to both prevent over-crowding in the city and make visitors internalize at least some of the marginal costs they place on the city's decaying infrastructure, the city of Venice is increasingly using digital technology to monitor the movement of visitors within the city. Alongside surveillance cameras—which were originally installed to monitor for crime and reckless boaters, but now also serve the purpose of spotting over-crowding—the city is collecting cellphone data following a deal struck with TIM, an Italian phone company.[5] Such information allows the city not only to track movement patterns of cellphone users but also to distinguish between different types of crowds: namely, residents and persons who stay overnight at city hotels versus day-trippers who stay only for a few hours.[6] Answering concerns voiced about invasion of privacy and lack of consent by those monitored, the city argues that the information is anonymized and aggregated and that it serves as an essential planning tool not only to control over-crowding but also to promote a policy that encourages visitors to stay overnight at city hotels and, more broadly, serves its permanent residents—while others have proposed what they consider to be less invasive means of promoting this policy, such as seeking to attract highly educated and creative professionals to the city.[7]

Venice may be an outlier in its irregular proportions between the numbers of residents and visitors or in the types of environmental concerns it deals with due to subsidence and rising waters, but many of the challenges it addresses typify many other cities across the world. In particular, such challenges attest to the shifting borders of physical and digital urban governance—and to the fact that cities often lack the tools and powers to deal with their highly dynamic current reality.

This Article seeks to identify the growing tension between the contemporary physical and digital reality of cities across the world and the formal, often archaic, body of norms that governs

---

by Tourists, Tries Tracking Them, N.Y. TIMES (Oct. 4, 2021), [https://perma.cc/34VA-2H7N].

5. Bubola, *supra* note 4.
6. *See id.*
7. *Id.*

city powers and duties vis-à-vis different types of persons and corporations: locals, non-local residents of the same nation-state, and foreigners. The nation-state's continuing dominance, both in the domestic division of power across various legal systems and in the international arena, often results in a systemic mismatch between formal governance tools and urban practices.[8] Although there is a growing body of literature that illuminates cities' governance deficit—including in their legal relations vis-à-vis upper-level governments in monist or federal nation-states,[9] the lack of a constitutional status in most legal systems,[10] and their traditional absence from international law[11]—a formal "governance overhaul" for cities along these various dimensions is not yet in the making.

The Article argues that many aspects of the governance mismatch embedded in the cross-border nature of urban life can be addressed through a reconsideration of the role of cities' property rights in assets, both tangible and intangible ones.[12] Cities face growing pressure on physical assets that are consumed by both residents and non-residents, often with little power to ensure that those who use such assets will internalize the costs involved.[13] This may call for reconsidering the tenets of cities' property rights in governing these types of assets. At the same time, cities—especially "smart cities"—are increasingly accruing a new type of valuable asset: aggregated data about everyday actions of residents and non-residents, such as their movement patterns across the city.[14] While cities must account for concerns

8. *See* discussion *infra* Section I.A.

9. *See, e.g.*, RICHARD SCHRAGGER, CITY POWER: URBAN GOVERNANCE IN A DIGITAL AGE 1-17 (2016) (arguing that cities in the United States should be given more formal power to govern because cities should work to promote not only economic growth but also other policies such as social welfare and other forms of equality).

10. *See* RAN HIRSCHL, CITY, STATE: CONSTITUTIONALISM AND THE MEGACITY 11-12 (2020) (highlighting "the bewildering silence of contemporary constitutional discourse with respect to cities and urbanization").

11. Helmut Philipp Aust & Janne E. Nijman, *The Emerging Roles of Cities in International Law - Introductory Remarks on Practice, Scholarship and the Handbook*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CITIES 1, 1-11 (Helmut Philipp Aust et al. eds., 2021) ("For most of the twentieth- as well as early-twenty-first-century international law scholarship, it is apt to speak of an 'invisibility' of cities.").

12. *See infra* Sections II, III.

13. *See infra* Section II.B.

14. *See infra* Section II.C.

about privacy, mostly by ensuring that data is anonymized, or about any other potential abuse of cities' formal power or practical ability to collect information, it may make sense to allow cities to make use of such data, and even monetize it, under certain constraints.

The normative case for allowing cities to exert their property rights in assets in a differential and creative manner to address effectively their cross-border reality is particularly strong when cities remain otherwise committed to promoting intra-local and inter-local openness, diversity, and tolerance. Much of the concern about parochialism derives from the exclusionary practices of suburbs or private communities. Cities are—and should be—different. The multiple dimensions of cross-border activities taking place in cities today attest to their key role as a forum for inter-local and international mobility, exchange, and heterogeneity. Any normative or doctrinal reconsideration of city legal powers must preserve these essential traits. Allowing cities to harness their property rights, while otherwise preserving their general cross-border openness, may aid them in addressing their mounting physical and digital challenges.

This Article proceeds as follows. Part I seeks to identify the "urban governance gap." This term stands for the growing disparity between the traditional mandate of cities and other types of local governments to govern "local" matters that relate mostly to those who reside within their territorial boundaries and the current reality of cities, which requires them to address not only thematic issues that have been traditionally conceived as a "national" matter, such as human rights,[15] but also a growing scope of activities that concern non-resident persons and corporations.[16] It shows that while cross-border activities in cities have been mostly associated with "global cities"—referring to cities that serve as a hub of transnational commercial or financial activity—or with "international cities" that are the object of international tourism or other types of cross-border interest (due

15. *See* Martha F. Davis, *Finding International Law 'Close to Home': The Case of Human Rights Cities*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CITIES, *supra* note 11, at 227, 227-37 (showing how various cities embrace human rights norms as a guiding framework for local governance).

16. *See infra* Section I.C.

to historical, cultural, or geopolitical reasons), many more cities around the world are currently facing challenges that transcend local and even national boundaries.[17] Part I then briefly surveys some doctrines that have been developed in different legal systems for defining the scope of legal rights and duties of cities in handling such challenges, including in examining under which circumstances a city may favor its residents over outsiders in granting access to assets or in charging fees for using them.[18] It argues that piecemeal rules that have been adopted in the matter are often archaic and do not capture the essence of contemporary cross-border urbanism—one that deals not only with physical aspects but also with digital or "smart" governance of city life.[19]

Part II suggests that at least some of the dilemmas that concern the ability of cities to function well in a physical and digital cross-border reality may be resolved through a reconsideration of the role of cities' property rights in tangible and intangible assets. Contemporary legal systems have generally moved away from medieval and early modern conceptions of cities as "associations" or collective entities based on corporate charters, adopting a public/private distinction by which local governments are "public corporations" that "exist for public political purposes only . . . although they involve some private interests."[20] This does not mean, however, that property rights that cities have in assets do not matter. The acquisition or creation of assets by cities is instrumental in allowing them to perform many of their functions and can be distinguished—although not hermetically—from the power of governance or regulation that cities have over privately owned assets located in their territory.[21] Accordingly, the design of property law for city-owned assets must address the growing pressure on physical city assets, such as streets, parks, or cultural institutions. At the same time, it should conceptualize the proprietary features of a new

17. *See infra* Section I.A.
18. *See infra* Section I.B.
19. *See infra* Section I.C.
20. This distinction was adopted by the U.S. Supreme Court in *Trustees of Dartmouth College v. Woodward*, 17 U.S. 518, 668-69 (1819) (Story, J., concurring); *see also infra* notes 113-14 and accompanying text.
21. *See infra* notes 127-34 and accompanying text.

type of asset that many cities, and particularly "smart cities," are increasingly accruing and processing: mass digital data.

Part III sets out the normative framework for identifying the rights and duties that should be attached to city-owned assets. It argues that cities should generally promote openness, diversity, and tolerance.[22] In contrast to suburbs or private communities—and especially "gated communities"[23]—the case for validating cities' property rights is particularly strong when cities serve as a forum for inter-local and international mobility, exchange, and heterogeneity and when property entitlements are harnessed as a mechanism that enables the internalization of costs and benefits of serving different crowds.[24] If this is done properly, property law can solve many of the everyday challenges that cities face, ones that cannot—or should not—be resolved by a political overhaul of the city/state allocation of power.[25] Any reconsideration of the role of property law in the context of city assets must address the unique position of cities in identifying the scope of *in rem rights* and *in rem duties* regarding such assets and how these may apply to different categories of persons: locals, non-local nationals, and foreigners. Once property theory is employed to handle the current mismatch between governance powers and the cross-border physical and digital reality of cities, we can delineate the legal boundaries that would determine under which circumstances cities may take a differential approach in exercising the different "sticks" in the bundle of property.

I. IDENTIFYING THE URBAN GOVERNANCE GAP

A. Global Cities, International Cities, Megacities, and Other Cities

We live in an increasingly urbanizing world. While at the beginning of the twentieth century, only 12% of persons lived in

---

22. *See infra* Section III.A.

23. *See, e.g.*, Rowland Atkinson & Sarah Blandy, *International Perspectives on the new Enclavism and the Rise of Gated Communities*, *in* GATED COMMUNITIES: INTERNATIONAL PERSPECTIVES, at vii, vii-xi (Rowland Atkinson & Sarah Blandy eds., 2006).

24. *See infra* Section III.B.

25. *See infra* note 205-08 and accompanying text.

cities, today most of the world's population resides in them, and the pace of urbanization is steadily increasing.[26] In 2018, 55% of the global population lived in cities, and this number is expected to rise to 68% by 2050.[27] This increase is even more staggering considering the overall growth in the world's population—for example, within a single generation between the mid-1990s and late-2010s, the number of city dwellers nearly doubled to 4.3 billion, and this number is expected to increase to 6.7 billion by the year 2050.[28] Accordingly, while the world map is still formally divided into nation-states in the aftermath of the 1648 Peace of Westphalia, in today's world, most significant activities take place in cities.[29] They are the center of the world economy, responsible for 80% of global gross domestic product (GDP)—accounting for most knowledge production, data collection, and tech innovation but also for most pollution.[30] Cities are also the meeting place for cultural exchanges and many other interpersonal activities.[31]

Cities were never disconnected from their immediate and outer surroundings—although over the course of history, we can identify a significant growth in the qualitative and quantitative dimensions of cross-border activities that typify cities. This process is more dominant nowadays than ever before due to a variety of economic, social, geopolitical, and technological reasons.[32]

With their revival around the eleventh century, European cities and towns sought to establish a substantial degree of

---

26. Julia Zinkina et al., *The Nineteenth-Century Urbanization Transition in the First World*, *in* GLOBALISTICS AND GLOBALIZATION STUDIES 164, 166 (Leonid E. Grinin et al. eds., 2017); ORG. FOR ECON. COOP. & DEV., THE METROPOLITAN CENTURY: UNDERSTANDING URBANISATION AND ITS CONSEQUENCES 1 (2015), [https://perma.cc/4S7Q-BDN5] (summarizing the full publication, *The Metropolitan Century: Understanding Urbanisation and Its Consequences*).

27. *See Urban Development*, THE WORLD BANK, [https://perma.cc/FN5W-UME3] (last visited Jan. 2, 2023); Dept. of Econ. & Soc. Affs., World Urbanization Prospects 2018, U.N. Doc. ST/ESA/SER.A/421, at 5 (2019), [https://perma.cc/JB86-MUHZ].

28. HIRSCHL, *supra* note 10, at 2.

29. *See* RAFFAELE MARCHETTI, CITY DIPLOMACY: FROM CITY-STATES TO GLOBAL CITIES 1-2 (2021).

30. *Id.* at 2.

31. *Id.*

32. *See* Saskia Sassen, *The Global City: Introducing a Concept*, BROWN J. WORLD AFFS., Spring 2005, at 27, 27.

autonomy.[33] However, it was not the autonomy of a political institution, such as in contemporary systems of government, but that of a "complex economic, political, and communal association."[34] The medieval city or town was essentially a group of people, most dominantly merchants, who sought protection against outsiders to promote the interests of the group as a whole.[35] The autonomy of the group and its ability to establish its own rules were recognized in the legal status of the town, whether by a city charter or other arrangements made with the rulers of the larger territory in which the city was located.[36] While English towns had less autonomy vis-à-vis the king and his lords, in other places, such as Flanders or today's Italy and Germany, this autonomy allowed cities or towns to govern their own matters.[37]

Accordingly, each medieval city was governed according to its specific urban laws, which were usually granted to it as a privilege by the territorial king or lord who confirmed its foundation and awarded the urban community the right to modify or complement the communal statutes—although this power to change rules was at times subjected to limits placed by the king or lord.[38] Because each such city charter or equivalent instrument was granted to cities on an individual basis, along with the power to amend such rules over time, urban legal regimes were often very different from one another, which in turn required cities to engage in some type of coordination, especially to promote trade between cities.[39] One prominent example is the late Middle Ages' network of trade relations and political alliances between the cities of Northern Germany, known as the Hanse.[40]

33. *See* GERALD E. FRUG, CITY MAKING: BUILDING COMMUNITIES WITHOUT BUILDING WALLS 27 (1999).

34. *Id.*

35. *See id.* at 27-28.

36. *See id.*

37. Charles Angelucci et al., *How Merchant Towns Shaped Parliaments: From the Norman Conquest of England to the Great Reform Act*, 112 AM. ECON. REV. 3441, 3476-78 (2022).

38. Tobias Boestad, *Legitimizing Interurban Cooperation in the Middle Ages: The Legal System of the Hanse*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CITIES, *supra* note 11, at 29, 30.

39. *See id.* at 30-36.

40. *Id.* at 33-36.

Therefore, even during the high days of autonomous cities, including Italian city republics—such as Florence, Milan, and Venice—cities have always engaged in complex political, economic, and legal relations with neighboring or outer-area cities.[41]

More broadly, as Martin Loughlin notes, regardless of the formal level of self-governing powers that a certain city has attained vis-à-vis upper-level rulers at a certain point in time, the city "always existed in an integral relation to its surrounding area" such that it was never truly a "self-sufficient organism."[42] "In modernity, the city as a discrete unit of government" is practically overwhelmed—again, regardless of the specific formal allocation of power within a nation-state—by massive processes of urbanization that transformed the entire society, such that current urban society can be viewed as a "series of intersecting center-periphery networks."[43] This is particularly the case given the growth of "megacities" around the world and the "extensive urban agglomeration and population growth" in such cities.[44] The remarkable feature of contemporary megacities is not only the size of their population but also the concentration of people in them in proportion to the overall population of the country in which they are located—for example, the Santiago de Chile and the Taiwan-Keelung metropolitan areas are home to over one-third of their respective nations' overall populations.[45] Cities are also the hub of economic, cultural, and social life in nation-states, which means that beyond their residents, cities—and megacities in particular—engage daily with masses of out-of-city workers and visitors, such that cross-border action is ever-present in these cities. Accordingly, the geographical parameters of daily out-of-city commutes are ever-expanding, with major cities around the world seeing an increased number of "supercommuters"—people

41. *See* Susanne Lepsius, *The Legal System Among Italian City Republics*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CITIES, *supra* note 11, at 41, 44-50.

42. *See* Martin Loughlin, *The City in the Constitutional Imagination*, 72 UNIV. TORONTO L.J. 356, 359 (2022) (reviewing HIRSCHL, *supra* note 10).

43. *Id.* at 360.

44. HIRSCHL, *supra* note 10, at 7; *see also* Loughlin, *supra* note 42, at 356, 360.

45. HIRSCHL, *supra* note 10, at 6-7. There is no single definition for the term "megacity," with some sources placing the benchmark at five million residents, and others at ten million. *See id.*

who spend ninety minutes or more each way in their daily trek from outer-area jurisdictions into the city.[46]

The cross-border reality of cities exceedingly goes beyond national boundaries. Here, too, while some cities, such as Athens, Rome, and Alexandria, were a locus of political, economic, and cultural cross-border activity in ancient times,[47] in contemporary cities, this cross-border reality is typical of numerous cities, albeit to varying degrees. Saskia Sassen differentiates between "international cities,"—such as Venice and Florence, which are sources of cross-border interest and, accordingly, sites of international tourism due to their historical, cultural, or leisure-time importance—and "global cities" (or the similar concept of "world cities").[48]

According to Sassen, the term "global cities" refers to "strategic sites for the management of the global economy and the production of the most advanced services and financial operations that have become key inputs for that work of managing global economic operations."[49] Different from a world economy that was dominated by trade in goods, in a global economy currently dominated by international investment and financial instruments, the "[o]rganizing and servicing work once carried out by governments has since shifted" to global markets and specialized service firms (such as legal, accounting, finance, and insurance

---

46. Danielle Furfaro & Tamar Lapin, *NYC Has More than 600K 'Supercommuters'*, N.Y. POST (Apr. 25, 2018, 9:46 PM), [https://perma.cc/FV87-FB6H] (reporting that, in 2018, New York City had the highest overall number of such supercommuters, who constituted 6.7% of the city's overall workforce—the fourth highest rate in the country).

47. *See* GREG CLARK, GLOBAL CITIES: A SHORT HISTORY 38-39 (2016) (describing the role of global cities in the various continents in pre-modern times); *see also* LEWIS MUMFORD, THE CITY IN HISTORY: ITS ORIGINS, ITS TRANSFORMATIONS, AND ITS PROSPECTS 95 (1st ed. 1961).

48. *See* SASKIA SASSEN, THE GLOBAL CITY: NEW YORK, LONDON, TOKYO 350-51 (2d ed. 2001). For the use of the term "world cities," see John Friedmann & Goetz Wolff, *World City Formation: An Agenda for Research and Action*, 6 INT'L J. URB. & REG'L RSCH. 309, 310-11 (1982). As for international cities being sites of international tourism, it may be noted that in 2019, just prior to the COVID pandemic, the world's most visited city, Bangkok, Thailand (with 22.78 million visitors), was not considered a financial/economic "global city." *See* Alison Millington, *The 19 Most Visited Cities Around the World in 2019*, BUS. INSIDER (Sept. 5, 2019, 9:16 AM), [https://perma.cc/5UFM-GWBT]; Vsupsai, *Is Bangkok Ready to Become a 'Global City' Yet?*, ASIAN CITIES (Jan. 30, 2017), [https://perma.cc/7JV8-WKE4].

49. *See* SASKIA SASSEN, CITIES IN A WORLD ECONOMY 32 (5th ed. 2019).

firms).[50] Moreover, a growing number of corporate headquarters' functions are now being outsourced to such specialized corporate service firms.[51] This means that "today the production of headquarter functions of global firms has two sites: one is the corporate headquarters proper, and the other is the specialized service sector [that is] disproportionately concentrated in major cities."[52] These cities thus have an important physical component: they are the locus of both higher-end and lower-end jobs for such corporate service firms as well as for other types of businesses that indirectly serve these specialized firms.

This means that the current economy is not entirely fluid and a-local but rather that place and place-ness still matter and that the territorial dispersal of economic activity at the national and world levels generated by globalization creates "new forms of territorial centralization."[53] Global cities thus address the prospects but also the challenges of running a place that must provide a proper physical and digital infrastructure while serving diverse crowds beyond their residents.

Importantly, this urban reality is no longer the province of a small number of "established global cities," such as New York, London, Paris, or Tokyo.[54] Alongside them, commentators have identified at least two other categories. The first category is "emerging global cities," referring to capital cities and other big cities in large- or medium-sized emerging economies.[55] These cities function as the business capitals of their domestic economies and at the same time serve as gateways for international firms, trade, and investment.[56] The traded specialization of these emerging cities, and accordingly the pace at which they are developing as genuinely global hubs, are highly differential. Cities such as Shanghai, Shenzhen, and Taipei specialize in financial services and are becoming prominent global actors.[57] Other cities, especially in East Asia, depend

50. *Id.* at 32.
51. *Id.*
52. *Id.* at 32-33.
53. *Id.* at 33-35.
54. *See* CLARK, *supra* note 47, at 119-22.
55. *Id.* at 122.
56. *Id.*
57. *Id.* at 122-23.

heavily on their hardware and engineering capacities, while still others, such as São Paulo or Moscow, serve as regional centers of asset management, real estate, and research and development.[58] Then there are cities such as Mumbai, Manila, Jakarta, and Dhaka, which have global-city aspirations but are still low-wage economies and face acute infrastructure challenges.[59]

The second category is that of "new global cities"—which include "smaller, more specialized but highly globally oriented cities" with strong infrastructure and a high quality of life.[60] Melbourne and Boston are considered as quintessential examples.[61] Some "new global cities," such as Brisbane, Vienna, Munich, and Tel Aviv, increasingly excel in high-tech or innovation, while others are becoming globally competitive in sectors such as art and entertainment.[62]

But for all types of cities, processes of agglomeration, digitalization, and globalization should not be confused with fluidity, complete mobility, and a-localism. Various types of global cities, international cities, or national- or metropolitan-level urban hubs remain physical spaces that must provide assets and services, and the full scope of city life, while engaging in cross-border action.

B. Legal Boundaries of Governance vis-à-vis Non-Residents

Many cities around the world engage in daily activities that affect not only residents but also large and diverse groups of "outsiders": out-of-city or foreign workers, day-commuters, visitors, and, more generally, persons located in other nodes of global networks of which the city is part.

This reality brings forward numerous issues about the power of governance that a city has over non-residents and, particularly, under which circumstances a city can engage in a differential policy toward residents vis-à-vis some or all other groups of affected parties. This Article points to a growing mismatch

---

58. *Id.* at 123.
59. CLARK, *supra* note 47, at 123-24; *see also* MARCHETTI, *supra* note 29, at 29-30.
60. CLARK, *supra* note 47, at 126.
61. MARCHETTI, *supra* note 29, at 30.
62. CLARK, *supra* note 47, at 126-27; MARCHETTI, *supra* note 29, at 30.

between the everyday physical and digital reality of cities and much of established doctrine across many legal systems.[63] While this Article obviously cannot cover all major legal dilemmas that pertain to the scope of authority that local governments have toward such different groups, or otherwise offer an exhaustive comparative-law overview, the key point made here is that current doctrine is often captured by a paradigm that focuses almost solely on the city's residents as the subjects of legal norms (which I term a "resident-focused paradigm") while overlooking many other dimensions of urban cross-border reality.[64] Accordingly, when political decision-makers and courts deal with specific issues that come up in the context of non-residents, they often stick to principles that were developed under this resident-focused paradigm.[65]

In addition, cross-border issues relating to city governance are also regularly viewed through conventional accounts of the division of power between different types of governments—whether between cities and upper-level governments or between neighboring cities.[66] In the context of city versus upper-level governments, such disputes often relate to the subject matters that a city can decide on when such issues do not fall within a conventional "local" ambit. In the United States, such disputes are often channeled via the concept of "home rule," i.e., whether cities hold a general power to govern within their borders beyond specific types of responsibilities that have been delegated to them by upper-level governments.[67] This type of discourse also governs disputes that essentially relate to cities' legal powers regarding certain types of non-residents, as was the case with the political and legal clash between "sanctuary cities" and the Trump

63. *See* discussion *infra* Section I.C.

64. *See* discussion *supra* Section I.A.

65. *See infra* notes 72-85 and accompanying text.

66. For an analysis of potential frictions between neighboring local governments in the context of land-use decisions that create externalities across local borders, see Amnon Lehavi, *Intergovernmental Liability Rules*, 92 VA. L. REV. 929, 963-64, 967 (2006), which suggests that adversely affected local governments should be granted a right to monetary compensation for certain decisions.

67. For an analysis of the "home rule" concept (and competing approaches) in the United States and a call to grant cities more governance power to address current challenges, see Richard Briffault et al., *Principles of Home Rule for the 21st Century* 10-17 (Va. Pub. L. & Legal Theory Rsch. Paper No. 2020-16, 2020), [https://perma.cc/P2DV-9XXB].

administration.[68] But here too, the discourse often disregards the respective rights of persons that live across borders.

In this Section, I exemplify briefly how current legal line-drawing about the governance power of cities regarding residents, as opposed to some or all non-residents, relies essentially on a "resident-focused paradigm" and that it may fail to address the cross-border reality that many cities face today. Moreover, in considering current doctrine, it is important to bear in mind that designing a timely and inclusive legal policy that applies to non-residents is particularly important because this aspect of city power cannot be otherwise disciplined by the regular political process. When a city takes a certain decision that applies to its residents, the latter can react to such a decision by exercising their voice through elections or by other political tools available to them as residents.[69] This does not apply, certainly not on the same scope, to large and diverse groups of non-residents who may have little or no collective clout with local decision-makers or an upper-level government.[70] At the same time, adhering to the "resident-focused paradigm" may also work to unduly limit cities from ensuring that different categories of non-residents, who take

---

68. *See* Ronald Brownstein, *Trump's Battle with Sanctuary Cities Is the Next Phase of His Confrontation with Urban America*, CNN: POLS. (Apr. 16, 2019, 1:20 AM), [https://perma.cc/GBQ7-BSJD]. The term "sanctuary cities" refers in this context to U.S. local jurisdictions that were reluctant to embrace and implement federal immigration enforcement actions advanced by the Trump administration. *Id.* According to some commentators, this clash was part of a broader political agenda by the Trump administration to promote rural areas at the expense of cities. *See id.*

69. It is a different question whether the local political process is actually successful in holding cities accountable for such actions. Consider, for example, the longstanding debate in the context of land-use decisions whether majoritarian or minoritarian biases are likely to dominate local land use decisions or if the different political interests would be well represented by "voice" and "exit" mechanisms. For a more skeptical approach, see, for example, Robert C. Ellickson, *Suburban Growth Controls: An Economic and Legal Analysis,* 86 YALE L.J. 385, 404-10 (1977); see also Einer R. Elhauge, *Does Interest Group Theory Justify More Intrusive Judicial Review?*, 101 YALE L.J. 31, 39 (1991), for a discussion of the disproportionate impact of certain interest groups. For a more favorable approach, see Carol M. Rose, *Takings, Federalism, Norms*, 105 YALE L.J. 1121, 1131-35 (1996) (reviewing WILLIAM A. FISCHEL, REGULATORY TAKINGS: LAW, ECONOMICS, AND POLITICS (1995)) (criticizing process theory's "localism bashing" and arguing that "exit and voice options constrain local government" decisions).

70. Lehavi, *supra* note 66, at 941-942; *cf.* WILLIAM A. FISCHEL, THE HOMEVOTER HYPOTHESIS 184-86, 205-06 (2001) (arguing that inter-local externalities in such decisions are smaller in scope than inter-regional or inter-state ones).

part in city life, internalize at least some of the costs and benefits that their actions generate for the city.

To illustrate the currently lacking doctrine on city power vis-à-vis non-residents, this Section looks briefly at U.S. case law regarding a city's ability to favor its residents over outsiders. It does so in the context of access to city-owned property—by showing that instead of developing an updated jurisprudence on the normative features of proprietary rights and duties in a cross-border reality, this case law relies on often archaic concepts based on a "resident-focused paradigm."[71]

In its decision in *County Board of Arlington County v. Richards*, the U.S. Supreme Court held that a local ordinance that limited parking in a residential area to local residents did not violate the Equal Protection Clause of the Fourteenth Amendment.[72] Referring to the county's stated policy goals of reducing "air pollution and other environmental effects of automobile commuting" and of ensuring convenient parking for "residents who leave their cars at home during the day," the Court reasoned that "[t]he Constitution does not outlaw these social and environmental objectives, nor does it presume distinctions between residents and nonresidents of a local neighborhood to be invidious. The Equal Protection Clause requires only that the distinction drawn by an ordinance like Arlington's rationally promote the regulation's objectives."[73] Later decisions reviewing the legality of differentiating between residents and non-residents for purposes of street parking based on state constitutional equal-protection clauses or the common law "public trust doctrine"[74] have generally validated such measures.[75] New York's Court of Appeals adopted a different approach, however, by holding that "residents of a community have no greater right to use the highways abutting their land—whether it be for travel or parking—than other members of the public," and it is for the state

71. *See infra* notes 71-84 and accompanying text.
72. 434 U.S. 5, 6-7 (1977) (per curiam).
73. *Id.* at 7.
74. For a discussion of the public trust doctrine, see *infra* text accompanying notes 223-27.
75. *See, e.g.*, Martell's Tiki Bar, Inc. v. Governing Body of Borough of Point Pleasant Beach, No. 13-5676, 2015 WL 132559, at *1, *8-9 (D.N.J. Jan. 9, 2015) (discussing decisions which upheld the distinction based on New Jersey's constitution and common law).

legislature to create an exception or delegate the power to localities to do so.[76] Thus, in the context of parking privileges, it seems that federal courts and most state courts focus on whether the marginal effects of into-the-city commutes justify placing some of these costs on commuters—adopting an approach that I consider pragmatic and connected to cross-border reality.

However, for other types of city assets, courts have taken a different approach, which moves away from a pragmatic approach that looks to the cross-border reality of cities and relies instead on legal categories that were developed essentially under the "resident-focused paradigm." Thus, in *Leydon v. Town of Greenwich*, the Supreme Court of Connecticut held that a municipality was prohibited from allowing access to municipal parks only to its residents and their guests and that such a restriction placed on non-residents violated the First Amendment to the U.S. Constitution and corresponding provisions in the Connecticut Constitution.[77] The reason for invalidating this differential policy does not rely, therefore, on considerations of urban justice, resource distribution, or marginal effects of additional users but rather on freedom of speech.[78] The court resorted to the traditional public forum doctrine—one that distinguishes between "the traditional public forum, the designated public forum, and the nonpublic forum."[79] Identifying parks as belonging to the category of "the traditional public forum," the court reasoned that:

> In places which by long tradition or by government fiat have been devoted to assembly and debate, the rights of the State to limit expressive activity are sharply circumscribed. . . . [Such locations include] streets and parks which have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions.[80]

---

76. N.Y. State Pub. Emps. Fed'n v. City of Albany, 527 N.E.2d 253, 256 (N.Y. 1988).
77. 777 A.2d 552, 557 (Conn. 2001).
78. *See id.* at 559.
79. *Id.* at 566-67 (citing United States v. Frandsen, 212 F.3d 1231, 1237 (11th Cir. 2000)).
80. *Id.* at 567 (citing Hague v. CIO, 307 U.S. 496, 515 (1939)) (alteration in original).

Accordingly, the Connecticut court ruled that "the forum-based approach for First Amendment analysis subjects to the highest scrutiny the regulation of speech on government property traditionally available for public expression" and that "[t]he government can exclude a speaker from a traditional public forum only when the exclusion is necessary to serve a compelling state interest and the exclusion is narrowly drawn to achieve that interest."[81] The court concluded that in this case "the town ha[d] failed to explain why the ordinance's virtual ban on nonresidents is a reasonable time, place or manner restriction on the use of the park by such nonresidents."[82]

Adversely, in *Wright v. Incline Village General Improvement District*, the U.S. Court of Appeals for the Ninth Circuit held that because beaches are not "a traditional public forum," one that is "devoted to assembly and debate," the decision of the district to grant access to the beach that it owns and operates only to people who own or rent property within the boundaries of the district did not violate First Amendment rights of non-residents.[83] In addition, in reviewing an equal protection claim, the court held that such a limit on access is not "based on the content of speakers' messages" and does not infringe a fundamental right because "beaches are not a public forum."[84]

I do not entirely dismiss the possibility that non-residents who use various types of city-owned assets also find them as instruments for public expression and debate. But at the same time, this type of legal distinction, which was largely conceived in the context of a city's ability to regulate access to and use of city-owned properties mostly vis-à-vis its residents—especially in the context of using public spaces for political or social protest[85]—should not serve as the prominent legal benchmark in

81. *Id.* at 568, 571.
82. *Leydon*, 777 A.2d at 572.
83. 665 F.3d 1128, 1132, 1134 (9th Cir. 2011).
84. *Id.* at 1141.
85. For the history of the public forum doctrine, see, for example, Richard B. Saphire, *Reconsidering the Public Forum Doctrine*, 59 CIN. L. REV. 739, 739-44, 739 n.1 (1991), which suggests that "the doctrine's formalism provides both its underlying justification and its primary appeal"); Don Mitchell, *Political Violence, Order, and the Legal Construction of Public Space: Power and the Public Forum Doctrine*, 17 Urb. Geography 152, 152-55

drawing the boundaries between residents and non-residents. In the contemporary reality of cities, the various crowds of non-residents who have an interest in accessing tangible or intangible assets owned or controlled by the city are driven by a multitude of purposes that go well beyond their interest in political expression and public debates.

Accordingly, city policy toward various types of groups should reflect this current reality and primarily address matters such as the nature and scope of city agglomeration, the efficiency and equity of intra-city and inter-city allocation of resources, or how different types of users can be made to internalize the costs and benefits of their physical or digital participation in city life. While there is room for normatively considering which types of assets, or mechanisms for access/exclusion or determination of fees/prices, may be particularly prone to overt or covert forms of class-based discrimination or infringement of fundamental rights,[86] identifying the freedom of expression as the principal normative criterion seems to miss the mark in today's cities. Relying on such a criterion exemplifies the growing disparity between the traditional legal rules on urban governance and the physical and digital cross-border features of contemporary city life.

C. The Digital Mismatch in Cross-Border Urbanism

The discussion in the previous Sections has sought to highlight some of the changing features of many cities across the world—not only "global cities" or "international cities"—in dealing with the everyday mobility of persons in and out of the physical territorial borders of the city and the growing

(1996) (arguing that the development of the doctrine was made in relation to social struggles over and in public space that necessitated legal decision-making).

86. Consider, for example, a recent decision by the Israeli Supreme Court, by which a local council's decision to entirely prohibit access to a city-owned fitness center to non-residents, as opposed to the possibility of setting up a differential scale of fees for residents versus non-residents, should be measured by a multitude of empirical factors, such as the asset's scarcity, existing demand, availability of similar assets in neighboring jurisdictions, and so forth, such that an outright ban may amount to a disproportionate and discriminatory measure. CivA 8956/17 Mansour v. Local Council Kochav Yair Zur Yigal, Nevo Legal Database (Jan. 14, 2021) (Isr.).

implications that this may have for designing legal policy for various groups of persons.[87] Yet, what has made these cross-border dynamics particularly extensive over the past few decades has been the use of various forms of digital technology and related know-how in operating cities. In turn, this increases the mismatch between the formal scope of a city's governance power and the actual reach of its actions.

While many idioms have been used to portray the use of such digital technology, the term "smart cities" is the most prevalent term used to conceptualize cities that are increasingly composed of what Rob Kitchin has described as:

> [P]ervasive and ubiquitous computing and digitally instrumented devices built into the very fabric of urban environments (e.g., fixed and wireless telecom networks, digitally controlled utility services and transport infrastructure, sensor and camera networks, building management systems, and so on) that are used to monitor, manage and regulate city flows and processes, often in real-time, and mobile computing (e.g., smart phones) used by many urban citizens to engage with and navigate the city which themselves produce data about their users (such as location and activity).[88]

The underlying idea behind the use of such technology is that "[c]onnecting up, integrating and analysing" the information produced by various electronic devices and information-collection methods "provides a more cohesive and smart understanding of the city that enhances efficiency and sustainability" and that "rich seams of data . . . can [be] used to better depict, model and predict urban processes and simulate the likely outcomes of future urban development."[89] Accordingly, the idea of smart cities is linked to concepts such as "smart people, smart economy, smart governance, smart mobility," and "smart

87. *See* discussion *supra* Sections I.A, I.B.

88. Rob Kitchin, *The Real-Time City? Big Data and Smart Urbanism*, 79 GEOJOURNAL 1, 1-2 (2014). Kitchin refers also to another meaning of the term, by which "a smart city is one whose economy and governance is being driven by innovation, creativity and entrepreneurship, enacted by smart people." *Id.* at 2. Kitchin focuses his attention, however, on the technological meaning of the term—as does most of the other academic literature on smart cities. *See id.* at 1-2.

89. *Id.* at 2.

environment."[90] In particular, the concept of "smart governance" is one that "encourag[es] people or citizens to participate and collaborate in smart cities," and accordingly, "[t]he successful government in smart cities depends on providing city services, channels, smart mobile services, and network integration to the citizens."[91]

Moreover, the scope of digitalization pertaining to assets, services, and actions taking place throughout the city is such that many cities now aspire to set up a "digital twin," meaning that all physical layers of the city will be complemented by and synchronized with a digital layer.[92] Further, the city of Seoul announced its plans to enter the metaverse and establish a platform for "contactless communication."[93] By 2023, the city will open its "Metaverse 120 Center," which will serve as a "virtual city hall" where residents meet (in their avatar forms) with local officials.[94]

What does this new type of digital reality mean for the actual scope of city action and its potential application to various crowds of non-residents—and how disconnected is this new reality from traditional modes of formal territorial governance? I suggest that the mismatch, or the "urban governance gap," is even more pronounced here than in the case of physical cross-border action. This is so because the actual scope of control and information-gathering in regard to both residents and non-residents is qualitatively and quantitatively different from physical modes of city action.

The most pressing concern with the practices of "smart cities" has to do with the potential ill effects of digital

90. Ayca Kirimtat et al., *Future Trends and Current State of Smart City Concepts: A Survey*, 8 IEEE ACCESS 86448, 86450 (2020).

91. *Id.* at 86451; *see also* Kai Cao et al., *Big Data, Spatial Optimization, and Planning*, 47 ENV'T & PLAN. B 941, 941-43 (2020).

92. *See, e.g.*, Hui Zeng et al., *Your City Is Mirrored by a Digital Twin - Advanced Version 2.0 of Interactive City: Integrating Physical World and Digital World*, *in* FUTURE CITIES, NEW ECONOMY, AND SHARED CITY PROSPERITY DRIVEN BY TECHNOLOGICAL INNOVATIONS 46, 46-47 (Lei Guo et al. eds., 2020).

93. Linda Poon, *Navigator: After the Pandemic Comes . . . the Metaverse?*, BLOOMBERG (Nov. 13, 2021, 8:40 AM), [https://perma.cc/BZZ7-TLGS].

94. *Id.*

surveillance—specifically the infringement of privacy.[95] While city life has been traditionally defined by at least some level of "public anonymity,"[96] in smart cities today, literally every step that persons make around the city can be tracked. Beyond identifying general movement patterns, digital data can expose many other personal details, with almost no legal regulation curtailing the city's ability to collect such information.[97] Moreover, unlike pieces of information that are genuinely "volunteered" by persons, such as when they participate in online polls or other forms of "crowdsourcing" initiated by the city, or alternatively, when persons are actively aware that personal data is being gathered from them, such as when they are questioned by a police officer, most forms of data collection in the "smart city" are "automated."[98] These forms include mobile phones and other devices that record and communicate their location and history of their use, "clickstream data" that records how people navigate through a website or app, transportation "smart cards," automatic number plate recognition (ANPR), automatic meter reading (AMR) that communicates utility usage without the need for manual reading, and so forth.[99]

While the city's residents have political means to address concerns over the invasion of privacy—such as by elections, referenda, or other opportunities for exercising "voice"—these measures are practically nonexistent for non-residents, especially for dispersed crowds of work-commuters, visitors, and so forth.[100] Accordingly, scholarly accounts about the prospects and perils of

95. *See, e.g.*, Lilian Edwards, *Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective*, 2 EUR. DATA PROT. L. REV. 28, 28-29 (2016); *see also* Liesbet van Zoonen, *Privacy Concerns in Smart Cities*, 33 GOV'T INFO. Q. 472, 472-73 (2016).

96. *See* Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213, 214-18 (2002) (calling to protect "the right to public anonymity" from camera surveillance in public places).

97. *See* JOSHUA A. T. FAIRFIELD, OWNED: PROPERTY, PRIVACY, AND THE NEW DIGITAL SERFDOM 67 (2017) ("Massive public infrastructural surveillance is disturbing and the reach and scale of the technology is new. Legal rules are not in place, and the potential for abuse is significant.").

98. Kitchin, *supra* note 88, at 4-5 (distinguishing between "directed, automated and volunteered" data).

99. *Id.*

100. *See supra* notes 68-69 and accompanying text.

"citizenship in smart cities" tend to focus on the political community of residents-voters.[101]

Moreover, as far as non-residents are concerned, not only are such digital devices able to identify pieces of data that are particularly relevant to cities, such as by tracking out-of-city commuting patterns for work and non-work destinations within the city,[102] but also, more generally, such devices can collect data on activities that are taking place outside the territorial borders of the city. Thus, for example, the automatic reading of transportation "smart cards" allows the city to track the entire itinerary of the user, well beyond transit points within the city's physical borders.[103] Data aggregation and control, therefore, exceed political boundaries, with no corresponding accountability by the city.

All of this means that the mismatch between the digital scope of city activity and traditional legal rules of territorial governance is especially pronounced. However, as already noted in the Introduction, we are unlikely to see a political reshuffle of governance powers across local borders on the national level, and even more so on the international level, to handle this governance gap.

At the same time, the case of mass digital data in smart cities also exemplifies a potential avenue to rescale city actions without waiting for a local-national-international "governance overhaul." This is so because many cities are now beginning to consider if and how to monetize the aggregated data—meaning that such cities now look at such data not only as a tool for effective

---

101. *See, e.g.*, PAOLO CARDULLO, CITIZENS IN THE 'SMART CITY': PARTICIPATION, CO-PRODUCTION, GOVERNANCE 58-59 (2021); Martijn de Waal & Marloes Dignum, *The Citizen in the Smart City. How the Smart City Could Transform Citizenship*, 59 INFO. TECH. 263, 265-69 (2017); Sofia Ranchordàs, *Nudging Citizens Through Technology in Smart Cities*, 34 INT'L REV. L. COMPUTS. & TECH. 254, 255 (2019).

102. *See, e.g.*, Roberto Ponce-Lopez & Joseph Ferreira Jr., *Identifying and Characterizing Popular Non-Work Destinations by Clustering Cellphone and Point-of-Interest Data*, 113 CITIES 103158, at 1-2, 4-10 (2021) (identifying, through cellphone tracing and a Google Maps application, the most popular places in Singapore that attract the greatest number of non-work visitors during each hour of the week).

103. *See, e.g.*, Victor Chang, *An Ethical Framework for Big Data and Smart Cities*, 165 TECH. FORECASTING & SOC. CHANGE 120559, at 1-2, 6 (2021) (illustrating the wide-ranging competencies of big data analytics in public transportation and addressing ethical concerns that these capabilities raise).

governance or policy-making but also as a revenue-generating asset that can be sold or otherwise traded with third parties.[104] This means that aggregated data is viewed as a potential asset owned or controlled by the city. As shown in Section II.C, while any normative consideration of whether to validate any such claim of entitlement by cities must address the legitimate interests of all relevant parties, a property analysis that establishes *in rem rights* but also extensive *in rem duties* on cities in dealing with such data may have the advantage of operating on the actual scales in which data is gathered.[105]

II. RECONSIDERING THE ROLE OF CITIES' PROPERTY RIGHTS

A. City Property and the Public/Private Distinction

As shown in Section A.1, medieval European towns and cities acted as associations in which the collective entity held both political privileges and economic entitlements.[106] The town or city defended its inhabitants from outsiders but also controlled all resources within its boundaries.[107] In England, the city's corporate power was based on a corporate charter granted by the king, and the rights acquired by the charter were considered to be property rights that were deemed essential for protecting both political and economic interests.[108] Over time, and particularly with the rise of Parliament as the dominant political force in England, the status of the city charter and the rights granted by it became more controversial, especially over the political aspect of self-governance.[109]

In colonial America and the early United States, cities and towns came to be viewed as corporations, although they had not been initially formed under a charter, as was the case in

104. *See, e.g.*, Benjamin Freed, *'Smart Cities' Contemplate Turning Big Data into Big Money*, STATE SCOOP (Apr. 3, 2019), [https://perma.cc/GT73-JV67]; see also *infra* notes 157-62, for additional resources.

105. *See infra* notes 186-93 and accompanying text.

106. *See supra* notes 33-41 and accompanying text.

107. *See* FRUG, *supra* note 33, at 27-28.

108. *See id.* at 34-35.

109. *Id.* at 32-36.

England.[110] Moreover, cities were unlike business corporations that had been granted specific charters by the colonies, and later by state legislatures, entitling them to engage in certain kinds of enterprises, such as building and operating bridges, canals, water supply, or banking activities.[111]

As courts gradually developed protections for investors' property in business corporations—leading, accordingly, to pressure on the legislature to extend the opportunities for incorporation from a favored few enterprises to the more general population—the actual conflation of business corporations and of cities as corporations became more problematic. State legislatures wanted to assert more political control over cities and towns and realized that granting them the type of strong protection from interference that was awarded to business corporations would run counter to this goal.[112] This process led to the development of the public/private distinction, anchored in the 1819 U.S. Supreme Court's decision in *Trustees of Dartmouth College v. Woodward*, in which the Court distinguished between private corporations that are founded by individual contributions of property by investors and public corporations that are founded by the government without such individual contributions.[113] As Justice Story explained: "Another division of corporations is into public and private. Public corporations are generally esteemed such as exist for public political purposes only, such as towns, cities, parishes, and counties; and in many respects they are so, although they involve some private interests . . . ."[114]

Cities, as "public corporations," were therefore reconceptualized as entities that are part of the political system of government—and are accordingly subject to the political hierarchy of state and federal entities—with ensuing controversies that continue to date about the scope of power that

110. *Id*. at 36-40.

111. *See id.*

112. *See* FRUG, *supra* note 33, at 38-39.

113. *Id.* at 40-42; *see also* Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 518, 562-63 (1819).

114. *Woodward*, 17 U.S. (4 Wheat.) at 668-69 (Story, J., concurring).

cities and other local governments should have to act without explicit state/federal authorization.[115]

This does not mean, however, that property rights that cities hold in assets no longer matter or that they are simply a derivative of the scope of the city's political power of governance. As Chancellor Kent wrote in his *Commentaries on American Law* as early as 1826, although the powers of governance of local governments are subject to the control of the state legislature, "[t]hey may also be empowered to take or hold private property for municipal uses, and such property is invested with the security of other private rights."[116]

Accordingly, property rights that cities acquire or otherwise hold in assets generally include the same kind of protection that private owners enjoy under the relevant legal system such that these rights also apply vis-à-vis upper-level governments. For example, in the United States, courts have held that when the federal government exercises the power of eminent domain over property owned by state or local governments, it is required to pay compensation according to the U.S. Constitution's Takings Clause, although this Clause does not make specific reference to publicly owned property.[117] In *United States v. 50 Acres of Land*, the Supreme Court reasoned that "[w]hen the United States condemns a local public facility, the loss to the public entity, to the persons served by it, and to the local taxpayers may be no less acute than the loss in a taking of private property."[118]

At the same time, in applying its property rights in assets vis-à-vis private parties—whether individuals or firms, residents or non-residents—the city may be subjected in principle to certain public or constitutional law duties, as demonstrated in Section I.B in the context of establishing differential rules for street parking or access to city-owned parks or beaches.[119] But any such limits on exercising the city's property rights should not be simply

115. *See* Briffault et al., *supra* note 67, at 13-17, 20 (calling to broaden the principle of home rule for cities in U.S. law).

116. 2 JAMES KENT, COMMENTARIES ON AMERICAN LAW 275 (New York, O. Halstead 1826).

117. *See* U.S. CONST. amend. V, cl. 4; *see, e.g.*, United States v. 50 Acres of Land, 469 U.S. 24, 31 (1984).

118. 469 U.S. at 31.

119. *See supra* notes 72-84 and accompanying text.

conflated with the general scope of its political power of governance or with the city's power to regulate privately owned properties.

Cities' property rights should have meaning and content that are not simply a derivative of such governance or regulatory powers. At the same time, their content should reflect the unique role of cities as public corporations, which is distinguished from that of private firms.[120] As I show in Part III, this calls for a normative reconsideration and re-delineation of cities' property rights, which identify not only the scope of *in rem rights* but also that of *in rem duties*.[121] Such an approach also has the practical advantage of working on the same scales—physical and digital—in which city-owned assets operate in practice in a world that is increasingly typified by cross-border urbanism without undermining political or regulatory realms governed by the political system.[122]

This approach resonates with the general mandate, awarded to a municipal corporation across many different legal systems, to "acquire needed property, real or personal, for its use and benefit as a local governmental organ."[123] Such power may also extend to "the holding of land or other property located beyond the limits of the municipal corporation."[124] Accordingly, under U.S. law, "[t]he two general classes of property which a municipal corporation may hold include, first, that property essential or convenient for it to function; second, property held for general convenience, pleasure and profit."[125] That said, it is otherwise within the province of the relevant upper-level government (namely, the state or federal legislature) "to declare what is a municipal purpose."[126]

What this means, more fundamentally, is that unlike private corporations, which are generally entitled to engage in any kind of business and are at liberty to acquire any type of asset they

120. *See* Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 518, 562-63 (1819).

121. *See infra* notes 197-200 and accompanying text.

122. *See infra* note 205-06, 242-45 and accompanying text.

123. 10 EUGENE MCQUILLIN ET AL., MCQUILLIN MUN. CORPS. § 28:2, Westlaw (database updated July 2022).

124. *Id.*

125. 10 MCQUILLIN ET AL., *supra* note 123, § 28:10.

126. *Id.*

deem fit (subject to specific limits, such as relevant rules of antitrust law),[127] the purposes for which cities may acquire, use, and profit from assets should be in line with the general purposes for which local governments operate. Beyond the abovementioned power of upper-level governments in various legal systems to set the broad outlines of "municipal purpose[s],"[128] the operation of cities in acquiring and holding assets should be subject to a more fundamental normative analysis. Such an analysis should consider, on the one hand, the multiple tasks that cities are expected or practically required to take up in current times and, on the other, the need to guard against abuses of property when the city leverages its governance or monopolistic powers.[129]

Thus, for example, the current approach toward allowing cities to engage in certain for-profit business activities, by providing services such as a cable TV station or a garbage collection service that competes with private businesses that operate within the territory of the city, is generally permissive—as long as the city does not abuse its governmental or regulatory powers to drive out or otherwise impair the ability of private businesses to compete with the city-owned enterprise.[130] In contrast, a city would be off-limits in acquiring real estate merely to make a speculative profit.[131]

More broadly, cities engage in an increasing number and variety of transactions involving assets that are owned or controlled by them. While such actions do not fall within the formal ambit of their governance or regulatory powers—and are part of the city's "proprietary" capacity—these dealings cannot be hermetically detached from the city's public-governance roles.

127. *See* Herbert Hovenkamp, *The Sherman Act and the Classical Theory of Competition*, 74 IOWA L. REV. 1019, 1044 (1989) (explaining that U.S. antitrust law was created as a tool to restrict anticompetitive behavior in the modern capitalist economy, and it therefore justifies placing certain limits on businesses' rights to buy or sell property).

128. *See* 10 MCQUILLIN ET AL., *supra* note 123, § 28:10.

129. For the scope of applicability of antitrust law rules to local governments in the United States, see Max Schanzenbach & Nadav Shoked, *Reclaiming Fiduciary Law for the City*, 70 STAN. L. REV. 565, 600-02 (2018).

130. *See generally* GERALD E. FRUG ET AL., LOCAL GOVERNMENT LAW 732-33, 742-61 (4th ed. 2005) (discussing the concept of increasing municipal economic power through "greater use of the entitlements that cities have as property owners").

131. *See* 10 MCQUILLIN ET AL., *supra* note 123, § 28:10.

One prominent example is the city of Chicago's decision in 2008 to sell to an investors' consortium the right to generate all revenues from its parking meters for a period of 75 years against a lump-sum payment of $1.157 billion.[132] While this deal raised public attention because of the contention that the deal severely under-valued this asset's true worth (a contention validated when the consortium already recouped the price in 2021, with 62 years left in the 75-year lease),[133] it points more broadly to the types and magnitude of city assets that are the object of market transactions with private corporations or of other types of asset monetization—such as public-private partnerships and other forms of privatization of activities formerly undertaken by the city.[134]

In other words, in addition to the key type of "urban governance gap" identified in this Article—that which concerns the cross-border nature of city life—I argue that devising a comprehensive proprietary framework of cities' *in rem rights* and *in rem duties* in tangible and intangible assets can also handle another aspect of the urban governance gap. This latter aspect refers to the growing scope of asset production, utilization, and monetization that does not fall within cities' governmental or regulatory powers but cannot be entirely divorced from the underlying normative considerations that should guide cities in acting vis-à-vis both residents and non-residents.

What this means is that there is growing importance in reconsidering the underlying normative parameters for enabling cities to acquire or otherwise hold certain types of property and the general purposes for which such assets may be used or monetized so that cities can exercise their property rights in a way that also conforms to their general roles as public-governance entities. This point is demonstrated in the next Sections, which

---

132. For the details of this deal, and the (unsuccessful) legal and public challenges that followed, see Schanzenbach & Shoked, *supra* note 129, at 567-70, and Indep. Voters of Ill. Indep. Precinct Org. v. Ahmad, 13 N.E.3d 251, 252-55 (Ill. App. Ct. 2014).

133. *See* Fran Spielman, *Chicago Parking Meter Investors Rake in $13M in Profit Despite Pandemic*, CHI. SUN TIMES (June 7, 2021, 5:30 AM), [https://perma.cc/HKU2-T5Q5] (reporting that by June 2021 the consortium recouped the investment plus $500 million more).

134. *See* Schanzenbach & Shoked, *supra* note 129, at 570-72 (pointing to numerous types of such dealings).

deal, respectively, with physical city assets and the new type of digital assets that "smart cities" are accruing and considering to monetize: data.[135] Part III will then lay out the proposed principles for rescaling cities' in rem rights and in rem duties in regard to city property.[136]

B. The Growing Pressure on Physical City Assets

The general trend of increasing urbanization throughout the world is placing a corresponding pressure on the physical infrastructure of cities and on the demand for services that are provided to a considerable degree through assets that are owned or controlled by cities. While some types of services and assets are intended primarily for residents—such as education, at least under systems of government that assign these tasks primarily to local governments—other types of services and city assets practically serve diverse crowds of both residents and non-residents.

One prominent example is transportation. Across all countries, but especially in developing countries in which the pace of urbanization has been significantly higher than in developing countries,[137] population growth has turned transportation from, to, and within urban areas into a major challenge.[138] The need to address the growing traffic of persons and freight within and across city boundaries requires cities to engage in schemes of "city logistics" that are intended to optimize the use of city resources to facilitate the goals of "mobility, sustainability, and liveability" while lowering "costs for customers as well as reducing negative environmental impacts

135. *See* discussion *infra* Sections II.B, II.C.

136. *See* discussion *infra* Section III.B.

137. *See* HIRSCHL, *supra* note 10, at 4 ("[A]pproximately 88% of urban population growth since 1960 has taken place in the developing world, meaning that about 9 of every 10 new urban dwellers since 1960 reside in Asia, Africa, or Latin America.").

138. Evans Mwamba et al., *Dynamic Effect of Rapid Urbanization on City Logistics: Literature Gleened Lessons for Developing Countries*, 3 J. CITY & DEV. 37, 38 (2021). Similar challenges for cities may apply to different types of services and city assets, such as energy supply, with developing countries facing particularly strong pressure in the face of the rapid pace of urbanization. *See, e.g.*, Nina Savela et al., *Rapid Urbanization and Infrastructure Pressure: Comparing the Sustainability Transition Potential of Water and Energy Regimes in Namibia*, 1 WORLD 49, 49-50 (2020) (examining the case of Namibia).

and improving safety."[139] The city-logistics strategy is proving particularly essential for megacities[140]—which, as already noted in Section I.A, are typified not only by a high number of residents but also by the constant inward and outward flow of multiple categories of non-residents (such as work commuters, persons who require onsite services from government offices or other types of businesses located in the megacity, shoppers, visitors/tourists, etc.).[141] Accordingly, regardless of various questions that may arise about the power of governance of cities over non-residents—questions that may remain unresolved or otherwise ambiguous—cities are practically required to acquire, manage, and operate transportation-related assets in the service of various crowds.

Particular attention should be paid in this context to various schemes of road-traffic pricing that are currently operated by various cities across the world. The key purpose of such pricing is to address various types of externalities that are caused by motor vehicle traffic—most prominently congestion (namely, the externalities that drivers inflict on one another by causing time delays and related costs) but also social costs related to accidents, air pollution, and so forth.[142]

Academic literature has been dealing with road traffic pricing since William Vickrey's work, which called to introduce road pricing based on the principles of Pigouvian welfare economics.[143] Under this theory, road pricing has an advantage over command-and-control policies such as outright bans on driving or restrictions on the days when a vehicle can be driven.[144] This is so because road pricing induces adjustments in trip frequencies, destination, mode (e.g., moving to public

---

139. *See, e.g.*, Eiichi Taniguchi et al., *Recent Trends and Innovations in Modelling City Logistics*, 125 PROCEDIA SOC. & BEHAV. SCIS. 4, 4-5 (2014).

140. *See id.* at 5 (pointing to the particularly complicated problems of city logistics that megacities may face).

141. *See supra* notes 43-46 and accompanying text.

142. *See* Alex Anas & Robin Lindsey, *Reducing Urban Road Transportation Externalities: Road Pricing in Theory and in Practice*, 5 REV. ENV'T ECON. & POL'Y 66, 66-68 (2011).

143. Alex Anas, *The Cost of Congestion and the Benefits of Congestion Pricing: A General Equilibrium Analysis*, 136 TRANSP. RSCH. 110, 111 (2020); *see also* William S. Vickrey, *Pricing in Urban and Suburban Transport*, 53 AM. ECON. REV. 452, 461-65 (1963).

144. Anas & Lindsey, *supra* note 142, at 67.

transportation), time of day, and long-run location decisions.[145] Moreover, road pricing can be changed according to the type and magnitude of the congestion externality.[146]

Moving from theory to practice, congestion pricing has become increasingly palatable because of the rising costs of congestion in big cities that are typified not only by a high overall volume of traffic but moreover by the constant inward and outward flow of traffic.[147] Accordingly, cities such as Singapore, London, and Stockholm have introduced congestion pricing in their central areas.[148] These mechanisms apply to either certain cordons leading to and from the central city or Central Business District (CBD), as is the case in Singapore and Stockholm, or for any car movement within a certain zone located around the city center, as is the case in London.[149] Charges vary based on the day and time of travel, with certain types of vehicles being exempt from tolls.[150]

Importantly, the congestion pricing schemes that apply in Singapore, London, and Stockholm are different from most other types of toll roads that are prevalent around the world. For the latter, "toll revenues are used either to cover maintenance and amortize construction costs or to make a profit for private operators."[151] As such, congestion-based, or more generally externalities-based, pricing is not only more effective in lowering congestion costs (and to a lesser degree, environmental costs),[152] but it also has merit in promoting efficient and fair cross-border urbanism and is normatively superior in promoting the goals that cities should serve as proprietary owners.

---

145. *Id.*

146. *Id.*

147. *See* Anas, *supra* note 143, at 111.

148. For a review of these specific road pricing schemes, see Anas & Lindsey, *supra* note 142, at 68, 71-77.

149. *Id.* at 71-76.

150. *Id.* at 72-73, 75.

151. *Id.* at 71.

152. Anas & Lindsey, *supra* note 142, at 77 (showing that environmental benefits of road pricing amount "to only a small fraction of the benefits from drivers' time savings"); *see also* Anas, *supra* note 143, at 120-24, 126 (showing, in the case of Greater Los Angeles region in California, that toll pricing has relatively small effects on long-run location decisions by persons and residents).

This is so because adopting a congestion-based pricing mechanism for using physical city-owned assets, such as transportation arteries, is generally more transparent, and it better promotes the role of the city as an open place that serves various crowds, including non-residents. A genuine externalities-based pricing mechanism undermines the ability of cities to engage in various forms of unwarranted parochialism by categorically favoring certain types of users regardless of an analysis of the marginal costs and benefits embedded in using city-owned resources (even if cities properly account for costs imposed on residents via local taxes or levies). Accordingly, empirical studies on the distributive impacts of congestion-based pricing show that the balance of winners and losers is far from being one that necessarily favors city insiders over outsiders or high-income households over low-income ones.[153] Thus, for example, the "availability of public transit limits the welfare losses of low-income groups," such that they may gain from such schemes when toll revenues are reinvested by the city to improve public transportation.[154]

Therefore, cities can act as asset-owners in a way that allows them to internalize the marginal costs and benefits they incur as property owners and operators of physical assets—requiring non-proprietary users to do the same through a pricing mechanism[155]—while at the same time maintaining their role as a hub for activities by diverse crowds of residents and non-residents. Identifying the growing pressure that cities face on assets they own or control while considering the broader goals that cities should promote in the current urban environments may set the stage for filling potential "governance gaps" through reconsidering the scope of cities' property rights.

153. *See* Benjamin Bureau & Matthieu Glachant, *Distributional Effects of Road Pricing: Assessment of Nine Scenarios for Paris*, 42 TRANSP. RSCH. 994, 994-95, 1005-06 (2008).

154. Anas & Lindsey, *supra* note 142, at 79.

155. *See* Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347, 348-49 (1967). Internalization serves as a key justification for the delineation of property rights in assets. Harold Demsetz's classic work discusses the evolution of property rights to handle externalities and related problems of under-investment and over-use of scarce resources when property rights do not exist or are not enforced. *See generally* Demsetz, *supra*.

C. The New Property for Cities: Aggregated Data

As shown in Section I.C, many cities around the world, typically dubbed "smart cities," engage in collecting mass amounts of data through various digital technologies about everyday actions taken by persons, whether residents or non-residents, located in and across city borders.[156] In so doing, cities collaborate with private entities, mostly technology companies, in collecting, analyzing, and employing such data to promote their respective interests.[157] While the stated purpose of cities in collecting such data has been to improve their governance capabilities and to allow for data-driven city planning, infrastructure development, effective regulation, and so forth,[158] technology companies are evidently interested in using such data to increase their profits.[159] Besides the often-blurred lines between cities and corporations about which party gets to use the data and to what extent—an issue that increases concerns over the publicly unaccountable exploitation of pieces of personal information that comprise the data[160]—cities are now beginning also to explicitly explore the possibility of monetizing the data, including by its sale to third parties.[161] City officials, alongside other policy-makers and commentators in various countries, are

---

156. *See supra* notes 88-89, 102-03 and accompanying text.

157. *See* Astrid Voorwinden, *The Privatised City: Technology and Public-Private Partnerships in the Smart City*, 13 Law Innovation & Tech. 439, 440-41 (2021); *see also*, *e.g.*, Albert Meijer & Manuel Pedro Rodríguez Bolívar, G*overning the Smart City: A Review of the Literature on Smart Urban Governance*, 82 INT'L REV. ADMIN. SCIS. 392, 393 (2016) (discussing Amsterdam as an "urban living lab" for businesses).

158. *See* Meijer & Bolívar, *supra* note 157, at 393-94, 400-01.

159. *See* Voorwinden, *supra* note 157, at 440.

160. *Id.* at 449-59 (discussing the need to safeguard public values in smart cities in light of such public-private partnerships, particularly to ensure the protection of personal data gathered by private corporations).

161. *See, e.g.*, Freed, *supra* note 104; *see also* Kalev Leetaru, *When Will Cities Begin to Monetize Their Residents' Data?*, FORBES (July 19, 2018, 6:05 PM), [https://perma.cc/W7TH-72H2]; Kitty Kolding, *One Approach for Cities to Recoup Lost Revenues Due to COVID-19*, MEETING OF THE MINDS (May 5, 2020), [https://perma.cc/AUU8-BRH4].

voicing their support for such a move,[162] with workshops being offered to cities on how to monetize their data.[163]

The first systematic effort at creating a public-private market platform for the exchange of such data was undertaken by the city of Copenhagen, Denmark, in collaboration with the technology company Hitachi between 2015 and 2018.[164] The City Data Exchange (CDE) program was set up to test the readiness of such a potential marketplace for data.[165]

According to a 2018 report published by the city of Copenhagen, the most sought-after datasets by both public and private actors concerned information on how people move around the city, including their various locations and frequency of movement, with such data labeled "people movement patterns."[166] This data is collected from numerous sources, such as "cell phone tracking, wireless connection counting, camera image counting, traffic sensors, visual surveying, [and] ticket purchases."[167] While the city is interested in such data primarily for traffic planning, use of public spaces, or health and safety issues, public utility providers and private corporations seek to improve business efficiency and profitability.[168] Retailers use such data for locating stores or engaging in targeted marketing.[169] Transportation companies "request[] information on the number of people traveling between different geographical locations to understand their market share, but also to adjust their offerings."[170] Tourism organizations seek data on the "flow of

162. *See* Freed, *supra* note 104 (citing Erik Caldwell, deputy chief operating officer for the city of San Diego, California, saying that while "[i]t's the people's data" and "[i]t's on us to keep it that way," he is "very interested in monetizing the data"); Sajeesh Kumar N. & Bibek Debroy, *City Data Monetization Could Help Our Development Road Map*, MINT (Sept. 18, 2019, 11:50 PM), [https://perma.cc/4D59-AQWV].

163. *See, e.g.*, *Ensuring Cities Are Getting Full Value and Meaningful Revenue from Their Valuable Data*, CHRYSALIS PARTNERS, [https://perma.cc/KC2X-GT5X] (last visited Jan. 8, 2023) (advertising a workshop for municipalities on monetizing their data).

164. *See* Kumar & Debroy, *supra* note 162; SMART CITY INSIGHTS, CITY DATA EXCHANGE - LESSONS LEARNED FROM A PUBLIC/PRIVATE DATA COLLABORATION 2 (2018), [https://perma.cc/VY9J-PTN3].

165. *See* SMART CITY INSIGHTS, *supra* note 164, at 2.

166. *Id.* at 3 (emphasis omitted).

167. *See id.*

168. *See id.*

169. *Id.*

170. *See* SMART CITY INSIGHTS, *supra* note 164, at 4.

tourists from different countries to inform [their] industry on how to provide better services for tourists."[171]

At the same time, the CDE project also attested to the then-immaturity of the market, the lack of specific-use cases exemplifying how selling or buying of data benefited companies, the fragmented data landscape, and the reluctance by many organizations to share data on an open data portal—because of unclarity about data ethics or simply to prevent competitors from gaining access to such data.[172]

Current attempts at creating digital platforms for data exchange, which could in turn enable cities and private corporations to monetize the data they collect, demonstrate that while some of the market immaturity problems may have been overcome since then, it is essential to address the fundamental legal and regulatory questions that concern ownership or control of the data and, in particular, to consider the rights and interests of the sources of the different pieces of raw data—namely, both residents and non-residents that move in, around, and out of the city.

Consider the Open Mobility Foundation (OMF),[173] set up by the city of Los Angeles in 2019 as a forum where local officials and private companies can "collaborate on mobility data sharing," with the assumption that such information should be a public resource and would thereby allow cities to "apply uniform requirements and work together" on mobility regulations.[174] Alongside about fifty cities in the United States and across the world, the OMF includes a number of private mobility providers, such as e-scooter and bike-share companies, and other public and private entities.[175]

According to the framework set up by OMF, the Mobility Data Specification (MDS)—the digital tool that "standardizes

171. *Id.*

172. *Id.* at 6-7.

173. *See About the Open Mobility Foundation*, OPEN MOBILITY FOUND., [https://perma.cc/F8PX-VV6M] (last visited Jan. 8, 2023).

174. *See* Laura Bliss, *Scooter Rides Have Turned into a Data Privacy Issue for Cities*, BLOOMBERG (Nov. 10, 2021, 3:00 AM), [https://perma.cc/XL65-SN79].

175. *See About MDS*, OPEN MOBILITY FOUND., [https://perma.cc/LKK8-CJMU] (last visited Jan. 8, 2023); *The Open Mobility Foundation Overview*, OPEN MOBILITY FOUND., [https://perma.cc/ZC7F-ZELG] (last visited Jan. 27, 2023).

communication and data-sharing between cities and private mobility providers"—enables cities to "share and validate policy digitally" and also provides private mobility companies with a "framework they can re-use in new markets, allowing for seamless collaboration that saves time and money."[176] This means that beyond the exploitation of data that each of the private companies is constantly gathering on its own users, such companies also have access to city-generated data that companies can then use for profit-maximizing. Moreover, critics of the OMF argue that a recently established startup company, Lacuna Technologies, Inc., "helped finance the OMF and recruited other cities and companies to join the consortium" without publicly disclosing its commercial interests in the matter.[177] In response, civil society organizations, such as the American Civil Liberties Union, have initiated legal proceedings—unsuccessful so far—arguing that the gathering and sharing of such data invades the privacy of users and that the Los Angeles Department of Transportation (LADOT) "has never articulated an adequate or reasonable justification for the collection of such sensitive location information en masse."[178]

Therefore, it seems that while the marketplace for the exchange and potential monetization of smart-city data is increasingly maturing—in the sense that both cities and private companies realize the potential of exchanging and trading in data, and that accordingly cities are viewing aggregated data as a form of "new property" that may go beyond a tool for better governance—it becomes essential to consider normatively and doctrinally the proprietary status of such data. In particular, such an analysis should inquire if cities are entitled to exploit such data not only as a tool for better governance and provision of services but also as a tradable asset—and if so, under what terms.

While a comprehensive analysis of the proprietary features of data is outside the scope of this Article,[179] it may be useful to

---

176. *See id.*

177. Bliss, *supra* note 174.

178. *Id.*

179. For a general analysis of data/information and property rights, see, for example, Andreas Boerding et al., *Data Ownership—A Property Rights Approach from a European Perspective*, 11 J. CIV. L. STUD. 323, 325-26 (2018); FAIRFIELD, *supra* note 97, at 2-12;

distinguish at the outset between the *conceptual* analysis of data as an intangible resource that is capable of being a potential object of property rights—identified and notionally carved out from the rest of the world and holding economic value—and a *normative* analysis about who should be entitled to control and profit from a certain piece or cluster of data.

Conceptually, the principles of property law generally require that the object would be clearly described and delineated and that the asset's outer limits—whether physical or conceptual—would be known publicly so that unauthorized parties would not encroach on it.[180] While the form of publicity may vary, distant parties should have a practical way of identifying the object and rights thereto so as to "[k]eep off."[181] Data poses particular conceptual challenges in the attempt to distinguish between general information about the world and a particular compilation of information that has been generated through a deliberate process and has a distinct value as such.

As Sjef van Erp shows in the context of COVID-19 tracing and tracking electronic apps, one can distinguish between a specific piece of data extracted about a person and the aggregation and cross-data analysis of data that is then undertaken by health authorities and corporations.[182] As shown below, this conceptual multi-layering of data may also have normative significance in that it illuminates the different production/compilation roles and interests of various stakeholders.[183] In the case of health-related data, this may relate to persons/patients, healthcare providers, public-health authorities, pharmaceutical companies, producers of medical hardware/software, etc.[184]

Lyria Bennett Moses, *Who Owns Information? Law Enforcement Information Sharing as a Case Study in Conceptual Confusion*, 43 UNIV. N.S.W. L.J. 615, 615-19 (2020); Sjef van Erp, *Ownership of Data: The Numerus Clausus of Legal Objects*, 6 BRIGHAM-KANNER PROP. RTS. CONF. J. 235, 235-36.

180. *See* THOMAS W. MERRILL & HENRY E. SMITH, THE OXFORD INTRODUCTIONS TO U.S. LAW: PROPERTY 4-11 (Dennis Patterson ed., 2010).

181. Felix. S. Cohen, *Dialogue on Private Property*, 9 RUTGERS L. REV. 357, 373-74 (1954).

182. Sjef van Erp, *Who "Owns" the Data in a Coronavirus Tracing (and/or Tracking) App?*, *in* CORONAVIRUS AND THE LAW IN EUROPE 131, 152-56 (Ewoud Hondius et al. eds., 2021) [hereinafter van Erp, *Data*].

183. *See* discussion *infra* notes 186-90.

184. *See* van Erp, *Data*, *supra* note 182, at 139.

In the context of smart cities, the process of complication, aggregation, and cross-cutting analysis of different pieces of data coming from multiple sources in and around the city is done primarily through a collaboration between the municipality and technology companies. Moreover, the process of *anonymizing* the data—to the extent that such a step can be reliably done, guaranteed, and monitored over time to protect the privacy interests of persons while meeting the requirements of legal norms such as the European Union's General Data Protection Regulation (GDPR)[185]—may add a further dimension to the potential conceptual distinction between the indefinite pieces of personalized data and between aggregated *and* anonymized data on people's movement patterns.

This conceptual analysis ties into the normative consideration of which party or parties should be entitled to collect, control, use, or profit from such data—and under what conditions. The potential distinction between a piece of information that is derived from a person and identifies him or her as the source of information and between an aggregated and anonymized cluster of data that has value in identifying collective patterns of behavior, such as movement patterns, may also have normative implications. The unbundling of a raw piece of data from a mass cluster can also allow for an unbundling of a single definition of "ownership" in data and parts thereof.[186] The more appropriate way to set out the entitlements to certain aspects of control, access, use, and profit-making in data could rather be one of relative title and use-specific entitlements, which involves considering the role of different stakeholders in collecting the data and their potential interests.

Thus, while a person's interest in a personalized piece of information can be validated not only in terms of the right of privacy but also in recognizing his or her proprietary entitlement

185. *See generally* Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119).

186. *See* van Erp, *Data*, *supra* note 182, at 152-55.

to it,[187] the normative justification for allowing cities to aggregate data—while anonymizing it—even if explicit consent by each one of the sources of the raw data cannot typically be assumed, may lie in the anticommons theory of property.[188] Under this theory, the over-fragmentation of private property rights can lead to inefficient and unjust results, from deadlocks among adjacent landowners about restructuring rights to allow for effective redevelopment of the entire area to the undersupply of biomedical innovation due to exclusive patents over fragments of knowledge.[189] "Too much private property" could be detrimental to asset governance and collective action.[190]

While the result of an anticommons dynamic should obviously not lead to a sweeping abolishment of private property rights or their preemption by the government or private parties that can allegedly generate more value from taking over all fragmented rights, this analysis could serve as a benchmark for reconsidering the concept of relative title or the unbundling of property rights. Accordingly, in considering which parties should be entitled to collect, integrate, and utilize clusters of a certain type of tangible or intangible resource, the analysis should look not only to the party that would be most effective in governing and utilizing such aggregation of assets from a social-welfare perspective but also to the one that could be held accountable for potential misuses.[191]

---

187. FAIRFIELD, *supra* note 97, at 99-118 (arguing that recognizing a person's property rights in information can help to "stop government snooping").

188. *See* Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698-99 (1998).

189. *See id.* at 699-700.

190. MICHAEL HELLER, THE GRIDLOCK ECONOMY: HOW TOO MUCH OWNERSHIP WRECKS MARKETS, STOPS INNOVATION, AND COSTS LIVES 2 (2008). This view has not remained uncontested. *See, e.g.*, Richard A. Epstein, *Heller's Gridlock Economy in Perspective: Why There Is Too Little, Not Too Much Private Property*, 53 ARIZ. L. REV. 51, 51-53 (2011); *see also* Jorge L. Contreras, *The Anticommons at 20: Concerns for Research Continue*, 361 SCI. 335, 337 (2018) (showing how certain information-sharing industry practices can alleviate the anticommons problem).

191. For an analysis of various anticommons dynamics and the types of proprietary settings that can address such problems (including by moving away from regular ownership to a model of trust governance), see Amnon Lehavi, *The Law of Trusts and Collective Action: A New Approach to Property Deadlocks*, 89 CIN. L. REV. 388, 390-91 (2021); *see also* LEE ANNE FENNELL, SLICES AND LUMPS: DIVISION AND AGGREGATION IN LAW AND LIFE 15, 22-26 (2019) (discussing the challenge of trying to secure cooperation to allow for either an aggregation or division of resources).

Therefore, in the context of digital data about activities taking place in and around a city—data that is gathered from an indefinite number of data sources (such as persons moving around the city) and numerous types of devices (e.g., GPS location, sensors, cameras, automatic meter reading)—the city itself can be deemed the party that should have a relative title or claim to aggregated data. The city can benefit from the aggregation and cross-cutting analysis of data in promoting its normatively legitimate goals as a local government—a point I address in detail in Section III.A[192]—while at the same time being held accountable for potential misuses of such aggregated data. Such accountability should be based not only on public law duties but also on defining the in rem duties that the city has toward different types of parties, including persons that are the source of raw data.

In other words, while recognizing certain in rem rights that a city may have in collecting, aggregating, analyzing, and utilizing smart-city data, it should also be subject to in rem duties in its proprietary capacity—ones that may obligate it, for example, to prevent the re-identification of anonymized data, including by third parties, such as technology corporations, transportation providers, retailers, or other cities with which the city shares data or to which such data is sold. This means, for example, that if a retail company that receives aggregated and anonymized data from the city engages in "reverse engineering" and re-identifies persons who are the sources of the raw data in order to commercially "target" them, then such an act can also be considered to be a violation by the city itself of its in rem duty to prevent violation of privacy in exploiting the data.

Accordingly, the scope of such proprietary in rem duties and of correlative respective rights—such as the right of a person not to be identified as the source of a piece of data when such identification would violate his or her privacy—should not depend on whether the potential right-bearer is a resident of the city. Namely, unlike the scope of political governmental authority of a city vis-à-vis persons—which may change in the case of residents, non-local nationals, or foreigners and in many cases may remain in a gray area because of the "governance gap"

192. *See infra* notes 197, 205-08 and accompanying text.

I identified earlier[193]—in the case of proprietary in rem duties, what matters is whether a person—*any person*—has been the source of a piece of data that was later aggregated or utilized by the city. In Section III.B, I discuss in detail what such an array of in rem rights and in rem duties may look like.[194]

III. RESCALING CITY PROPERTY IN A CROSS-BORDER REALITY

A. Why Cities Matter for Intra-Local and Inter-Local Openness

As Section II.A has shown, in the context of the development of the public/private distinction in U.S. law and similar conceptions developed in other legal systems around the world, local governments—including cities—are viewed as public or governmental corporations.[195] What this means, among other things, is that unlike private corporations that are generally entitled to engage in any kind of business and are at liberty to acquire any type of asset as they deem fit, subject to few constraints,[196] cities act as governmental entities within the authority granted to them under a certain legal system. Consequently, the purposes for which cities may acquire, use, and monetize assets should be in line with the broader purposes that local governments should promote.

Thus, beyond the need to abide by legal rules that govern the allocation of power between upper-level governments and local ones and general doctrines that apply to all types of public entities in the relevant legal system, a city should be subject more fundamentally to a normative evaluation of the goals it seeks to promote, even when it acts in a "proprietary" capacity—that is, for example, when it decides to sell or rent an asset that it owns to a private party or when it competes in the market for the

193. *See supra* text accompanying notes 15-16.
194. *See infra* notes 240-45 and accompanying text.
195. *See supra* notes 113-15 and accompanying text.
196. *See* Hovenkamp, *supra* note 127, at 1044 (explaining that U.S. antitrust law was created as a tool to restrict anticompetitive behavior in the modern capitalist economy and therefore justifies placing certain limits on businesses' rights to buy or sell property).

provision of services or products alongside private corporations.[197]

Accordingly, the normative analysis of legitimate city action should apply to all stages and aspects of its proprietary rights and duties. This includes the fundamental questions about which types of assets a city should be entitled to own or control—including, for that matter, aggregated data that it collects about patterns of movement of residents and non-residents across the city—or under which circumstances it may exclude certain persons from accessing or using city-owned assets. In other words, allowing cities to exert their property rights in assets in a differential and creative manner to account for their current cross-border reality—which is the key thesis being promoted in this Article—should be subject to identifying the underlying normative criteria that cities should meet and how such criteria affect the correlative rights and duties of diverse categories of persons.

This Section does not explore in detail the various arguments that have been made in the academic literature and public discourse over the past few decades about the ideologies that should guide cities, such as the idea of the "neoliberal city" that gained traction in the United States and elsewhere during the 1980s and 1990s[198] or competing models of welfare-oriented cities.[199] It seeks, rather, to focus on certain values that are particularly important in the context of cross-border urbanism—and that consequently have a bearing on the way in which cities should control and use their tangible and intangible assets—and how this should translate into in rem rights and in rem duties. Such values deal with intra-local and inter-local openness and, with it, diversity and tolerance.

---

197. *See, e.g.*, Schanzenbach & Shoked, *supra* note 129, at 594-95 (reasoning that the public trust doctrine in U.S. law "restricts cities' freedom in transacting with assets because the public is held to be the assets' beneficial owner").

198. *See* JASON HACKWORTH, THE NEOLIBERAL CITY: GOVERNANCE, IDEOLOGY, AND DEVELOPMENT IN AMERICAN URBANISM, at xi-xii, 9-13 (2007) (critically examining the evolution of this idea and its practice in various American cities).

199. *See* SCHRAGGER, *supra* note 9, at 247-55 (arguing that cities should be given more power to improve the health and welfare of their citizens and to ameliorate inequality, such as by imposing a municipal minimum wage).

What this essentially means is that in designing the proprietary rights and duties of cities in their assets—and particularly their right to exclude or otherwise control access to such assets or the right to collect and aggregate data about persons' movements while potentially monetizing it—cities should be generally committed to promoting openness, diversity, and tolerance, not only toward their residents-voters but also toward various crowds of workers-commuters, visitors, etc. Thus, while cities should be generally entitled, as asset owners, to require persons who use city assets to internalize the marginal costs they generate—including in cases where the city's political or fiscal power of governance is not clearly articulated to address various categories of users—they should do so in a manner that seeks to facilitate the role of cities as a hub of cross-border activities.

The view of cities as places of intra-local and inter-local openness, diversity, and tolerance seeks not only to recognize and validate the economic, organizational, and technological features of contemporary cities and their instrumental value in facilitating cross-border markets but also to promote other values. Gerald Frug sees the key role of cities as one that seeks "to increase the capacity of metropolitan residents to live in a world composed of people different from themselves."[200] Referring to the works of Iris Young,[201] Richard Sennett,[202] Jane Jacobs,[203] and others, Frug articulates the challenges but also the benefits of urban openness and heterogeneity. These benefits include the psychological contribution of a more open way of life for human development and growth, the social goal of overcoming interpersonal suspicion and fear, and the potential for alleviating the political divisions that characterize many metropolitan areas.[204]

Therefore, intra-local and inter-local openness, diversity, and tolerance—which generally distinguish cities from many

---

200. FRUG, *supra* note 33, at 115.

201. *Id.* at 11, 137 (citing IRIS YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (1990)).

202. *Id.* at 116, 119-22, 138-40 (citing RICHARD SENNETT, THE USES OF DISORDER: PERSONAL IDENTITY AND CITY LIFE (1970)).

203. *Id.* at 123-24, 141 (citing JANE JACOBS, THE DEATH AND LIFE OF GREAT AMERICAN CITIES (1961)).

204. *Id.* at 137-142.

suburbs, and even more so from "private communities," such as residential community associations[205]—are not only instrumental in allowing cities to work as cross-border economic hubs but are also self-standing normative virtues that should be promoted. According to Frug, as well as other authors, such as Richard Schragger, this calls for reconceiving and *strengthening* city power, rather than centralizing power in a regional or national government.[206]

In the context of the cross-border reality that is the focus of this Article, while a major political realignment of governance power between cities and national governments, or on the international scene, seems unlikely—thus leaving intact the urban governance gap detailed above—reconceiving cities' in rem rights and in rem duties in assets may allow them to utilize such tangible and intangible assets to promote the instrumental and deontological goals of cross-border openness. Such in rem rights and in rem duties could cover the actual geographical and digital scales within which cities operate and would accordingly apply to various groups of residents and non-residents.

This property re-scaling may prove effective not only on the domestic front but also internationally. Thus, while cities have no formal standing in international law,[207] cities should generally be allowed to trade in assets with cities in other countries, especially in the case of intangible assets, such as aggregated and anonymized data, in order to promote goals such as fostering economic, cultural, and technological collaboration, or addressing climate change—thereby allowing for efficient utilization of resources while promoting cross-border openness. The following Section sets out to consider how such in rem rights and in rem duties should be designed.[208]

---

205. For the characteristics of many suburbs and "private communities" as places aimed at homogeneity and exclusion, see, respectively, FISCHEL, *supra* note 70, at 202-06, 215-19; EVAN MCKENZIE, PRIVATOPIA: HOMEOWNER ASSOCIATIONS AND THE RISE OF RESIDENTIAL PRIVATE GOVERNMENT 56-78 (1994).

206. FRUG, *supra* note 33, at 118; SCHRAGGER, *supra* note 9, at 253-55.

207. Barbara Oomen & Moritz Baumgärtel, *Frontier Cities: The Rise of Local Authorities as an Opportunity for International Human Rights Law*, 29 EUR. J. INT'L L. 607, 621 (2018).

208. *See infra* Section III.B.

B. Redefining City Property: In Rem Rights, In Rem Duties

In light of the urban governance gap identified in this Article,[209] there is a pertinent need to establish a comprehensive legal framework that governs the proprietary rights and duties of cities in tangible and intangible assets—and the correlative rights and duties of diverse groups of persons who use such assets or are otherwise affected by them, whether they are residents or non-residents.

How should such in rem rights and in rem duties concerning city assets be designed? How would these diverge, if at all, from the general contours of rules on private property designed in different legal systems, on the one hand, and from national-government ownership, on the other? Such a potential blueprint requires us to briefly consider the in rem trait of proprietary rights and duties and how this feature establishes certain structural components in the design and enforcement of norms that govern the access, control, and use of an asset that would apply toward large, often indefinite numbers of heterogeneous norm-subjects—but without essentially dictating a single pre-fixed content of property norms for all types of resources and/or all types of owners.

In his seminal work, Wesley Hohfeld set out to challenge the traditional dichotomy, dating back to Roman law, between in rem (Latin: "against a thing") and in personam ("against a person").[210] Defining and analyzing the different attributes of in personam rights through a delineation of jural opposites and jural correlatives that govern legal relationships among persons (such that, for example, one person's right is correlated with a duty imposed on another person), Hohfeld argued that the same typology of jural opposites and jural correlatives applies to in rem rights—save for the large, indefinite number of persons who are bound by these interpersonal legal relationships.[211]

---

209. *See supra* Part I.

210. *See* Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 26 YALE L.J. 710, 720-29 (1917).

211. *See* Wesley Newcomb Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 YALE L.J. 16, 28-32 (1913); Hohfeld, *supra* note 210, at 710-17.

Whereas Hohfeld's enterprise was largely analytical-conceptual—one addressing the legal structure of property—the "subsequently developed metaphor of the 'bundle of rights'" served also a normative purpose, especially by twentieth-century legal realists and critical legal theorists, who sought to de-canonize the institution of property on an ideological basis.[212] Under this account, property is not a natural right but is rather a legal institution that is a creature of the State and could thus be designed to promote broad social goals beyond preserving the owner's dominion.[213]

The counter-movement to this line of argument, which I elsewhere dubbed "new essentialism," emerged in the late twentieth century and early twenty-first century.[214] According to James Penner, the fundamental interest protected by property law is grounded in the use of objects—thus returning somewhat to the original meaning of the term in rem—and characterized by the exclusion of others, meaning that property norms protect the interest of use through exclusion.[215] Under his characterization of norms in rem, the owner's in rem rights to property use are individuated and framed in terms of a duty in rem to exclude oneself from the property of others—meaning that the duty not to interfere with the assets of others applies universally to all non-owners.[216]

In a series of influential works, Thomas Merrill and Henry Smith argue that in rem rights are qualitatively different from in personam rights even if the property/contract borders are not always clear, reasoning that different legal systems continue to embrace a numerus clausus principle of property forms, that property rights retain at least a basic layer of a universal right of exclusion in favor of the owner good against the world, and that these distinctive traits of property law can be justified as socially

---

212. Amnon Lehavi, *The Property Puzzle*, 96 GEO. L.J. 1987, 2001 (2008).

213. For an analysis of this literature, see AMNON LEHAVI, THE CONSTRUCTION OF PROPERTY: NORMS, INSTITUTIONS, CHALLENGES 26-30 (2013) [hereinafter LEHAVI, CONSTRUCTION].

214. *Id.* at 46-49.

215. J. E. PENNER, THE IDEA OF PROPERTY IN LAW 68-74 (1997).

216. *Id.* at 128-152; *see also* Eric R. Claeys, *Property, Concepts, and Functions*, 60 B.C. L. REV. 1, 51 (2019) (arguing that when an in rem right is institutionalized, "it also establishes correlative *in rem* duties and disabilities on non-proprietors").

efficient in view of systemic information and enforcement costs.[217]

While I dispute Merrill and Smith's arguments about the essential *substantive* content of property, and especially the contention that the right to exclude is the inherent core of ownership regardless of the nature of the asset or the type of owner[218]—and I explain the implications of my viewpoint for city-owned assets below[219]—I share the notion that as a *structural* matter, property differs from other types of obligations and that property law is "shaped largely to reduce the informational burdens of the owners and non-owners who have to cope with the system."[220]

This is particularly so because of the way in which legal rights and duties regarding both specific assets and more generally categories of resources (such as land, chattels, financial instruments, intellectual property, or data) regularly implicate numerous parties with diverging features and interests. Beyond the fact that such parties are usually not identifiable to one another in advance—unlike parties that have a privity of contract among them—they often turn out to be more heterogeneous in their epistemological, cultural, and social attributes, as compared with typical contractual counterparts. What this generally means is that for property to function well in creating, allocating, and enforcing in rem rights and duties, it must facilitate broad-based understanding about the way in which property legal interests are structured and defined.[221]

217. *See* Thomas W. Merrill & Henry E. Smith, *The Property/Contract Interface*, 101 COLUM. L. REV. 773, 774-78, 792-99 (2001) [hereinafter Merrill & Smith, *The Property/Contract Interface*]; Thomas W. Merrill & Henry E. Smith, *Optimal Standardization in the Law of Property: The* Numerus Clausus *Principle*, 110 YALE L.J. 1, 3-9 (2001); Thomas W. Merrill & Henry E. Smith, *The Morality of Property*, 48 WM. & MARY L. REV. 1849, 1855 (2007) [hereinafter Merrill & Smith, *The Morality of Property*].

218. *See* Merrill & Smith, *The Morality of Property*, *supra* note 217, at 1858; Thomas W. Merrill, *Property and the Right To Exclude*, 77 NEB. L. REV. 730, 730-31 (1998); Henry E. Smith, *Exclusion Versus Governance: Two Strategies for Delineating Property Rights*, 31 J. LEGAL STUD. S453, S456 (2002).

219. *See infra* notes 221-30 and accompanying text.

220. *See* Robert C. Ellickson, *Two Cheers for the Bundle-of-Sticks Metaphor, Three Cheers for Merrill and Smith*, 8 ECON J. WATCH 215, 218-19 (2011) (analyzing Merrill and Smith's work—while offering certain "[f]riendly" critiques).

221. *See* LEHAVI, CONSTRUCTION, *supra* note 213, at 39-41 (discussing the third-party applicability of property rights).

That said, in rem rights and duties need not have a single content regardless of the type of asset that is the object of property rules or the identity of owners and different categories of non-owners. As long as sufficient clarity, transparency, and publicity can be achieved in articulating certain variations across the system of property for different types of assets and/or different types of owners—including private versus public owners—property law can maintain its essential structural traits without succumbing to a single substantive blueprint that cannot be normatively defended.

Thus, for example, the fair use doctrine in copyright law substantially limits the copyright owner's right to exclude others—in a manner that does not exist for other types of intellectual property such as patents or other types of assets such as land.[222] This substantive variation derives from normative considerations that address the balance between incentivizing creators of certain content, such as musical, literary, or architectural works, to innovate and enrich the world, and the public interest in allowing non-owners to use such content at a certain scope and for certain purposes such as education, thus also distinguishing between different categories of non-owners.[223]

Accordingly, the identity of the owner can also lead legal systems to design certain varieties in property law doctrines. In particular, the rules for public ownership may somewhat diverge from those for private ownership without undermining property law's general structural features. For example, the common law "public trust doctrine," which originated in England and developed in the United States during the nineteenth century—by which the government holds certain types of assets in trust for the public—may place certain limits on the public owner but also certain stronger entitlements vis-à-vis non-owners, as compared with private owners of the same kind of asset.[224]

---

222. *See, e.g.*, Neil Weinstock Netanel, *Making Sense of Fair Use*, 15 LEWIS & CLARK L. REV. 715, 719-20 (2011) (looking at the doctrine's recent historical development and arguing that it offers significant consistency and determinacy).

223. *See id.* at 719-20, 724.

224. *See generally* MOLLY SELVIN, THIS TENDER AND DELICATE BUSINESS: THE PUBLIC TRUST DOCTRINE IN AMERICAN LAW AND ECONOMIC POLICY, 1789-1920 (1987). It should be noted that while the public trust doctrine could also apply, at least in some U.S. jurisdictions, to privately-owned assets such as beachfront land, its scope is more limited

Thus, the public trust doctrine limited the ability of state governments, and even more so of local governments, to alienate or otherwise restrict public access to navigable waters and the land submerged under them, and later also to highways and streets, based on the view that the government held such assets merely as its residents' agent.[225] While this doctrine has somewhat changed course during the twentieth century, it still persists in Anglo-American law[226] and has even been reinvigorated over the past few decades, largely due to the work of Joseph Sax,[227] who called to extend the scope of the doctrine and the limits it places on government control over natural resources.[228]

At the same time, the rationale by which government holds certain types of assets in trust for its residents has also worked to give it increased protection versus certain types of non-owners. Under common law rules originating in England, adverse possession of land does not run against the government—local, state, or federal.[229] Although this immunity against adverse possession has been somewhat downscaled by some U.S. states, most courts adhere to this rule, reasoning that because government holds the land in trust for all people, the latter should not lose the land because of the negligence of government officials—thus distinguishing public ownership from a private one.[230]

Another type of distinction that may be relevant for different types of assets, or different types of owners, concerns the scope of in rem duties. Under the essentialist approach to property discussed above, in rem duties apply chiefly to non-owners, requiring them to exclude themselves from the property of

than for publicly owned assets. *See* JESSE DUKEMINIER & JAMES E. KRIER, PROPERTY 816-23 (5th ed. 2002).

225. Schanzenbach & Shoked, *supra* note 129, at 585-87.

226. *Id.* at 594-95.

227. *See* Joseph L. Sax, *The Public Trust Doctrine in Natural Resource Law: Effective Judicial Intervention*, 68 MICH. L. REV. 471, 474 (1970).

228. *See* Michael C. Blumm & Zachary A. Schwartz, *The Public Trust Doctrine Fifty Years After Sax and Some Thoughts on Its Future*, 44 PUB. LAND & RES. L. REV. 1, 1-4 (2021).

229. DUKEMINIER & KRIER, *supra* note 224, at 102.

230. *Id.* at 103.

others.[231] What about in rem duties that apply to owners? According to Merrill and Smith, in rem duties of property owners, such as the duty of a landowner to refrain from carrying out a nuisance activity, are very limited in scope and always "negative."[232] However, as Robert Ellickson notes, this approach is oversimplified.[233] As a matter of current doctrine, "[t]he law may affirmatively require a landowner, for example, to control natural vegetation or to contribute to the costs that an abutting neighbor has incurred to fence a common boundary."[234] Ellickson's approach is based mostly on the information costs theory advocated by Merrill and Smith by suggesting that many of the "affirmative duties of owners are similarly based on the likelihood that they have special knowledge" about a certain asset and its potential impact on other parties.[235]

More importantly, such a narrow approach ignores a growing number of normative arguments made by authors such as David Lametti,[236] Gregory Alexander,[237] and Joseph Singer,[238] by which property owners also owe certain affirmative duties toward other members of society. Even if a legal system does not embrace a comprehensive set of affirmative duties across property law, it may be justified to do so for particular types of

---

231. *See supra* text accompanying notes 214-16.

232. Merrill & Smith, *The Property/Contract Interface*, *supra* note 217, at 788-89.

233. Ellickson, *supra* note 220, at 220.

234. *Id.* (citations omitted).

235. *See* Robert C. Ellickson, *The Affirmative Duties of Property Owners: An Essay for Tom Merrill*, 3 BRIGHAM-KANNER PROP. RTS. CONF. J. 43, 52 (2014). Thus, when a municipality requires landowners to shovel snow from abutting sidewalks, it may be justified to do so because "[t]he owner of a lot knows best, for example, which of its shrubs can best withstand a pile of deposited snow" and also because the owner, who frequently uses the abutting sidewalk, is more incentivized to perform the task well. *Id*. at 56.

236. *See* David Lametti, *The (Virtue) Ethics of Private Property: A Framework and Implications*, *in* NEW PERSPECTIVES ON PROPERTY LAW, OBLIGATIONS AND RESTITUTION 39, 66 (Alastair Hudson ed., 2004) (discussing the ethical aspects of property law and arguing that "private property will have to be seen to be as much about duties and goals as it is about rights").

237. *See* GREGORY S. ALEXANDER & EDUARDO M. PEÑALVER, AN INTRODUCTION TO PROPERTY THEORY 80-97 (2012) (developing a social-obligation norm in property law, one that seeks to promote "human flourishing" that would enable all individuals to live lives worthy of human dignity, and analyzing the corresponding obligations on asset owners that should facilitate this).

238. *See* Joseph William Singer, *Democratic Estates: Property Law in a Free and Democratic Society*, 94 *CORNELL L. REV.* 1009, 1048 (2009) (arguing that property owners owe a duty of "attentiveness" toward other members of society).

assets or categories of owners. For example, patent law in the various legal systems requires patent applicants, as a condition for obtaining a patent, to disclose the content of the invention such that non-owners, including for that matter scientific or professional competitors, would be able to use it once the patent expires—meaning that the (temporal) in rem right of exclusive use is balanced by imposing an in rem duty that seeks to serve not only the interests of other innovators/manufacturers but also those of the public at large.[239] Such an affirmative duty placed on the owner of the patent derives, therefore, from the particular normative balance regarding the creation, access, and use of this type of asset—and it can be designed in a manner that does not undermine the structural traits of proprietary rights and duties.

Accordingly, affirmative in rem duties can be designed for certain categories of owners, while also considering that different categories of non-owners may have a particular normative claim for a correlative right vis-à-vis such owners. In the context of city-owned assets in an age of cross-border urbanism, such in rem duties may be owed not only to residents but also to affected non-residents.[240] This should be the case with data aggregated by the city about movement patterns or other pieces of information that are derived from persons who move in and around the city, whether they are residents or non-residents. Recognizing a principled proprietary right for a city to assemble, use, and monetize such data should be conditioned on establishing affirmative in rem duties on it. This means that beyond a general in rem duty to make use of such data to promote the general normative goals for which cities should operate, the city should have a particular in rem duty toward all actual and potential subjects of raw data to make sure that the aggregated data remains anonymized and is not being "reverse-engineered" against such persons.[241]

---

239. *See* Lisa Larrimore Ouellette, *Do Patents Disclose Useful Information?*, 25 HARV. J.L. & TECH. 545, 546-50, 554 (2012) (offering an empirical analysis of the scope of use in such revealed information and calling to expand the duty).

240. This approach also shows the advantage that the in rem rights/duties conceptual framework has over a "fiduciary" model that is in principle restricted to the local government's residents. *See* Schanzenbach & Shoked, *supra* note 129, at 585-86.

241. *See supra* notes 95-103, 192-93 and accompanying text.

Therefore, establishing a system of in rem rights and in rem duties that would apply to assets owned or controlled by cities in an age of cross-border urbanism need not undermine the basic features of prevailing property law concerning different types of assets but should be finetuned to account both for the specific features of cities, as compared with private owners on the one hand and national governments on the other, and for assets that exhibit particular traits when operating in the urban context—such as the type of data collected, used, and monetized in smart cities.[242]

What this means, on the side of in rem rights, is that cities should be entitled to exert their property rights in assets to require different categories of users to internalize the marginal costs and benefits they place on such assets. To the extent that the mode of governance of a certain asset is not governed by statutory law or regulatory provisions deriving from an upper-level government, the city should be generally at liberty to introduce cost-internalization mechanisms such as quota-setting, sorting, and pricing, as long as such criteria are transparent and consistent. Therefore, to the extent that such mechanisms differentiate between categories of users, including for that matter between residents and non-residents, the city should be able to justify such distinctions by demonstrating how each such category of users contributes differently to the acquisition and maintenance of assets (including by payment of local taxes) and/or how such category of users imposes a different scope or type of costs on the asset. In managing an asset that it owns or controls,[243] a city should be granted the legal power to do so in a differential and creative manner.

At the same time, the city should also be subject to a general, affirmative in rem duty that has special merit in an age of cross-border urbanism. The city should own, manage, and utilize its assets in a way that promotes both intra-local and inter-local openness, tolerance, and diversity, based on the normative

242. *See supra* Sections I.C, II.C.

243. For the argument that property rights are a management tool, meaning that property law should facilitate effective management, see Lynda L. Butler, *Property as a Management Institution*, 82 BROOK. L. REV. 1215, 1222 (2017).

analysis outlined in Section III.A.[244] What this means, in more concrete terms, is that cities should generally abstain from engaging in outright exclusion of non-residents from city assets, or from other rules that have the practical effect of keeping out underprivileged locals or the general category of non-residents, if such rules cannot be grounded in cost-internalization (notwithstanding the self-standing normative merit in subsidizing or promoting vertical equity). Cities should make sure that they remain generally open and inclusive in both theory and practice.

The same type of balance between in rem rights and in rem duties should also apply to the ability of the city to create, acquire, or assemble types of assets that are derived, at least to some extent, from its governance power or actual control over the city's physical and digital spaces. This is particularly so in the case of data that is aggregated, among other things, from surveillance cameras, sensors, automatic meter reading, and other tools placed or operating in its territory.[245]

While the anticommons analysis presented in Section II.C above[246] can justify the city's right to transform indefinite pieces of raw (and personal) data into an integrative asset of aggregated and anonymized data—one that may aid it in its governance capacities but that could also be monetized—such an in rem right must be accompanied by significant in rem duties. Beyond the general duty of the city to use such an asset in line with the general purposes that it should promote, including for that matter intra-local and inter-local openness, it also owes an affirmative in rem duty to all past, present, and future persons that are the sources of the raw data.[247] Such a duty should hold the city liable against any misuse of the aggregated data—whether it is done by the city itself or by third parties with which the city is transacting in collecting, processing, or monetizing the data.[248] Thus, an infringement of privacy concerning the raw and personal piece of data should be viewed as infringing also an in rem right that the subject of data has vis-à-vis the city that collected it.

244. *See supra* notes 197, 205-08 and accompanying text.
245. *See supra* notes 88-99 and accompanying text.
246. *See supra* notes 188-91 and accompanying text.
247. *See supra* notes 197, 205-06, 239-41 and accompanying text.
248. *See supra* notes 196-200 and accompanying text.

CONCLUSION

Cities today face a growing number of challenges but also prospects for change that could help them to move forward and enable them to function effectively and innovatively as hubs for economic, technological, and interpersonal exchange in the twenty-first century. Going beyond a previously limited group of "global cities" or "international cities," many cities in developed, emerging, and developing economies must address the effects of the constant movement of goods, capital, services, and persons across national and international borders, alongside the dramatic effects of digital technology and other innovations that defy both geographical borders and any attempt at neat divisions between local, national, and international matters. However, political institutions and arrangements do not follow up on such developments so quickly, if at all. The result is one of an "urban governance gap," a term that refers to the growing disparity between the traditional mandate of cities in national legal systems and the current reality of cities, which requires them to address not only thematic issues that have been traditionally left to other levels of government but also a growing scope of everyday activities that involve non-residents.[249]

This Article argued that many aspects of the governance mismatch embedded in the cross-border nature of urban life can be addressed through a reconsideration of the role of cities' property rights in assets, both tangible and intangible ones.[250] A reconfiguration of the in rem rights and in rem duties that cities should have regarding assets they own or control would not simply take us back to the pre-modern era of cities as "associations" or chartered enclaves.[251] Rather, articulating the broad-based normative goals that cities should promote is not only instrumental in assessing how they perform in the governance capacities they possess under the relevant legal systems, but it could also be of key importance in identifying proprietary rights and duties in regard to city assets so as to make

249. *See supra* Part I.

250. *See supra* Part II.

251. *See supra* notes 20, 33-41 and accompanying text.

sure that the urban governance gap will not leave cities unable to act effectively.[252] This is the case in regard to physical assets, such as local infrastructure used by both residents and non-residents, as well as digital assets—such as the aggregated data that cities are increasingly collecting about everyday action patterns of both residents and non-residents.[253]

The normative case for allowing cities to exert their property rights in assets in a differential and creative manner to effectively address their cross-border reality is particularly strong when cities remain otherwise committed to promoting intra-local and inter-local openness, diversity, and tolerance. The multiple dimensions of cross-border activities taking place in cities today attest to their key role as a forum for inter-local and international mobility, exchange, and heterogeneity[254]—meaning that any normative or doctrinal reconsideration of the array of in rem rights and in rem duties that cities should have in regard to assets must preserve these essential traits of urban life.

252. *See supra* Part III.
253. *See supra* Section II.B., C.
254. *See supra* Sections I.A, III.A.

RACIAL DIVERSITY AND LAW FIRM ECONOMICS

Jack Thorlin*

INTRODUCTION

There is an eternal temptation to think that if one recognizes a moral problem and does something about it, then one is blameless even if the action taken does not solve the problem. We usually recognize that it is absurd to credit intent when the disconnect from results is vast—consider the rightfully mocked tendency of people to respond to tragedies by declaring that their "thoughts and prayers" are with the victims rather than taking any meaningful step to ameliorate their suffering.[1] People still engage in such posturing because the behavior benefits them in several ways: (a) others see that the actor is doing something and think the actor is moral, and (b) the actor can assuage her own conscience by having done *something*.[2] But declared good intent divorced from positive consequences is no virtue, and if it goes on long enough with the full knowledge of the actor, it turns into vice.

Racial diversity in law firms is a classic "thoughts and prayers" situation. Legal practitioners are consistently among the most socially progressive professions in the country, a tendency even more pronounced among the top-tier law school graduates who populate law firms.[3] Attorneys frequently cite the

* Adjunct Professor of Law, Georgetown University Law Center.

1. *See* Ben Rowen, *What Science Says About 'Thoughts and Prayers*,*'* ATLANTIC (Oct. 5, 2017), [https://perma.cc/NMQ5-MKGF] (describing the widespread use of the formulation "thoughts and prayers").

2. *See* AJ Willingham, *How 'Thoughts and Prayers' Went from Common Condolence to Cynical Meme*, CNN (May 19, 2018), [https://perma.cc/S6CB-8S78] (criticizing the emptiness of thoughts and prayers compared to action of some kind).

3. Christina Pazzanese, *Gauging the Bias of Lawyers*, HARV. GAZETTE (Aug. 10, 2017), [https://perma.cc/78TE-G8QB] (noting that of lawyers who made political donations in the 2016 cycle, 68% donated more to Democrats, and 76% of lawyers who both donated and who went to top law schools donated more to Democrats).

profession's role in championing civil rights and upholding justice.[4] One searches in vain to find a major law firm that does not publicly tout its commitment to improving racial diversity.[5] That public commitment dates back decades.[6]

What outcomes have been achieved in that time? For reference, non-Hispanic Whites constitute 59.3% of the overall population, and Blacks constitute 13.6%.[7] From 2007 to 2019, Black representation among law firm partners rose from 1.9% to 2.2%.[8] White attorneys constitute 89.9% of equity partners, down from 93.7% in 2007.[9] At that rate of progress, Black attorneys will still not quite be proportionally represented among law firm partners when the American Bar Association (ABA) celebrates its half-millennium birthday in 2378.[10]

The real question regarding racial diversity in law is why things are improving so slowly. One school of thought is that greater social forces are at work.[11] For example, some scholars argue women face systematic barriers to being promoted at law firms because the gendered distribution of parenting duties outside of the legal profession hinders their perceived "commitment" to the firm.[12] Perhaps a similar dynamic of racially charged assumptions leads fewer Black associates to

---

4. *See, e.g.,* David L. Douglass et al., *Signposts in the Road: The Lawyer's Ethical Obligation to Promote Diversity in the Legal Profession*, IILP REV., 2019-2020, at 52, 54, 57, 60 (stating that "the first role of lawyers in society was selfless pursuit of the common good").

5. *See, e.g.*, *Diversity, Equity, & Inclusion*, WINSTON & STRAWN LLP, [https://perma.cc/L4VN-TKTS] (last visited Jan. 24, 2023).

6. *See, e.g.*, Paul M. Barrett, *Shearman & Sterling Strives to Keep Its Black Attorneys*, WALL ST. J. (July 8, 1997), [https://perma.cc/2XCY-FWK5].

7. *QuickFacts: United States*, U. S. CENSUS BUREAU, [https://perma.cc/B9P4-GZSW] (last visited Jan. 24, 2023).

8. VAULT & MINORITY CORP. COUNS. ASS'N, 2020 VAULT/MCCA LAW FIRM DIVERSITY SURVEY REPORT 16, [https://perma.cc/XNP2-3RSK].

9. *Id.* at 24.

10. Increasing in representation at a rate of 0.3% each decade, Black attorneys will be proportionally represented among law firm partners in 2392 (assuming for argument's sake that the Black share of the overall U.S. population remains static, which it will not). *Id.* at 16.

11. *See* Hilary Sommerlad, *Minorities, Merit, and Misrecognition in the Globalized Profession*, 80 FORDHAM L. REV. 2481, 2492-94 (2012) (discussing stratification theory in which broader social injustices feed into the legal profession).

12. SUSAN EHRLICH MARTIN & NANCY C. JURIK, DOING JUSTICE, DOING GENDER: WOMEN IN LEGAL AND CRIMINAL JUSTICE OCCUPATIONS 107, 121, 126-28 (2006) (describing the advantage men get from their "release" from domestic burdens).

eventually become partners. Similarly, some argue that socioeconomic disparities are the primary cause of later discrepancies in outcome at firms.[13] There are many ways in which this could be true. For example, economic disparities between races in the United States are truly massive. In 2019, median White household wealth was $189,100, compared to $24,100 for Black households.[14] These disparities in turn lead to academic disparities, decreasing the available talent pool for law firms to ultimately recruit.[15]

The problem with this line of reasoning is that similarly situated professions are doing significantly better than law firms and the legal profession. Consider that only 56% of active physicians are White, compared to 81% of attorneys.[16] Among accountants and auditors, non-Hispanic Whites constitute 68.4%, and Blacks constitute 9%.[17] As recently as 2015, an article in the *Washington Post* declared the legal profession to be one of the "least diverse" in the country.[18] These data points suggest there are one or more factors limiting racial diversity in law that is not as acutely present in similar fields.

Another telling data point is that racial diversity among lawyers is comparatively far better among government lawyers than among law firms.[19] Despite the fact that most legal jobs are in the private sector, Black and Hispanic lawyers are slightly more likely to work for local, state, or federal government than in a law firm.[20] For comparison, 40% of White lawyers are found at

---

13. *See, e.g.*, EMMA BIENIAS ET AL., INTELL. PROP. OWNERS ASS'N, IMPLICIT BIAS IN THE LEGAL PROFESSION 11-12 (2017), [https://perma.cc/6QFJ-Q99M] (describing bias at law firms arising from cultural differences attributable to socioeconomic differences).

14. *Racial Economic Inequality*, INEQUALITY.ORG, [https://perma.cc/WAM2-43CE] (last visited Jan. 8, 2023).

15. BIENIAS ET AL., *supra* note 13, at 11-12.

16. *Diversity in Medicine: Facts and Figures 2019*, ASS'N. OF AM. MED. COLLS., [https://perma.cc/297G-9QKW] (last visited Jan. 24, 2023).

17. *Accountants & Auditors*, DATA USA, [https://perma.cc/UXM2-9C69] (last visited Jan. 9, 2023).

18. Deborah L. Rhode, *Law Is the Least Diverse Profession in the Nation. And Lawyers Aren't Doing Enough to Change That.*, WASH. POST (May 27, 2015), [https://perma.cc/8P24-TYT7].

19. AM. BAR ASS'N, ABA PROFILE OF THE LEGAL PROFESSION 2020, at 43 (2020), [https://perma.cc/RU7B-GDFN].

20. *Id.*

law firms, and just 17% work in government.[21] Black and Hispanic lawyers are also slightly more likely than White lawyers to be solo practitioners.[22] While there are certainly broader social factors deterring the entrance of Black and Hispanic lawyers into the legal profession, there is something unique about law firms that is inhibiting racial diversity.

In this Article, I will discuss several aspects of law firm structure that likely contribute to the racial diversity deficit in the legal profession. From the economics of how firms make money, to the psychology of equity partners, to the game theory of how young associates get promotions, the law firm system as it currently exists dramatically slows progress toward proportional representation. None of these factors depend on racial animus among the vast majority of attorneys. Instead, each actor in the law firm system acts in accordance with their short-term interests, and the exclusion of attorneys of color from real decision-making roles is a negative externality of those actions.

Aligning the incentives of individual actors with the overall benefit of racial diversity would solve these problems. The question is how that can be done, a thorny issue I will also discuss. In the short run, entities like the ABA can start to align incentives by researching and reporting on more meaningful metrics of racial inclusion, such as the amount of time Black attorneys actually spend with clients at big law firms. Law firms do not want to be viewed as regressive, and if they actually start losing money because of the failure to promote racial diversity, they will figure out ways to change the incentives facing everyone within the firm, including associates.

In Part I, I discuss how law firm economics and individual incentives within firms stymie racial diversity. In Part II, I describe law firm efforts to increase diversity, why they have not succeeded, and possible better initiatives. I conclude that while firms can do better, some sort of coordinated external action will likely be necessary to sufficiently alter incentives to make an appreciable difference.

21. *Id.*
22. *Id.*

I. LAW FIRMS AND RACIAL DIVERSITY

We lawyers are accustomed to thinking about ourselves as "officers of the court" representing high ideals, like ensuring every person has zealous representation and ensuring the rule of law. Even opponents in the same case are ultimately working as part of the same system striving toward justice. From this perspective, the lack of racial diversity in the legal world generally and especially toward the top echelons of law firms is puzzling. If many—possibly most—lawyers recognize the value of racial diversity, why are we failing to make progress toward it?

The short answer is that commitment to racial diversity is outweighed at every level of law firm culture by less noble but much more tangible goals—the ambition of associates, the maintenance of firm profits, upholding the image of the firm, and preserving a myopic sense of moral rightness. These factors should not be understood as personal failings but as the inevitable consequence of the high level of competition in the legal world. It is not easy for individual attorneys—whether equity partners, fresh-faced associates, or anyone in between—to survive in the intensely competitive culture if they sacrifice any advantage. The grand competition rages on between individuals at the same firm and between firms. It is difficult for outsiders to differentiate between genuine commitment to diversity and the usual promotional websites touting diversity initiatives, so firms have little incentive to make genuine commitments.

A. How Intense Competition Can Undermine Social Goals

In his influential article *Meditations on Moloch*, Scott Alexander described a simple phenomenon he calls a "multipolar trap[]," also colloquially known as a "race to the bottom":

> In some competition optimizing for *X*, the opportunity arises to throw some other value under the bus for improved *X*. Those who take it prosper. Those who don't die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been—in other

words, until human ingenuity cannot possibly figure out a way to make things any worse.[23]

Obviously, not every competition is so dire, but the more intense the competition, the likelier it is to become a multipolar trap.[24] What makes competition more intense? Scarcity of resources, lack of coordination, and lack of external limiting forces all contribute.[25] Essentially, if an entity can survive without throwing other values under the bus for improved *X*, then the competition is not a multipolar trap.

Examples of multipolar traps abound. Any situation with high stakes and few limiting factors will face pressure to sacrifice other values. The "Two-Income Trap" identified by then-Professor Elizabeth Warren is one: competition for suburban houses in good school districts leads parents to debt and demanding jobs.[26] Another is pesticide and fertilizer use: in the absence of environmental regulation, intense competition among farms can lead to the use of destructive amounts of pesticide and fertilizer.[27] In the long run, it might not be sustainable or good for the world as a whole, but the owner of an individual farm will lose on price competition and die out if they do not keep up with their competitors.

As any reader with inductive powers might surmise, I will argue that the intensity of competition in and between law firms creates a multipolar trap. Law firms optimize for making money from fees for their services. The competition is intense because successful innovation is rare in law,[28] but there is a tremendous amount of money and prestige associated with being a successful lawyer.[29] While tech companies, for example, can secure

23. Scott Alexander, *Meditations on Moloch*, SLATE STAR CODEX (July 30, 2014), [https://perma.cc/ZJJ6-4PSE] (emphasis added).

24. *See id.*

25. *See id.*

26. ELIZABETH WARREN & AMELIA WARREN TYAGI, THE TWO-INCOME TRAP: WHY MIDDLE-CLASS MOTHERS AND FATHERS ARE GOING BROKE 6-9 (2004).

27. *See* L. Bakker et al., *Kicking the Habit: What Makes and Breaks Farmers' Intentions to Reduce Pesticide Use?*, 180 ECOLOGICAL ECON. 106868, at 2, 8-9 (2021).

28. *See* THOMPSON HINE LLP, CLOSING THE INNOVATION GAP 2 (2018), [https://perma.cc/VL5H-QUGJ] (discussing a study finding only 4% of corporate counsel respondents had seen "a lot" of innovation from law firms).

29. *See* Roy Strom, *Prestige Still Beats 'Quality of Life' in Big Law Talent War*, BLOOMBERG L.: BIG L. BUS. (July 29, 2021, 5:01 AM), [https://perma.cc/2HC3-NG5G]

temporary monopoly-like advantages for themselves by inventing (and patenting) new gadgets or methods, lawyers have little room to patent, say, a novel legal argument. Everyone has access to roughly the same set of cases from which to develop arguments, so there are not many ways to create monopoly-like profits. Adding to the competition is the fact that a lawyer cannot easily retrain for some other field, certainly not at a similar level of pay.[30]

Competition has increased in the legal world for a variety of reasons over the past several decades. Overall law school enrollment in the United States increased from 46,666 in 1963 to a peak of 147,525 in 2010, which represents a per-capita doubling of lawyers.[31] The widespread integration of women into the legal community is certainly a positive development, but the introduction of more talented people into the legal world increased competition as well.[32] The ubiquity of legal research software has arguably made it easier to conduct better legal research, allowing more firms to compete for clients.[33] Legal self-help software has begun taking some low-lying, profitable fruit from lawyers, such as basic drafting of wills.[34] Big

(discussing surveys finding that the ten most prestigious law firms are hiring more than the ten firms rated best to work for).

30. *Cf.* William Vogeler, *Two-Thirds of Lawyers Want Out of the Profession*, FINDLAW (May 14, 2019, 11:00 AM), [https://perma.cc/WML8-YCX3] (describing the high portion of lawyers who want to leave the profession).

31. The number of enrolled law school students has declined from its 2010 peak to 112,878 in 2019, still representing a 40% per-capita increase. *See Law School Enrollment*, LAW SCH. TRANSPARENCY, [https://perma.cc/8ED3-MRJD] (last visited Jan. 8, 2023).

32. *See* Amanda Weinstein, *When More Women Join the Workforce, Wages Rise—Including for Men*, HARV. BUS. REV. (Jan. 31, 2018), [https://perma.cc/XW9A-WJ36] (describing the general tendency whereby increasing participation by women in a given field increases the overall competitiveness of the labor market in that field); Jennifer Cheeseman Day, *More Than 1 in 3 Lawyers Are Women*, U.S. CENSUS BUREAU (May 8, 2018), [https://perma.cc/GG5Z-Q2LZ] (discussing women's steadily increasing share of lawyer jobs).

33. *See* Paul Hellyer, *Assessing the Influence of Computer-Assisted Legal Research: A Study of California Supreme Court Opinions*, 97 LAW LIBR. J. 285, 285, 287-88, 290, 298 (2005) (describing the debate over whether electronic research has improved attorney work product and concluding that electronic research has not had as great of an effect as its proponents argued it would have).

34. *See* Megan Leonhardt, *More Than Half of Americans Don't Have a Will—This App Wants to Change That*, CNBC (Nov. 13, 2019, 10:48 AM), [https://perma.cc/9VCW-72LF] (describing apps that offer free will-drafting services in exchange for anonymized data that the company behind the app can sell to third parties).

accounting firms have even begun taking on due diligence and investigative work that might have been performed by attorneys.[35]

These developments affect both individual attorneys and entire law firms. More attorneys bring more competition for summer associate and entry-level attorney positions. Higher quality attorneys at the associate level mean that associates have to do more to stand out from their peers. Average law school debt has nearly doubled since the turn of the century, meaning that the stakes for success for individual attorneys are much higher.[36] At the firm level, increased competition means profit is simply more difficult to come by.

If we have established the competition aspect of a multipolar trap for the legal world, we can identify values thrown under the bus in the name of that competition. At the firm level, if increasing racial diversity would impose a significant burden, firms will do their best to avoid it to the extent they can do so without endangering their public image. On an individual level, sobriety and mental health are sacrificed far more often than in other professions.[37] Divorce is anecdotally more common among attorneys at big law firms.[38] The return for these sacrifices is the short-term ability to cope with the stress of competition and the time to bill more hours in hopes of outperforming others at the firm.[39] This willingness to sacrifice on the part of individual attorneys at a firm will become important later when we discuss

35. *See* Meg McEvoy, *Analysis: The Big 4 Is Knocking—Are State Bars Answering?*, BLOOMBERG L. (Sept. 18, 2019, 4:01 AM), [https://perma.cc/84YR-9JUW].

36. Melanie Hanson, *Average Law School Debt*, EDUC. DATA INITIATIVE (Nov. 7, 2022), [https://perma.cc/4NH2-HZ63] (showing an increase in inflation-adjusted average debt among law school graduates from $87,900 in 1999-2000 to $160,000 in 2019-2020).

37. Elizabeth Olson, *High Rate of Problem Drinking Reported Among Lawyers*, N.Y. TIMES (Feb. 4, 2016), [https://perma.cc/WU38-JXXB] (noting that "[l]awyers struggle with substance abuse, particularly drinking, and with depression and anxiety more commonly than some other professionals" and that "lawyers working in law firms had the highest rates of alcohol abuse").

38. *See* Harrison Barnes, *Why Big Law Firms Attorneys Are So Likely to Get Divorced: Stressed, Tired, Mad and With Nothing More to Give*, LINKEDIN (Oct. 26, 2020), [https://perma.cc/ZUF7-QZLP] (discussing various reasons why law firm lawyers are particularly likely to get divorced).

39. It should be noted that substance abuse and mental health problems are generally intended to help cope with stress in the short term but come with obvious long-term problems.

individual incentives vis-à-vis perpetuating racial diversity. Consider that if an associate is willing to sacrifice things like their own family and health in the name of competition, how much guilt would they feel over sacrificing something as abstract as racial diversity?

B. How Law Firms Compete

The foregoing discussion suggests there is intense competition between law firms, and that firms are willing to sacrifice other values if doing so will help them stay competitive. This suggests an awkward question: why does increasing racial diversity impose a burden on law firms?

To answer that question, we must consider how law firms compete with each other. At the simplest level, law firms make a profit by taking in more in fees from clients than they expend in retaining attorneys, support staff, and the physical infrastructure of business. Assume, for the sake of this analysis, that firms have made the support staff and physical infrastructure components as efficient as they can be—there is no firm out there that could meaningfully outcompete the others by, say, cutting a really good deal on their office rent or computers.[40] Eliminating those aspects of the profitability equation leaves us with two factors to play with: fees taken in, and salaries paid to attorneys.

There are two dominant and countervailing considerations in maximizing fees: (1) clients fundamentally do not know whether they are being overbilled;[41] and (2) there are increasing returns to doing more work for existing clients.[42] The first point is not at all unique to law. Fields ranging from auto maintenance to medicine involve clients who hire specialists to perform a service where

---

40. Of course, these factors matter on the margins, but note that expenses related to employee turnover make up a much larger share of overhead costs than things like office space or technology. *See* Rikke Diget Fuglsang, *Law Firm Overhead: Understanding, Diagnosing, & Fixing Profit Killers*, ASKCODY (June 1, 2022), [https://perma.cc/4Q9R-UMTJ].

41. *See* Douglas E. Rosenthal, *Evaluating the Competence of Lawyers*, 11 LAW & SOC'Y REV. 257, 257 (1976) ("People do not how to choose a lawyer at the outset, or how to evaluate his or her performance during the lawyer-client relationship.").

42. *See* Tyler J. Replogle, *The Business of Law: Evolution of the Legal Services Market*, 6 MICH. BUS. & ENTREPRENEURIAL L. REV. 287, 289, 291-92 (2017) (describing close, long-term client relationships as producing a "Golden Era" for law firms).

they cannot easily verify whether they are paying for necessary work. To overcome that discrepancy in knowledge, some kind of third-party verification is generally helpful. For example, doctors have a limited ability to bill patients because they have to work with health insurance companies possessing a similar level of expertise and experience on payment for health care.[43] There is no direct third-party verification of that kind in law. While there are rules regarding the assessment of fair fees,[44] the entity judging whether the fees are fair—the client—has far less expertise than an insurance company. One cannot even use outcomes to judge whether a client was overbilled because there is no way to answer the counterfactual question of whether a different law firm would have done better.[45] The answer to this principal-agent dilemma in law turns out to be similar to that in auto maintenance: trust based on repeated interactions.[46]

On the second point, increasing returns to additional work for existing clients, both clients and firms profit from having a longstanding relationship. Attorneys at the firm better understand the client's preferences for work product, and they bring greater knowledge to the table that can help maximize the value of their work.[47] Clients come to know they can depend on the firm's expertise, and repeated interactions can at least theoretically reduce billing because the firm already knows the background in their area of law.[48] Firms get a large and constant stream of

---

43. *See* N. GREGORY MANKIW, THE ECONOMICS OF HEALTHCARE 8 (2017), [https://perma.cc/K4HX-42YF] (describing the use of payment rules by insurers to guide physicians' recommended treatments to patients).

44. *See* MODEL RULES OF PRO. CONDUCT r. 1.5(a) (AM. BAR ASS'N 1983) (listing factors relevant to determining whether a fee is reasonable).

45. Rosenthal, *supra* note 41, at 263-64.

46. *See id.* at 263, 265-66. Note additionally that this is inherently conservative criteria—a client surely knows there are other competent attorneys, but they do not know how to identify them, so they rationally prefer to stick with their current choice. This unfairly casts aspersions on other competent lawyers not selected.

47. *Cf.* Thomas Kollar & Stephanie Mills, *Perspective: Why Secondments Are Even More Valuable to Law Firms Today*, BLOOMBERG L. (Feb. 26, 2016, 1:09 PM), [https://perma.cc/H6F5-AFPS] (discussing how attorneys can do their job better with better understanding of their corporate clients).

48. *See* Peter D. Sherer, *Leveraging Human Assets in Law Firms: Human Capital Structures and Organizational Capabilities*, 48 INDUS. & LAB. RELS. REV. 671, 673-74 (1995) (describing the benefits of diversification and dedication that accrue from a long-term relationship with a full service law firm).

business.[49] Consequently, the optimal situation for a large firm is to leverage its size and diversity of expertise to become a "full service" firm, i.e., they can do all or nearly all work that a client needs.[50] This is similarly beneficial to clients because the value of the overall relationship is an implicit promise of high-quality work in each issue area—the relationship is too valuable for the firm to do anything less than a satisfactory job.[51]

These two considerations suggest the way for law firms to profit is to maximize client trust in order to secure as long a relationship as possible. Thus, maximizing client trust is what law firms are truly optimizing for. The factors that build client trust in the firm are not difficult to surmise: (1) discernible quality of work; (2) results; (3) responsiveness; (4) familiarity and predictability in key personnel.[52]

C. The Effect of Inter-Law Firm Competition on Racial Diversity

For a variety of unjust economic and psychological reasons, increasing racial diversity among senior members of a big law firm can undermine the four factors described above in building client trust. On discernible quality of work, as discussed above, clients do not possess the expertise to perfectly monitor performance, but they are human, and so tend to exhibit predictable biases. For example, several psychological studies have indicated that people are less inclined to view someone they perceive to be an "affirmative action hire[]" as competent.[53]

49. *See* Patrick Smith, *The Art of Developing and Maintaining Make-or-Break Client Relationships*, AM. LAW. (June 29, 2021, 8:04 AM), [https://perma.cc/A7EJ-6JPU] ("No billion-dollar law firm could keep churning without having a few whales in the mix. And most Big Law firms have more than a few—the kind of clients with an enterprise so expansive that their need for outside counsel seems limitless.").

50. *Id.*

51. *See* Ronald J. Gilson & Robert H. Mnookin, *Sharing Among the Human Capitalists: An Economic Inquiry into the Corporate Law Firm and How Partners Split Profits*, 37 STAN. L. REV. 313, 365-66 (1985) (comparing the full-service firm to a department store where the brand's perception of quality rests on quality across the board in every service area).

52. *Cf.* Jeremy M. Evans, *Building a Client Base: Top Tips from Practitioners*, AM. BAR ASS'N, [https://perma.cc/WE4T-TC9F] (last visited Jan. 9, 2023).

53. *See* Madeline E. Heilman et al., *The Affirmative Action Stigma of Incompetence: Effects of Performance Information Ambiguity*, 40 ACAD. MGMT. J. 603, 603 (1997) (finding

People also often judge attorneys based on superficial qualities—how they talk, what their accent is, how they dress, etc.[54] One can immediately see where economic disparities would particularly disadvantage attorneys of color in this regard.

Measuring "results" is a tricky business for both clients and law firm management.[55] The "discernible quality of work" discussion in the preceding paragraph casts a shadow over this factor as well. In most legal contexts, there is a spectrum of outcomes. Yes, a case may be won or lost, but damages or settlements could be greater or smaller. Even the binary outcome of win/loss does not reliably connote quality work—perhaps the case seemed closer than it ought to have been. The less faith the client has in the attorneys she works with, the less charitably she will judge their results.

Responsiveness is a seemingly benign quality, but a simple thought experiment suggests why it ends up inhibiting racial diversity. Assume there are two equally competent attorneys, *A* and *B*, at a large law firm. *A* is a single man coming from a wealthy family. *B* is a single mother coming from a poor family. Attorney *A* has few, if any, limitations on responsiveness. Attorney *B* has countless limitations—she may not be able to check her phone constantly and respond while caring for children, helping relatives, etc. A cold-blooded, rational client who wants

---

that test subjects rated female affirmative-action hires as less competent than men or women not associated with affirmative action even if disconfirming performance information was provided); *see also* David B. Wilkins, *From "Separate Is Inherently Unequal" to "Diversity Is Good for Business": The Rise of Market-Based Diversity Arguments and the Fate of the Black Corporate Bar*, 117 HARV. L. REV. 1548, 1572 (2004) ("As Paul Barrett discovered when doing interviews for his revealing book about race in large law firms, several partners hold the view that the real reason why firms have so few black partners is that there are too many incompetent black lawyers who have been 'polished up' by affirmative action to look like Harvard Law School graduates.").

54. *See, e.g.*, Adrian Furnham et al., *What to Wear? The Influence of Attire on the Perceived Professionalism of Dentists and Lawyers*, 43 J. APPLIED SOC. PSYCH. 1838, 1847 (2013) (finding that male lawyers in professional and formal attire were rated as more suitable, capable, easier to talk to, and friendlier).

55. Some argue that the partner structure is necessary for law firms because non-expert owners and managers cannot monitor professionals whose work they do not understand. *See, e.g.*, Larry E. Ribstein, *The Death of Big Law*, 2010 WIS. L. REV. 749, 754-55. I believe that partners, while better situated than non-lawyers to understand an associate's work, cannot verify things like whether an important case was missed in research. Consequently, I think partners are largely subject to the same information asymmetry as clients and hypothetical non-lawyer managers.

maximum responsiveness will always choose Attorney *A*. As a knowledgeable reader can guess, White attorneys are far likelier than attorneys of color to resemble Attorney *A*, and attorneys of color are far likelier than White attorneys to resemble Attorney *B*.[56]

Familiarity and predictability in key personnel also weighs against having attorneys of color in client-facing roles for a variety of unfair reasons. First, there is a chicken-and-egg problem: if attorneys of color are underrepresented, by definition including them more often in client relationships will require change in who the clients see.[57] Change is the very antithesis of familiarity. Second, undoubtedly due in part to lack of advancement to partner, attorneys of color leave firms more often, making them a riskier steward for a client relationship in the eyes of partners.[58] Third, racial disparities in wealth mean that large-firm clients are likelier to be White.[59] A mountain of psychological evidence indicates most people, not just Whites, *ceteris paribus* trust people of their own race more than others.[60]

None of these factors *justifies* law firms' failures to meet longstanding goals for racial diversity in the legal profession. Rather, they explain what firm management might view as the cost of increasing racial diversity. Firms, or more specifically the management of firms, absolutely could choose to bear that cost. They could promote more attorneys of color to partner. They

56. *See* DESTINY PEERY ET AL., LEFT OUT AND LEFT BEHIND: THE HURDLES, HASSLES, AND HEARTACHES OF ACHIEVING LONG-TERM LEGAL CAREERS FOR WOMEN OF COLOR, at ix (2020), [https://perma.cc/J6DP-STMB] (describing studies finding, among other things, that women of color were more likely than White women and men to report having extended family responsibilities).

57. Some attorneys of color report feeling used as a token in client interactions, being "trotted out to clients only when it would help the firm look good but not necessarily in ways that helped them further their own careers." *Id.* at viii. Clearly, my analysis here means attorneys of color having substantive relationships with clients, not just being literally in the room.

58. *Id.* at ix (finding that women of color were more likely to report that they were seriously considering leaving their law firms).

59. *See* Debra Cassens Weiss, *Would-Be Clients with White-Sounding Names Got 50% More Lawyer Responses in California, Report Says*, ABA J. (June 6, 2019), [https://perma.cc/V597-9S34].

60. *See, e.g.*, Adam Okulicz-Kozaryn, *Are We Happier Among Our Own Race?*, 12 ECONS. & SOCIO. 11, 13 (2019) (discussing several studies finding that racially homogeneous areas have more civic engagement, more trust, and more redistributive policies).

could explain to clients that they are consciously empowering attorneys of color, and if it costs a few clients on the margins, then so be it. But the data suggest most firms do not do this.[61]

One important wrinkle to this analysis is that all available evidence suggests that diversity helps businesses generally and law firms in particular in the long run.[62] The general rationale is that a more diverse team will bring complementary skills to the fore.[63] This is entirely plausible in the legal context, particularly given that large firms provide a wide range of services. The more varying the work, the more likely that a diverse team would do better than a homogeneous one.[64] To understand why this sort of potential long-term benefit is overshadowed by short-term client retention concerns, we must look at the incentives of individual decision-makers within the firm.

D. Incentives for Individuals Within Law Firms

The incentives of attorneys at law firms work against racial diversity for reasons completely unrelated to firm-level considerations. To understand why, we should consider the ways in which attorney and firm incentives are misaligned. But the incentives facing attorneys vary tremendously within a law firm depending where in their career they are. The relative importance of reputation, money, and prestige vary predictably with age and socioeconomic status.[65]

In this discussion, we should be aware of the distinction between financial and personal incentives. Most professionals are motivated by some mixture of both. Virtually no one would work for free, and absurd hypotheticals aside, no amount of

---

61. *See* ROY S. GINSBURG, AM. BAR ASS'N, DIVERSITY MAKES CENTS: THE BUSINESS CASE FOR DIVERSITY 1 (2014), [https://perma.cc/654T-6FT6] (finding that law firms with higher diversity attract and retain better lawyers, devise better strategies for clients, and can attract new business).

62. *Id.*

63. *Id.*

64. *Cf.* David Rock & Heidi Grant, *Why Diverse Teams Are Smarter*, HARV. BUS. REV. (Nov. 4, 2016), [https://perma.cc/9VUF-EXM2] (describing higher innovation and more focus on facts on diverse teams).

65. *See* Chang-ming Hsieh, *Money and Happiness: Does Age Make a Difference?*, 31 AGEING & SOC'Y 1289, 1289 (2011) (finding that money is not significantly correlated with happiness in older adults).

money will make most people do personally repulsive work. It is far easier to identify and dissect financial incentives than personal incentives in an objective way, but in some instances, we can identify broadly shared personal goals distinct from finances. For example, many lawyers care greatly about the esteem of their colleagues and professional advancement.[66] For deep-seated psychological reasons, they may crave the approval of firm leadership. As we will see, these less tangible considerations can compound already powerful financial motivations.

1. Associate-Level Incentives

The essence of an associate is aspirational competitiveness untempered by the attachments of older attorneys. Their financial situation is complex and nuanced. They generally do not have a relationship with their clients; rather, they do work for the partners of the firm.[67] Often, they have considerable debt from law school.[68] They are typically paid a handsome salary with some limited form of performance incentive (e.g., a bonus for meeting a billable-hours goal).[69] Because they tend to live in urban areas, associates have high cost-of-living, taking a little of the shine off their handsome salary.[70] An associate at a firm in, say, New York, has a salary that almost—but not quite—puts them in the top 10% of earners in the city.[71] While this is certainly

66. *See, e.g.*, Walker v. City of Mesquite, 129 F.3d 831, 832 (5th Cir. 1997) (describing an attorney's professional reputation as "a lawyer's most important and valuable asset").

67. *See* Jordan Rothman, *Biglaw Associates Should Interact with Clients More*, ABOVE THE L. (June 2, 2021), [https://perma.cc/TZ6E-FZZ3].

68. *See* Hanson, *supra* note 36 (reporting average law school debt at $180,000 in 2020-2021).

69. *See* Staci Zaretsky, *Associate Compensation Scorecard: Biglaw's 2020 Bonus Bonanza*, ABOVE THE L. (Nov. 25, 2020, 10:28 AM), [https://perma.cc/FAP2-YQQU] (sharing a full spreadsheet of associate compensation at large law firms, many of which have bonuses).

70. *See Cost of Living Data Series*, MO. ECON. RSCH. & INFO. CTR., [https://perma.cc/49R2-RM4N] (last visited Jan. 9, 2023) (showing a much higher cost of living in and around urban areas).

71. *See New York City, New York Population 2023*, WORLD POPULATION REV., [https://perma.cc/L8KJ-T2TS] (last visited Jan. 9, 2023). Note that starting salary at the law firm Cravath, Swain, and Moore is approximately $200,000. *See* Meghan Tribe, *Cravath Tops Rival Davis Polk's Associate Pay Scale, Up to $415k*, BLOOMBERG L. (Feb. 28, 2022), [https://perma.cc/HQ7H-2ZVY].

noteworthy, given where they live, associates cannot help but encounter wealthier people, including the partners they work for.

As for personal incentives, we can surmise several important factors based on demographic facts. Associates are less likely than partners to be married or have children.[72] People with children tend to be less ferociously competitive or monomaniacal, so associates have a greater tendency to focus on career progress.[73] Because they are younger, associates are more likely to believe in statements like "I have a destiny."[74] The personal incentives of associates are thus intently focused on a one-dimensional but achingly imprecise metric: success at the firm.

In a more abstract sense, associates have been primed by a decade of intense competition to optimize for individual competence and pleasing superiors. Associates have recently completed a long educational process in which they competed and competed to get into the best, in this order: college, internships, law school, academic journal, summer associateship, law firm, judicial clerkship. In that process, everything from studying for the SAT to sending out clerkship applications is geared toward individual aspiration, ambition, and competition. There is only marginal or incidental organization-wide benefit for things like making the law review, but there are intense personal benefits—namely, increased competitiveness for the next competition.

The ways in which associate incentives diverge from firm-wide incentives are clear. Baldly, the associate is focused on her own success because she does not wield sufficient power to significantly alter her firm's fate.[75] The time horizon of associates is quite limited—most associates will leave the firm

72. *See* Jeena Cho, *Family Way*: *Lawyers on Balancing Motherhood or Choosing a Child-Free Life*, ABA J., Nov. 2018, at 26, 26; *see also* Esther Lee, *This Was the Average Age of Marriage in 2021*, THE KNOT (Feb. 15, 2022), [https://perma.cc/B8RN-CLCV].

73. *See* Art Markman, *How Do People's Values Change as They Get Older?*, PSYCH. TODAY (Sept. 1, 2015), [https://perma.cc/JWG3-5XSR].

74. *Cf.* Kelly A. Durbin et al., *Optimism for the Future in Younger and Older Adults*, 74 JS. GERONTOLOGY 565, 565 (2018) (describing optimism about one's own future decreasing with age).

75. *See* PETER-J. JOST, THE ECONOMICS OF MOTIVATION AND ORGANIZATION 284 (2017) (describing opportunistic behavior in employees when their individual interests diverge from those of the firm).

within eight years, unsuccessful in their pursuit of partnerdom.[76] Associates thus naturally focus on personal success over firm prosperity.

2. The Effect of Associate Incentives on Racial Diversity

Associates may seem a strange group to discuss when considering the lack of racial diversity at law firms. After all, associates do not choose their peers, and they do not decide which of their number reach the rank of partner. What associates could, in theory, control is the way in which they compete to become a partner, for therein lies the problem. Regardless of whether racial diversity would benefit the firm, it is quite irrelevant to an associate how racially diverse her firm is. Indeed, if the associate is White, racial diversity is an active impediment to winning the next round of competition and becoming a partner. Under those circumstances, it should not be surprising that associates engage in what is ordinarily benign behavior, but which has the tendency of forestalling racial diversity.

Many first-hand accounts describe the dominant strategy to becoming a partner: do your work well and forge relationships with partners.[77] The former is obvious enough, and aside from prejudicial racial aspects of the "responsiveness" factor discussed earlier in this Article, this area is not generally a limitation on racial diversity.[78] Forging relationships with partners is seemingly benign as well, but such a reaction depends on a naïve view of how such relationships work. In the words of one corporate law partner: "Partners often see themselves in the associates they assign work to. That means if you didn't share

---

76. *See* Nicole Donnelly, *The Shelf Life of a Law Firm Associate*, LINKEDIN (May 13, 2015), [https://perma.cc/M2MC-94PK].

77. *See The Allocation of Work*, HARV. L. SCH. CTR. ON THE LEGAL PRO. (2017), [https://perma.cc/X37A-7XBL] (extensively discussing the value of an associate's relationships with partners to succeed at a law firm).

78. *But see* Melanie Lasoff Levs, *The Partnership Track: Everything You Didn't Learn in Law School*, DIVERSITY & BAR, May-June 2005, [https://perma.cc/8YX2-6USJ] ("[W]hereas a white associate will maybe not know exactly what he or she is doing and the work product is not the best, the response will be, 'He's green. He needs more training . . . .' When an African American or other minority associate makes the same error, has work product needing improvement or has the same issue of not knowing, the presumption is sometimes incompetence.")

similar characteristics or interests, it was that much harder to build relationships of trust. There was nothing actively malicious about it, but we can't deny that it was occurring."[79]

It does not take much imagination to see where this goes wrong. The upper echelons of firms are predominantly affluent White men.[80] If we expand the focus to affluent White people generally, it is an overwhelming majority.[81] White associates will already naturally share more "characteristics and interests" with White partners. They will also be in a better position to feign shared interests without scorn or ridicule—imagine the varying reactions if a White associate pretends to be interested in sailing as opposed to an associate of color from a disadvantaged background. Even putting any conscious discrimination aside, a White person in their late-twenties and a White person in their, say, mid-fifties are far likelier to share interests than a Black person in their late-twenties and a White person in their mid-fifties.

The high incentives for competitiveness among associates also lead to competition in billable hours. As discussed earlier, economic disparities between races mean that it is easier for White associates to take on as many billable hours and matters as necessary to stand out.[82] Similarly, even for associates, much legal work depends on presentation, opening another advantage for children of wealth. Whether it is simple upper-class diction or the use of highfalutin vocabulary, there are many ways to make substantively identical arguments *sound* better to a partner reviewing an associate's work. Under such circumstances, the partner may not even be consciously aware of respecting the associate's work more, and the associate is merely writing the way she learned how to over decades of schooling.

Associates could manage these problems, at least in theory. They could agree among themselves not to engage in the sort of

79. *The Allocation of Work*, *supra* note 77.

80. *Id.*

81. *Id.*

82. *See supra* Section I.C.; Richard H. Sander, *The Racial Paradox of the Corporate Law Firm*, 84 N.C. L. Rev. 1755, 1800 (2006) (noting the wide gap in respondents who said they had worked on nine or more matters over the past six months—59% among White men; 33% among Blacks; 38% among Hispanics).

toadying that unfairly benefits White associates. They could jointly agree to limit their billable hours. Individual associates could decide on their own not to take advantage of the benefits. Whether one can fairly ask a twenty-six-year-old associate to take on such a responsibility is debatable, but taking the incentives facing associates into account, there is no realistic possibility of associates voluntarily relinquishing such an advantage. These problems are particularly thorny because they involve behavior that is almost always positive and good. Indeed, it sounds ludicrous to talk about friendliness, socialization with older coworkers, and hard work as things that forestall racial diversity. It is unlikely every White associate in a firm would forgo their advantage, and even if they claimed they would, there would be no way to verify that they were doing so. Very few associates behave in that way with the conscious knowledge that it is an unfair advantage.[83]

It is difficult to quantify the effects of these factors, but available anecdotal and scientific data suggest a strong hindrance to racial diversity. As discussed earlier in this Article, people of color are dramatically underrepresented at the partner level.[84] The increased difficulty for associates of color in their career path at the firm predictably leads to higher attrition of those associates. Sure enough, Black and Hispanic associates leave law firms at a higher rate than White associates.[85] As of 2019, the average governance and compensation committees at a law firm were composed of twelve people and had only one person of color.[86] Law firms have increasingly hired associates of color, but that progress has barely made a dent in the disparity at the partner level.[87]

---

83. *See supra* Section I.

84. *See supra* text accompanying note 80; *The Allocation of Work*, *supra* note 77.

85. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 8.

86. NAT'L ASS'N OF WOMEN LAWS., 2019 SURVEY REPORT ON THE PROMOTION AND RETENTION OF WOMEN IN LAW FIRMS 7 (2019), [https://perma.cc/3H5G-LJDK].

87. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 6.

3. Partner-Level Incentives

In both financial and personal terms, incentives for partners weigh against increasing racial diversity. However, the financial aspect is arguably more important and the personal aspect less important in impeding racial diversity in firms. Partner compensation is tied directly to revenue creation in a way that associate compensation is not. Partners make far more money than associates, but it is taken as a share of the firm's revenues rather than a fixed salary.[88] This theoretically should create a broader perspective whereby partners will do what is good for the firm overall.[89] A partner should, as the theory goes, be a dispassionate, mature presence, willing to forgo the self-promoting passions of associates in order to do the right thing for the firm.[90] In practice, partners are often paid based on self-focused metrics like billable hours and revenue generation, meaning that their financial incentive is squarely fixed on short-term gain.[91]

Much has been written about how to better align partner incentives with the overall well-being of the firm, but the variety of practices between law firms suggests that no particular solution works significantly better than the others.[92] Most feature a mix of claiming revenue directly from fees taken in from the partner's clients, billable hours, and reaching some firm-set goal.[93] At one extreme, in an "eat-what-you-kill" system, partner compensation is mostly or entirely based on fees taken in from the attorney's own clients.[94] That system creates a strong incentive not to refer

88. Shari Davidson, *Law Firm Compensation: How Are Partners Paid, Compensated*, JD SUPRA (June 16, 2021), [https://perma.cc/497J-HWWU]. This is, of course, an "equity" partner. *Id.* There are non-equity partners in many firms now as a sort of intermediary position between an associate and the classic "partner." *Id.*

89. *See* Ribstein, *supra* note 55, at 754-55 (describing compensation schemes for partners as intended to "reward[] [partners] for the firm's overall success to motivate them to contribute to this success by monitoring the other worker-owners").

90. *Id.*

91. *See* Paul Floyd & Nick Ryan, *Thinking Outside the Black Box: Reimagining Attorney Compensation for the 21st Century*, BENCH & BAR MINN., Mar. 2020, at 16, 16, 18 (2020) (describing the different current approaches to attorney compensation at large law firms).

92. *Id.* at 17-19.

93. *Id.* at 17.

94. *Id.* at 19.

clients to other attorneys in the firm—the partner profits from his own clients, not the clients of other attorneys at the firm. This dynamic weakens the value of having a "full service" firm because each partner is incentivized not to call upon the rest of the firm's expertise if it is even remotely possible that he can provide the service.[95] At the other extreme is "lock-step" compensation, where salary is based on seniority within the firm, and everyone at the same seniority level makes the same amount.[96] While this can enable a partner to focus on what is good for the firm, it also removes a strong incentive to generate business.[97] While it is difficult to generalize, observers have noted a decline in lock-step systems, driven by a desire to attract lawyers who are particularly outstanding at generating revenue.[98]

At the personal level, partners are overwhelmingly wealthy older White men—90% of equity partners are White, and 78% are male.[99] Demographically, older White men are the least likely to be receptive to making a sacrifice on behalf of racial diversity.[100] Older White people generally, but particularly men, are less likely to believe Blacks are treated less fairly than Whites.[101] In one 2019 survey, about 60% of White men in Generation Z believed Blacks were treated less fairly versus 40% of the Baby Boomer generation.[102] Perhaps even more notable, nearly 20% of White male Baby Boomers believe Whites are treated less fairly than Blacks.[103]

Finally, it is worth noting that partners are much more sensitive to client wishes than others at the law firm. Partners manage the relationship and profit most directly from it,

---

95. *Id.*

96. Floyd & Ryan, *supra* note 91, at 18.

97. *Id.*

98. *See* John Roemer, *In the Money*, ABA J., Apr.-May 2021, at 24, 24-25 (describing the competitive disadvantage of lockstep compensation, namely that lateral hiring of "rainmakers" is a necessary practice in the modern legal marketplace).

99. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 7.

100. *See* Philip Bump, *Most Young White Men Are Much More Open to Diversity than Older Generations*, WASH. POST (Jan. 20, 2019, 1:29 PM), [https://perma.cc/4BR6-NYNQ] (showing various polling indicating that White men of the boomer generation were significantly more likely to say that increasing racial and ethnic diversity is not a good thing).

101. *Id.*

102. *Id.*

103. *Id.*

especially in eat-what-you-kill systems.[104] To the extent that clients have racial or other bias, partners will be the ones who have to either accommodate it or figure out some way to neutralize it.

4. The Effect of Partner Incentives on Racial Diversity

Partners can inhibit racial diversity at their firms in three main ways: (1) they can steer work toward White associates; (2) they can prevent significant client interaction with minority attorneys; and (3) they can choose not to promote minority attorneys to the partner level. The data at hand strongly suggest (3) is happening currently and has been for some time—there is far more racial diversity at the associate level than at the partner level.[105] I contend that the incentives facing partners, primarily financial, militate against the inclusion of minority attorneys, which then snowballs into a lack of mentorship and access to senior firm leaders, culminating in the systematic failure to promote minority associates to the partner level. The incentives depend somewhat on the kind of partner-compensation strategy used by the firm, but the end result is always that racial diversity is undervalued.

In an eat-what-you-kill system, virtually all of the short-term financial sacrifices of increasing racial diversity are borne by the partner, and virtually none of the long-term benefits accrue to him. The benefits of racial diversity might be seen in better work product and a more diverse client base for the overall firm, but not for a run-of-the-mill older White male partner. Clients cannot easily discern higher quality work product, and the central flaw of the eat-what-you-kill system is that the partner does not necessarily care whether the firm gains clients if they are not *his* clients.[106] The short-term costs of increasing racial diversity might be less apparent stability for clients used to working with White attorneys—precisely what the partner wants most in order to keep his clients as comfortable as possible with the relationship.

---

104. *See* Floyd & Ryan, *supra* note 91, at 18-19.
105. *See* VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 7.
106. *See* Floyd & Ryan, *supra* note 91, at 18.

In a lock-step compensation system, increasing racial diversity brings different costs for partners. Recall that the major drawback of the lock-step compensation system is lack of incentive for creating new business.[107] The best scenario for a partner in a lock-step compensation system is to do as little work as one can get away with and be surrounded by partners who are, to use the parlance of the profession, "rainmakers." The worst scenario is to be the most productive partner in such a firm because it means you are being compensated below your productivity. Surveys of law firms indicate that revenue generated by White partners is nearly 60% greater than that generated by Black partners.[108] White partners generated nearly double the revenue of Hispanic partners.[109] We can surmise a variety of reasons for that discrepancy—bias by clients, lack of support from others at the firm, etc., but incumbent partners at a lock-step law firm would, presumably, predict diminished income from increased diversity.

On the personal level, there are few counterweights for partners to the financial disincentives to racial diversity. As discussed above, demographically, the White men who comprise a supermajority of partners are more likely to be skeptical of claims of racial injustice.[110] Some partners have ambitions to, say, become a judge or political figure, but confirmation hearings or television interviews are unlikely to delve into detailed specifics about whether the partner voted to promote minority associates or made a real effort to mentor them. Some partners are content to remain at the firm, perhaps because they are interested in the subject area they work in. Such partners would be similarly indifferent to the plight of minority associates. The only partners likely to ignore financial incentives to improve racial diversity would be those who care tremendously about racial injustice, did not leave to work directly on the issue in the

---

107. Mustaqim Mohammad Iqbal, *Long Live the Lockstep*, THE STUDENT LAW. (Oct. 21, 2019), [https://perma.cc/LN67-PV22].

108. *See* JEFFREY A. LOWE, MAJOR, LINDSEY & AFRICA, 2020 PARTNER COMPENSATION SURVEY 29 (2020), [https://perma.cc/ZFK8-SQM5].

109. *Id.* (listing White, non-Hispanic average originations per partner at $2,888,000 and Hispanic average originations per partner at $1,471,000 in 2020).

110. Bump, *supra* note 100.

nonprofit sector, and stuck with the firm long enough to realize that vision. Those partners doubtless exist, but given the paucity of partners of color, their quest for diversity has proven quixotic.

5. Short-Term vs. Long-Term Incentives

One attribute common to everyone at a law firm is a shorter time horizon than the firm itself. Associates are not likely to last a decade at the firm where they start their legal careers.[111] Even the longest-term partners cannot realistically hope to be at the firm for more than, say, forty years. Many of the largest firms are already significantly older than that and in all likelihood will not shut down anytime soon. Cravath, Swaine, and Moore, consistently ranked among the top law firms in the country, was founded in 1819.[112] The difference in time horizons creates a natural incentive to paper over problems rather than solve them.

In the racial diversity context, there is an obvious short-term cost in attempting anything ambitious: admitting there is a problem in the first place creates a negative impression in the short term. The firm itself might be better off in, say, ten years, but in the short run, clients will not want the reputational risk of being associated with a law firm known to have a racial diversity problem. As we have seen, partner compensation is much more directly tied to what clients want than the long-term well-being of the firm.[113]

A recent ABA publication aimed at spurring in-house law departments to retain outside counsel of color described the problem with this memorably awful circumlocution in a section titled "Highlight Law Department's Lack of Diversity":

> An in-house law department that is not already diverse must ensure that its operations reflect D&I [diversity and inclusion] principles before implementing a law firm diversity program. If a law department does not prioritize D&I in its operations, majority-owned firms

111. *See, e.g.*, Link Christin, *Confronting Lawyer Turnover in Law Firms*, ATT'Y AT WORK (Mar. 27, 2021), [https://perma.cc/6KHL-5LHW].

112. *See Our Story*, CRAVATH, SWAINE & MOORE LLP, [https://perma.cc/D2L5-37YP] (last visited Jan. 9, 2023).

113. *See* Floyd & Ryan, *supra* note 91, at 18-19.

cannot be expected to follow a department's law firm diversity guidelines.[114]

While the meaning of this guidance remains unclear, it seems to indicate that in-house lawyers who retain outside law firms should not do so unless they themselves have a sufficiently diverse workforce. Presumably, the fear is that unless one's house is already in order, any significant action relating to diversity could bring unwanted attention.

E. Why Clients Do Not Demand Diversity at Law Firms

If the incentives of firms, and in particular partners at firms, revolve around pleasing clients, we must ask why clients have not pressured the law firms they hire to improve racial diversity. The mystery deepens when we realize that most major corporations profess a commitment to racial diversity, and many at least appear to walk the walk, with far more minority representation at higher levels than in law firms.[115] While it is impossible to look into the hearts of clients and generalize, we can examine a few relevant considerations to try to address this mystery.

First and likely foremost, clients benefit from internal diversity at their firms but have a far less direct interest in creating a diverse workplace from a law firm that they employ. Perceived short-term costs in quality and responsiveness accrue directly to the client; whatever benefits the law firm's future clients reap from a more diverse workforce do not redound to the current client. While this is a compelling argument, the one nagging doubt is that if we believe the legal world has particularly acute competition that it is suffering from a multipolar trap, can it really be the case that clients are generally subject to the same dynamics even though they are generally in less competitive industries?

While clients may work in less competitive fields, they often have high stakes in the matters for which they retain counsel, and

114. David O'Connor, *Increasing Law Firm Diversity*, DIVERSITY & INCLUSION COMM. (Am. Bar Ass'n, Chi., Ill.), Winter 2020, at 1, 7, [https://perma.cc/BFC5-D3MW]. The diligent reader may also note that several of the cartoon women of color in the graphic above the article appear to have been copied and pasted, with only the color of their shirt changed.

115. Tracy Jan, *The Legal Profession is Diversifying. But Not at the Top.*, WASH. POST (Nov. 27, 2017, 8:08 AM), [https://perma.cc/XKW3-M7VE].

therein lies the incentive against diversity. It is easy for attorneys to forget just how anomalous legal services are as a product in today's world. Most products and services are somewhat predictable (e.g., you can build 10,000 cars if you put in *X* amount of money), insured against loss, and not particularly dichotomous in outcomes (e.g., you might make cars of a slightly lower quality than intended, but they will still generally work). If one hires, say, an accounting firm, the likelihood of outright disaster is low because accountants' struggle is against inanimate numbers. While accountants can and do make big mistakes, they are relatively rare and can be insured against.[116] A law firm is often hired to undertake an important and highly uncertain mission. Under those circumstances, clients understandably optimize for sheer competence and victory in their particular case above all else. They are willing to pay exorbitant fees because no one knows how high of quality the legal work must be to end in victory.[117] Because they optimize for competence, they understandably want their law firm to do the same. And because they cannot easily judge whether the end product was high quality, they look for ancillary factors like responsiveness as proxies for quality representation, bringing us right back to the same incentives that work to hinder racial diversity inside law firms.

Against this backdrop, we can see how the outrageously unfair stereotypes of attorneys of color at major law firms end up neutralizing clients' desire for racial diversity in the law firms they employ. As discussed earlier, they cannot discern quality, but they think they know that a responsive White man is working on their case with younger responsive White men, so they are getting their money's worth.[118]

---

116. *See* Roger Russell, *It's a Good Time to Buy Liability Insurance*, ACCT. TODAY (Oct. 2, 2017, 11:52 AM), [https://perma.cc/A9DJ-SCZN]. Obviously, lawyers have malpractice insurance as well, but the difference is that you can lose and be competently represented by an attorney. An accountant's "loss" is much more inherently due to a fault of the accountant's.

117. *See, e.g.*, Harry S. Margolis, *Why Are Lawyers So Expensive?*, MARGOLIS & BLOOM, [https://perma.cc/H3Y9-3E8G] (last visited Jan. 9, 2023) (providing a client-side view of legal services: "They are willing to pay for the right representation because so much is at stake.").

118. *Id.*

F. How Law Firms Attempt to Manage Their Racial Diversity Problem

While misaligned incentives explain much of why firms have not made progress on racial diversity, the next logical inquiry is what do firms actually *do* about racial diversity and why has it failed to produce results. Understanding what firms have done and why requires looking past stated or even honestly held intentions to hard reality. The dominant solution hit upon by major law firms involves a mixture of diversity committees, formal mentorship programs, affinity groups, and some manner of social outreach. Unsurprisingly, these efforts have not significantly altered the trajectory of racial diversity in law because they do not address the incentives at the heart of the problem. While they may be well-intentioned, these efforts largely function as public relations management, whether to aid in the recruitment of associates of color or to have something to point to when critics inevitably point out the dismal lack of racial diversity in law.

1. Law Firm Diversity Committees

The single most widely adopted measure to increase racial diversity at large law firms is the diversity committee.[119] For large law firms, a diversity committee is usually comprised of about twenty attorneys, at least one of whom is a partner.[120] The committee's exact responsibilities and activities are usually unclear to an outside observer because firms do not generally publish meeting minutes, agendas, or the like.[121] Business organizations like the U.S. Chamber of Commerce advise that diversity committees should gather data and advise on policies that should be modified or eliminated to achieve diversity-related

---

119. *See, e.g.*, *Diversity Committee*, CRAVATH, SWAINE & MOORE LLP, [https://perma.cc/B8RS-5445] (last visited Jan. 9, 2023); O'Connor, *supra* note 114.

120. *See Diversity Committee*, *supra* note 119 (noting that 85% of firm diversity committees now have at least one partner).

121. *See, e.g.*, *id.*

goals.[122] As examples, the Chamber noted that adding employee benefits, creating affinity groups, and considering "more diverse" company events could fall in the ambit of the diversity committee.[123]

This summary suggests some of the shortcomings of diversity committees. Committees are so inherently and famously ineffective at rallying consensus that a U.S. spy agency in World War II trained agents undercover at German factories to create committees and refer as many matters to the committee as possible to slow production.[124] In the limited literature on the effectiveness of diversity committees, the clear theme is that they can work if they have authority to impose accountability for diversity goals.[125] Law firm diversity committees tend not to do that. Lacking both clear goals and authority and comprised mostly of the least senior and most racially isolated people at the firm, they cannot hope to achieve much.[126] Including at least one partner, as many firms do, is not the same thing as substantial buy-in from firm management. At best, it means that if there is something easy the firm can do to improve its diversity situation, there is a designated partner to hear about it.

Just because the committees are ineffective at their stated goal does not mean they are useless, however. Distilled down, diversity committees are popular because they constitute doing

---

122. *See* Sean Peek, *6 Essential Steps to Creating a Diversity, Equity & Inclusion Committee at Your Company*, U.S. CHAMBER OF COM. (Aug. 26, 2020), [https://perma.cc/B2Q6-SF9D].

123. *Id.*

124. *See* U.S. OFF. OF STRATEGIC SERVS., SIMPLE SABOTAGE FIELD MANUAL 28 (1944), [https://perma.cc/4UBF-4GMB] ("When possible, refer all matters to committees for 'further study and consideration.' Attempt to make the committees as large as possible—never less than five.").

125. Savita Kumra, *Busy Doing Nothing: An Exploration of the Disconnect Between Gender Equity Issues Faced by Large Law Firms in the United Kingdom and the Diversity Management Initiatives Devised to Address Them*, 83 FORDHAM L. REV. 2277, 2288 (2015) (citing studies of diversity management for the proposition that accountability is a key determinant of whether diversity initiatives succeed).

126. *See* CHRISTOPHER L. MEAZELL, THE BUSINESS OF CONTEMPORARY LAW PRACTICES 19 (2021) (describing many respondents to a survey saying that law firm diversity committees lacked "central focus," and "coordinated firm-wide" goals, and "lacked the authority to make a real difference").

something, but in a controlled, low-commitment way.[127] They offer the maximum ratio of public relations benefit to actual investment. The public relations benefit in question is not great in magnitude, but the investment is essentially zero. A firm can always point to the creation or continued existence of the committee to show that firm leadership takes the issue of diversity seriously. Indeed, most firms take exactly this tack. The typical firm has a webpage for their diversity committee noting its general task and at least some of its composition. It does not note any actual agenda or specific goals.[128]

To delve more deeply into the benefits of diversity committees, we can start with their low cost. Generally, firms do not hire anyone specifically to be on a diversity committee, and they do not pay anyone a bonus to serve on one.[129] Indeed, some Black employees in tech companies have criticized the widespread use of diversity committees, seeing it as extra unpaid work that inherently lessens their ability to compete with White workers not similarly burdened.[130]

An attractive aspect of diversity committees for firms is that they provide a tightly controlled outlet for complaints that might otherwise go to newspapers, blogs, or some other public venue that could potentially embarrass the firm.[131] Everyone on the committee works for the firm and has a vested interest in maintaining the firm's reputation. While many committee participants doubtless care about achieving equitable outcomes whenever possible, they are inherently limited in just how

127. *Cf.* Kumra, *supra* note 125, at 2298 (describing the motive of some firms as "positive company image" and noting that at those firms there is a greater likelihood that the gestures toward diversity are an "empty shell").

128. *See, e.g.*, *Diversity Committee*, *supra* note 119; *Diversity Committee*, SULLIVAN & CROMWELL, [https://perma.cc/75SH-782M] (last visited Jan. 9, 2023); *Diversity Overview*, GIBSON DUNN, [https://perma.cc/64QK-Q9DR] (last visited Jan. 9, 2023).

129. Staci Zaretsky, *25 Biglaw Firm Offers Bonus-Eligible Diversity & Inclusion Billable Hours*, ABOVE THE L. (Feb. 8, 2021, 1:48 PM), [https://perma.cc/TMW8-4PWL].

130. *See* Nitasha Tiku, *Tech Companies Are Asking Their Black Employee Groups to Fix Silicon Valley's Race Problem—Often for Free*, WASH. POST (June 26, 2020, 6:00 AM), [https://perma.cc/DFZ6-73PV].

131. *See, e.g.*, Patrick Dorrian, *Davis Polk Says Black Lawyer's Suit Defamatory, Wants Sanctions*, BLOOMBERG L. (Jan. 27, 2021, 9:41 AM), [https://perma.cc/ZXW3-QNCV] (discussing a lawsuit filed by a Black attorney alleging racial discrimination and retaliation for complaints about racial bias at a major law firm).

adversarial they can be without throwing their future career at the firm in jeopardy. At the same time, lawyers of color with a race-related grievance against the firm could do tremendous damage to the firm's image if they file a lawsuit instead of a diversity-committee complaint.[132] If having the diversity committee as an outlet results in even one fewer lawsuit against the firm, it will have been worth whatever marginal commitment of billable hours the firm has devoted to it.

Diversity committees provide everyone involved—associates and partners—with a line on their resume that could prove useful when eventually moving to another firm. Associates of color, facing many obstacles to forming bonds with partners, can take positions in the diversity committee, where there is likely to be at least one partner, and get some face-to-face time with firm leadership. Partners can claim important management experience by serving on a diversity committee. Everyone else at the firm benefits from the improved public image of the firm.

2. Affinity Groups

By all available evidence, affinity groups help associates of color, but do not make a decisive difference in promoting diversity. Affinity groups are voluntary associations of firm members sharing some characteristic.[133] They are not at all limited to racial affinity groups; some firms have groups based on veteran status, gender identity, etc.[134] They are distinct from diversity committees because their activities are less structured. While that does limit their potential to achieve specific goals, anecdotal evidence suggests affinity groups do help minority attorneys feel a sense of community.[135] For example, the affinity groups can organize events, group chats, and even email chains that provide a common space for members of the designated community to express themselves frankly.[136] Helping associates

132. *Id.*

133. *See* Mishell Parreno Taylor, *Today's Affinity Groups: Risks and Rewards*, SHRM (Oct. 11, 2019), [https://perma.cc/96YX-TA3U].

134. *Id.*

135. *Id.*

136. *Id.*

of color cope with law firm life is certainly an accomplishment and can indirectly lead to greater representation in partner ranks if it reduces the attrition rate of associates of color.[137]

One negative aspect of affinity groups from a racial-diversity standpoint is that they are still an outgrowth of the firm itself. They are intended to achieve the firm's goals.[138] Employment law attorneys frequently stress the need to align an affinity group's goals with the firm's.[139] It does not help firms if, for example, the affinity group suggests that members of the affinity group need additional support in order to succeed at the firm.[140] Another shortcoming of affinity groups is one shared with diversity committees: if they do not build bridges to partners, they are unlikely to directly lead to promotion of attorneys of color. Indeed, some White employees at major firms claim to feel discriminated against because they do not have an affinity group of their own.[141] If a White partner feels that way, then an associate's being in an affinity group could unfairly harm their prospects for advancement. While claims occasionally arise that the creation of affinity groups inherently discriminates against any set of individuals not represented in the groups, such claims have not fared well in court.[142]

3. Formal Mentorship Programs

The vast majority of firms have adopted formal mentoring programs, but these can fail to produce true mentorship. The problem arises from a misdiagnosis of why mentoring helps in the first place. Mentorship in its most basic form involves providing

137. Virginia G. Essandoh, *Why Law Firm Affinity Groups Are a Valuable Resource*, LAW.COM (Oct. 23, 2008, 12:00 AM), [https://perma.cc/AAU4-2KQ3].

138. *See* Sandra S. Yamate, *Affinity Groups in Large Law Firms: What to Consider*, AM. BAR. ASS'N: LITIG. NEWS, [https://perma.cc/7T8T-PXZ5] (last visited Jan. 24, 2023) (advising firms to be "careful in how the purpose and goals of [affinity] groups is expressed" and to "[b]e clear about the group's purpose from the firm's perspective").

139. *Id.*

140. *Id.*

141. *See* Anne-Marie Vercruysse Welch et al., *Legal Traps Associated with Affinity Groups*, 33 ABA J. LAB. & EMP. L. 267, 269 (2018) (describing a lawsuit at General Motors alleging precisely this kind of discrimination).

142. *See id.* ("Articulating and following a process to treat affinity groups equally will greatly reduce [the] risk [of a lawsuit].").

advice and guidance to help newcomers gain expertise in the most efficient manner possible.[143] While receiving information from a partner doubtless has some value to associates, the major benefit in a law firm is differentiation from the mass of associates in the eyes of the partner.[144] In the literature on mentorship, this is often described as the mentor becoming a "champion" for the mentee.[145] If the idea is to give every associate a "champion," there are some obvious game theory problems to firm mentorship programs. First, everyone having a champion logically means no one has a champion. That might seem like a good, if roundabout, way to weaken the advantage of White associates in forming bonds with firm leadership. However, the reality is that all officially declared champions are not equally dedicated to the task of promoting their mentee. One survey found that only 27% of firm mentors described themselves as actively advocating for their mentees.[146] Another study found that informal mentorship is more effective than formal, and informal mentorship is precisely what one would expect to arise from the commonalities between White associates and partners.[147] This is particularly noteworthy because partners are more resistant to mentoring than associates, presumably due in part to the greater tradeoff between billable hours and mentoring time.[148] Informal mentorship will be voluntary on the partner's part, whereas formal mentorship is less voluntary.

A more subtle problem with firm mentorship is that it does not bridge the divide between associates of color and White

143. *See* Rafael X. Zahralddin-Aravena, *The Evolution of Mentorship in Legal Professional Development*, AM. BAR. ASS'N: BUS. L. SECTION (Mar. 19, 2020), [https://perma.cc/UA3G-MUGK] (describing firm mentorship programs).

144. *See* Levs, *supra* note 78 (describing a usual firm process where an associate is nominated by the partner they work for and the key role played by personality).

145. *See, e.g.*, Michael L. Nguyen & Apoorva J. Patel, *Strategies for Effective Mentorship and Sponsorship*, AM. BAR. ASS'N, [https://perma.cc/3EK9-J379] (last visited Jan. 24, 2023).

146. *See* Zahralddin-Aravena, *supra* note 143.

147. *See* Sonia R. Russo, *Be the Change: How Mentoring Can Improve Diversity in the Legal Profession*, LAW PRACTICE TODAY (July 14, 2016), [https://perma.cc/537J-R8SH].

148. *See* Bruce Epstein, *Mentoring in Law Firms: A Survey of Current Practices*, LINKEDIN (Sept. 27, 2017), [https://perma.cc/VK8U-R3NN] ("A small percentage of junior associates resist mentoring; a larger percentage of partners (perhaps 20%) also resist being mentors.").

partners. One study found 71% of firm mentors are the same race or gender as their mentees.[149] Superficially, this arrangement makes sense—if we want the partner and associate to forge a bond, they may be more likely to do so if they are of the same race. However, one can logically surmise that partners of color are not the roadblock to the advancement of associates of color in the first place. If a mentorship program is not bridging divides with White partners, it has limited ability to help associates of color become partners. This tracks with anecdotal complaints by associates of color about firm mentorship programs, namely that Black associates often have Black mentors.[150] If the Black mentors have less power in the firm, their mentorship will not be as valuable, and they will not be able to champion their mentees in the way a White partner would.[151] Further, since there are few partners of color and relatively more associates of color,[152] one would expect the ratio of mentees to mentors to be higher among partners of color. This would dilute whatever personal connection is made with the mentor.

Some of these factors would apply in any professional context, but the intense competitiveness and lack of alternative channels of competition for associates creates a measurable problem specific to law. Recall from earlier in this Article that the medical community has a far less severe racial diversity problem.[153] Doctors also do not have a disparity in mentorship between clinicians of different races.[154]

Given that the younger ranks of both law and medicine are composed of ambitious, hard-working, smart people, the diverging outcomes on racial diversity generally, and mentorships in particular, suggest there is something about how the profession

---

149. Zahralddin-Aravena, *supra* note 143.

150. *See* Natalie Runyon, *Why Inclusion Matters: A Story About the Different Experiences Between White and Black Attorneys*, THOMSON REUTERS (Feb. 6, 2020), [https://perma.cc/SXV6-G8HA].

151. *Id.*

152. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 7.

153. *See supra* note 16 and accompanying text.

154. *See* Mitchell D. Feldman et al., *Does Mentoring Matter: Results from a Survey of Faculty Mentees at a Large Health Sciences University*, 15 MED. EDUC. ONLINE 5063, at 1 (2010) (describing a study finding no difference in having a mentor based on gender or ethnicity at one of the largest and most comprehensive mentoring programs in medicine at the University of California, San Francisco).

works causing the difference. It would take a more rigorous empirical analysis to authoritatively identify the causes, but I speculate that the true root cause is how doctors and lawyers compete for career advancement. In medicine, good and bad outcomes are much easier to discern and much easier for superiors to monitor. Deciding whether, say, Patient *X*'s foot surgery went well can be answered quite authoritatively—can she walk? Is she in constant pain? If so, what exactly did the surgeon do? A commonality of experience and knowledge allows one doctor to easily pass judgment on another. Because judgment can be passed easily and frequently, a supervising doctor can easily measure a new doctor's performance on the merits.[155] Socializing with the supervising doctor, while advisable, is ancillary to substantive performance.

By contrast, as we have discussed, measuring lawyer performance is difficult. Aside from obvious grammatical problems or completely missing well-known case law, it is hard to know when a memo "fails" for reasons identifiable to an outside observer. Memos generally analyze unique combinations of facts and law.[156] Unless a partner possesses encyclopedic knowledge of that area of law *and* keeps up to date with all the published and unpublished opinions in the area, he has no way of knowing whether the associate missed something important. Winning or losing an individual case is often beyond the control of an individual associate, so the sort of binary outcomes typical in medicine are not available to judge associate performance at large law firms.

4. Community Outreach

Community outreach programs at major law firms, while commendable, barely count as an attempt at solving the racial diversity shortfall. Instead, they represent an entirely natural reaction to a difficult problem: attempting to solve a different

---

155. *Id.*

156. *Memorandum*, CORNELL L. SCH. LEGAL INFO. INST. (Sept. 2020), [https://perma.cc/89EX-C9VB].

problem, even if it is something that has already been solved.[157] It is reassuring to be able to do *something* tangible, both because it is a signal of commitment and because it provides a false sense that the problem is being addressed in some way. A more cynical view would be that flashy solutions that are not strictly related to the real problem allow public relations professionals to convince outsiders that the problem is well in hand.

In the ubiquitous "diversity" section of major firm websites, many firms advertise their charitable work with local communities in which they promote legal education or seek minority summer associates from local schools, etc.[158] It is certainly not a bad thing for a firm to give to charity or have outreach efforts to improve recruitment, but these sorts of programs should not be confused with a systematic effort to fix the problem. The diversity problem among associates is nowhere near as bad as with equity partners.[159] Similarly, diversity at law schools, while not representative of the population as a whole, is not anywhere near as skewed as in firms.[160] Much like diversity committees, these initiatives are easy for the firms. They require only slightly altering summer associate classes or issuing potentially tax-deductible charitable donations.

5. The Mansfield Rule

The Mansfield Rule is a well-intentioned step that will almost certainly fail to solve the problem of racial diversity in law firms. The rule is simple: a law firm must "affirmatively consider[]" at least 30% women, lawyers of color, LGBTQ+ lawyers, and lawyers with disabilities for leadership- and governance-committee positions, equity partnerships, "formal

157. Examples include professional sports teams that heavily advertised disinfecting of seats at the height of the coronavirus pandemic despite the fact that the primary transmission mechanism was in-person transmission.

158. *See, e.g.*, *Diversity, Equity, & Inclusion*, *supra* note 5 (noting a summer associate recruitment program for minority law students and a scholarship program).

159. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 8.

160. *Compare id.*, *with Law School Enrollment by Race & Ethnicity (2019)*, ENJURIS, [https://perma.cc/86ME-R7E3] (last visited Jan. 9, 2023) (showing racial-ethnic minority law students comprising around 30-46% of all law students in large states).

client pitch opportunities," and senior lateral positions.[161] The rule was created by a group called the Diversity Lab, which certifies firms that follow the rule.[162] The rule is rooted in the theory that promotion committees can ameliorate their conscious or subconscious biases by making their promotion pool more diverse.[163] As of this writing, 117 firms, including many of the biggest and most well-known, are Mansfield-certified.[164] According to the Diversity Lab, 76% of participating firms said their equity-partner promotions pool was more diverse.[165]

One obvious problem with the Mansfield Rule is that firms can meet it very easily without improving racial diversity. Obviously, simply "affirmatively considering" candidates of color does not ensure they are selected. But more subtly, the rule requires 30% of candidates to fall into one of several underrepresented categories. For example, one way to meet the Mansfield Rule is to have 30% of the promotion pool be White women. Women currently represent 22% of equity partners,[166] so firms only need to be slightly more inclusive on the gender of promotion pools to meet the rule without changing racial representation a jot.

The Mansfield Rule was based on the Rooney Rule in the National Football League (NFL),[167] so it is worth a brief digression into the NFL to predict potential long-term results of the Mansfield Rule. The Rooney Rule requires teams with a head coaching vacancy to interview at least one racial minority candidate.[168] The rule has been in place since 2003,[169] so it may be instructive to consider the extent to which the rule has worked. There was a significant bump in the racial diversity of head

161. *An Open Letter from the 2020-2021 Mansfield Law Firms' Chairs & Managing Partners*, DIVERSITY LAB, [https://perma.cc/Y3MU-ABJH] (last visited Jan. 24, 2023).

162. *Id.*

163. *Id.*

164. *Id.*

165. *Id.*

166. Meghan Tribe, *Women Law Partners Weigh In on How to Close the Pay Equity Gap*, BLOOMBERG L. (June 14, 2022, 3:45 AM), [https://perma.cc/7TPU-AQSY].

167. Press Release, Diversity Lab, 44 Law Firms Pilot Version of Rooney Rule to Boost Diversity in Leadership Roles (Sept. 25, 2017), [https://perma.cc/G837-CU9A].

168. Press Release, Nat'l Football League, NFL Expands Rooney Rule Requirements to Strengthen Diversity (Dec. 12, 2018), [https://perma.cc/5JHQ-7YUQ].

169. *Id.*

coaches from about 2003 to 2016.[170] The share of games coached by non-White head coaches increased from about 9% to a peak of 25% in 2011.[171] However, as of 2020, non-White head coaches coached roughly the same percentage of games as before the Rooney Rule began.[172]

Analysis of the Rooney Rule has raised a number of problems that will likely also limit the effect of the Mansfield Rule. The Rooney Rule appears to have increased representation temporarily, possibly by allowing individual applicants the opportunity to make an impression on team owners who otherwise would have viewed them through a stereotypical lens. That is precisely what the Mansfield Rule hopes to accomplish. However, in college football, Black candidates who were subsequently hired after the Rooney Rule was put in place were given fewer chances to overcome disappointing seasons.[173] NFL owners also increasingly made up their minds before the hiring process began so that the nominal hiring process was merely for show.[174] The fundamental problem of the Rooney Rule is the illusion that one simple, mechanistic requirement can overcome wider institutional problems.

The most important difference between the Mansfield and Rooney Rules is that the former is voluntary and the latter is not.[175] Paradoxically, this suggests the Rooney Rule should be *more* effective. Logically, the most forward-leaning firms would be the most likely to seek Mansfield certification because the "price" of doing so (i.e., how much the firm actually needs to change) would be lower. We should then logically expect that many firms will successfully gain Mansfield certification without

170. RICHARD E. LAPCHICK, UNIV. OF CENT. FLA., THE 2022 RACIAL AND GENDER REPORT CARD: NATIONAL FOOTBALL LEAGUE 47 (Asia Ervin et al. eds., 2022), [https://perma.cc/KQ4V-V7ZJ].

171. *Id.*

172. Neil Paine, *The Rooney Rule Isn't Working Anymore*, FIVETHIRTYEIGHT (Jan. 14, 2020, 3:17 PM), [https://perma.cc/J8UK-NFD2].

173. *Id.*

174. *Id.*

175. *See* Press Release, *supra* note 168; Lenore Pearlstein, *Mansfield Rule for Law Firms Builds on the Rooney Rule*, INSIGHT INTO DIVERSITY (June 25, 2020), [https://perma.cc/RZV4-XZQ4].

changing much, and the firms we do not hear about where a Mansfield rule would be more helpful will also change nothing.

In addition to the practical problems identified above, I submit that the Mansfield Rule has not and will not work because it has misidentified the key drivers of underrepresentation at law firms. The Mansfield Rule envisions a world in which partners sitting on promotion committees consciously or unconsciously decide not to interview an otherwise good candidate because they are a racial minority. If this were the case, the problem of racial diversity in law firms should be no different from the problem of racial diversity in, say, medicine. In either case, the issue would be overcoming bias. However, we know that racial diversity is worse in law firms,[176] which strongly suggests something other than (or at least, in addition to) bias is afoot. I submit that incentives facing firm lawyers fill that gap and better explain the data. The Mansfield Rule, for example, has no explanatory power if a firm interviews 30% racial minority candidates, virtually none of whom had the same mentorship opportunities as their White peers, and ultimately promotes none of the minority candidates. However, we can immediately understand that structural incentive problems explain the failure to promote minority candidates.

G. How Morally Culpable Are Firms for Diversity Failures?

I have been critical of law firm efforts to promote racial diversity, but a fair rejoinder might be that ineffective options may still be the best available and better than nothing. More dramatic action to bring about racial diversity at a particular firm might be possible, but if law firm competition has produced a true multipolar trap, a firm taking on the costs of dramatic action would not survive the competitive effect. Surely, it is utopian to expect a firm to do something noble that would result in its demise.

Here it is important to distinguish between the concept of a multipolar trap and the theoretically *perfect* multipolar trap. The

---

176. Debra Chopp, *Addressing Cultural Bias in the Legal Profession*, 41 N.Y.U. REV. L. & SOC. CHANGE 364, 364-65 (2017).

general notion of a multipolar trap is that important values must be sacrificed in order to survive a high level of competition.[177] In a perfect multipolar trap, the smallest shred of value contrary to the factor being optimized results in complete destruction of the firm that attempts it.

I contend that firms and the people in them are in a multipolar trap, which limits but does not completely eliminate their ability to incorporate values other than competition. Obviously, people who work at firms still do have families and personal lives, even if their work-life balance is generally poor. However, the multipolar trap limits the ability of firms to uphold values, and because enhancing racial diversity directly contradicts the incentives of individuals within the firms, firms tend not to enhance racial diversity at senior levels.

The exclusion of Jewish lawyers from top law firms in the early twentieth century reveals how competition and discrimination did not always work the way they currently do. White Anglo-Saxon Protestants dominated elite New York law firms in the late nineteenth and early twentieth centuries.[178] By the latter half of the twentieth century, Jewish attorneys were overrepresented at top New York law firms.[179] Note that the legal field was nowhere near as competitive in the early twentieth century—there were far fewer law firms in a far poorer economy than exists today.[180] The racist preferences of elite law firms could only be indulged because the lack of competition allowed space for "values," even bad values like antisemitism.

I believe the opposite is happening now. Most law firm attorneys have at least a weak preference for racial justice,[181] but competition has reached a pitch where it takes a meaningful preference for racial justice to be willing to bear the short-term costs of increasing diversity. The incentive structure of law firms

177. *See* Alexander, *supra* note 23.

178. *See* Eli Wald, *Jewish Lawyers and the U.S. Legal Profession: The End of the Affair?*, 36 TOURO L. REV. 299, 308 (2020).

179. *Id.* at 308-11.

180. *See* Philip Gaines, *The "True Lawyer" in America: Discursive Construction of the Legal Profession in the Nineteenth Century*, 45 AM. J. LEGAL HIST. 132, 132-34 (2001); James Willard Hurst, *Lawyers in American Society*, 50 MARQ. L. REV. 594, 594, 596-97, 600-01, 606 (1967).

181. Douglass et al., *supra* note 4, at 53, 62-63.

separates the short-term costs from the long-term benefits, and so individual attorneys do not take on the sacrifices that would be necessary to increase racial diversity.

Whether law firms and their employees should face moral opprobrium for refusing to make these sacrifices is a subjective question outside the scope of this Article. I think in such situations, particularly good people can do the right thing—e.g., a senior partner can genuinely make a special effort to include associates of color in client interactions. It is also understandable that ordinary people take the easy route. For instance, a senior White partner might be glad her firm has community outreach programs but continues to mainly mentor White associates. To make real progress on an issue like racial diversity requires aligning incentives at least to the extent that an ordinary person feels comfortable taking on the necessary sacrifice to produce the outcome we all want.

II. SOLUTIONS

Understanding the multipolar trap that has forestalled racial diversity at law firms suggests an entirely different set of solutions than we typically consider. If one thinks that individual racial prejudice is what stops the advancement of attorneys of color at firms, then some combination of specific education of attorneys and general progress in society writ large will eventually solve the problem. However, if my contention in this Article is correct and racial diversity is being slowed by a multipolar trap, increasing diversity will require changing the incentives of firm attorneys. The multipolar trap by its very definition suggests that individual firms cannot escape the trap on their own—the level of competition is so high that even a short-term disadvantage could be fatal for the careers of the partners involved. The only way to disarm a multipolar trap is from the outside, whether through government regulation, ABA-driven regulation, or third-party pressure.

A. Why is Racial Diversity Among Partners Important?

We have taken it as given to this point in the Article that racial diversity is something worth some level of sacrifice to achieve, but to determine what level of intervention in the traditional firm model is justifiable, we should have some understanding of the benefits to be gained through intervention. It is worth asking at the outset: why should we want there to be a legal-profession-specific solution? One could argue that economic inequality between races is a society-wide problem. Not only is it impossible for the legal profession to solve this problem on its own, but attempting to do so uses up resources that could be directed against solvable problems. Perhaps law firms use affinity groups, diversity committees, and the like because there is no option that is as effective per dollar of cost. Here, it is crucial to recall that the lack of racial diversity in the legal world is not common to all professions.[182] While racial injustice exists outside the legal world, it is worse at law firms because of identifiable dynamics and incentives.[183] Some kind of coordination is necessary to change the terms of competition to allow escape from the trap.[184] Even that would not guarantee fully equitable representation of minorities in law firms, but it would at least enable the profession to meet the low threshold set by other professions.

There is, of course, a moral case to be made for increasing diversity, but morality-based arguments have not proven effective. The morality case for diversity, distilled down, is that racial minorities have often in the past been excluded from both the practice and the protection of law, much to their detriment.[185] This anti-competitive exclusion benefitted White attorneys, both directly and because the increased resource base created better circumstances for White children, who were then better prepared

182. Allison E. Laffey & Allison Ng, *Diversity and Inclusion in the Law: Challenges and Initiatives*, AM. BAR ASS'N (May 2, 2018), [https://perma.cc/4XR7-WL6M]; Chopp, *supra* note 176, at 364-65.

183. *See supra* Section I.D.

184. *See* Alexander, *supra* note 23.

185. *See, e.g.*, Kasper Lippert-Rasmussen, *Affirmative Action, Historical Injustice, and the Concept of Beneficiaries*, 25 J. POL. PHIL., 72, 72-74 (2017) (describing the historical injustice-based arguments for affirmative action and increasing racial diversity).

to compete.[186] Requiring some level of sacrifice by current partners, who are overwhelmingly White, to fix the situation has some basic moral appeal. But moral arguments are inherently limited and subjective. It is easy to imagine how, say, a White associate from an impoverished background would feel about the argument that associates of color should be promoted ahead of her. My point is not that those feelings justify doing nothing about diversity, but that the persuasiveness of moral arguments will depend on the vantage point of the listener, and as such may not carry the day in arguing for racial diversity. Of course, as discussed above, firms can also defuse the morality argument by agreeing with it, but then not acting in meaningful ways to reduce diversity.

Many scholars have argued that diversity is its own reward, that businesses with diverse leadership outperform businesses with non-diverse leadership.[187] However, there are no detailed studies of this phenomenon in law. There are also so few major firms with racially representative equity partners that a detailed study controlling for confounding factors would be difficult.[188] As discussed earlier, there is an important distinction between long-term benefits and short-term profits. Presumably, if firms expected a significant short-term profit increase from increasing diversity among partners, more firms would have done it, especially given the high level of competition between firms. One can certainly imagine long-term benefits at the firm level to more racial diversity—different experiences leading to more creativity, appeal to a broader range of clients, etc. If partners do not capture that gain, however, it plays a much smaller role in decision-making.

Putting aside arguments that increasing diversity is a moral necessity or an obvious financial win, we can offer a third, less-explored rationale: increasing diversity would help the legal profession as a whole, ultimately redounding to the benefit of firms and the lawyers within them. Increasing diversity would

186. *Id.* at 72-73.

187. *See* Robin J. Ely & David A. Thomas, *Getting Serious About Diversity: Enough Already with the Business Case*, HARV. BUS. REV., Nov.-Dec. 2020, at 116, 118 (describing the historical and current "business case" for diversity).

188. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 7-8, 10, 13-15,17-19.

likely have the long-term benefits discussed above.[189] If most or all major firms are doing so at roughly the same time, firms will not be able to gain much of a competitive advantage by dragging their heels. The profession as a whole would of course gain a public relations win, but it would also gain an advantage over professions that are at least partial substitutes for law firms. This would include accounting firms, auditors, consultants, lobbyists, and other entities that can offer strategic advice involving law. While those substitute professions cannot provide all legal services the way a firm can, they can eat into the profit margins of firms by taking on quasi-legal issues that a firm could handle, like internal investigations.

There are also real risks to the profession as a whole from failing to diversify. The practice of law depends on barriers to entry—bar admissions, law school accreditations, etc. These barriers are sufficiently aged and ubiquitous that we lawyers take them for granted, but there are substantial and growing economic forces that could undermine them. Imagine if a behemoth like Walmart or Amazon decided there was considerable profit in, say, providing simple or large-volume legal services for things like wills, contesting parking tickets in court, or drafting employment contracts.[190] Consider that excluding major companies from providing important services at a low price to indigent constituents would not be a popular position for state legislators eternally bereft of campaign funds. A lobbyist for those companies would not hesitate to use the lack of diversity at law firms as a talking point to justify reducing or eliminating barriers to entry. That point can be further emphasized by noting the dramatic shortage of legal services—reportedly near 80% of the public does not have access to a lawyer.[191] By contrast, the more law firms are seen to be positive factors in the legal world, the easier it will be to maintain a pro-licensure and pro-regulation coalition. Further, if law firms start losing the battle to exclude companies from the marketplace for legal representation, the

189. O'Connor, *supra* note 114.

190. *See, e.g.*, *Appeal Parking Tickets in Any City*, DONOTPAY, [https://perma.cc/HJ3G-XDN2] (last visited Jan. 9, 2023).

191. Mary Juetten, *How Can We Reform Legal Education? Try Spotlighting the Outcomes*, ABA J. (Nov. 9, 2018), [https://perma.cc/6JFZ-LWHC].

multipolar trap would likely grow more intense. Less business divided over the same number of attorneys is a recipe for increasing competition, which could exacerbate the multipolar trap hindering racial diversity in a downward spiral.

B. How Should We Measure Success?

Before discussing what a good specific goal for law firm diversity would be, it is worth asking why a specific goal is important in the first place. The answer is simple: accountability. As discussed above, firms currently suffer few, if any, consequences for their lack of results on diversity.[192] So long as they acknowledge diversity as a worthy goal and have diversity committees and affinity groups, they need not worry about their lack of diversity in senior roles. The first step in holding firms accountable for diversity is being able to discern true commitment to diversity.

The academic debate over how to measure success in racial diversity in law is reminiscent of the often-mocked tendencies of Soviet revolutionaries to endlessly argue over what the precise contours of society would be once they had eradicated capitalism.[193] Would there be no government? An all-powerful, all-benevolent government? These questions were silly because the revolutionaries had so many immediate problems—corruption, famine, crippling poverty—and were so far from addressing any of them.[194] Similarly, we are currently so far from addressing racial disparities at law firms that debates over measuring success are almost entirely theoretical. There is much debate among diversity experts about the difference between "[f]ormal diversity" (i.e., the distribution of attorneys by race roughly matches the distribution of race in the general population) and "[s]ubstantive diversity" (i.e., equality across a broader range of measures, such as compensation rates).[195]

---

192. *See supra* Section I.G.

193. *See Revelations from the Russian Archives*, LIBR. OF CONG., [https://perma.cc/9L76-XAH8] (last visited Jan. 9, 2023).

194. *Id.*

195. *See* Jason P. Nance & Paul E. Madsen, *An Empirical Analysis of Diversity in the Legal Profession*, 47 CONN. L. REV. 271, 274 (2014).

I believe this debate is largely abstract and, given the current extent of the diversity problem, premature. Only a truly naïve and obtuse formal measure of diversity would actually create problems, such as if one examined a firm composed exclusively of White attorneys and minority secretaries and concluded there was no diversity problem if overall firm employment reflected underlying population distribution. Of course, if "formal diversity" is attained someday, we scholars should confirm that firms have not manipulated diversity numbers in subtler ways. However, at this point, mere "formal diversity" would be a major improvement.

One simple way of determining success, and the one I endorse here, is that the rate of improvement in the share of racial minorities among equity partners increases to the point that firms would have roughly representative numbers of minorities in our lifetimes. That is not overly ambitious. Currently, racial minorities represent 9.1% of firm equity partners and about 40% of the U.S. population, so meeting this goal would require tripling the share of racial minorities among equity partners.[196] Thus, an improvement of 1% each year would solve the problem in thirty years. Within the 100 highest-grossing U.S. law firms, there were 21,258 equity partners in 2020.[197] In a firm of about 200 equity partners, going from twenty to eighty partners of color in thirty years requires adding two equity partners of color per year. Ambitious, but not unreasonable—the prestigious law firm Cravath, Swaine & Moore, with about 100 equity partners, promoted two Black partners in 2020 (albeit, the second and third in the firm's history after their first left in 2017).[198]

A results-based goal is critical to changing firm incentives, but it is also more effective than advocating a single set of best practices that every firm should use. While I do think there are generally better practices that firms should consider, which I will

196. VAULT & MINORITY CORP. COUNS. ASS'N, *supra* note 8, at 16; Nicholas Jones et al., *Improved Race and Ethnicity Measures Reveal U.S. Population Is Much More Multiracial | 2020 Census Illuminates Racial and Ethnic Composition of the Country*, U.S CENSUS BUREAU (Aug. 12, 2021), [https://perma.cc/G3D6-48TC].

197. *Total Number of Equity Partners Working at the Leading U.S. Law Firms 2016 to 2020*, STATISTA, [https://perma.cc/76TJ-QDBF] (last visited Jan. 9, 2023).

198. Staci Zaretsky, *Cravath Makes History in Diversity with Its New Partnership Class*, ABOVE THE L. (Oct. 30, 2020), [https://perma.cc/L52X-ZCPP].

discuss in more depth below, firm structure and culture can vary sufficiently that solutions will vary in success depending on the firm. There doubtless are law firms with an exemplary commitment to diversity committees that have made significant progress. If that is the case, and those firms are advancing quickly toward representation in their equity partners, then they should not tinker with a successful formula. The only universal in which I am confident is that firms will perform better on diversity measures if it is in their interests to do so. The weaker the incentives for diversity, the more reliant diversity is on partners willing to sacrifice to make it happen, the less firms will improve.

C. What Can Individual Firms Do?

The multipolar trap law firms find themselves in and their incentives structure all but precludes them from more aggressively pursuing racial diversity. Equity partners have not chosen to make the short-term sacrifice to increase racial diversity. They have settled on a basic suite of cheap pro-diversity measures and left it at that. The logic of the efficient market hypothesis suggests there are no easy and effective pro-diversity measures that firms could adopt at this point.[199] If there were, one of the thousands of other law firms in the United States or around the world would have figured it out, and then most of the others would copy the solution.

Still, the experience of antisemitism in law suggests that there could be widespread cultural factors that are producing a non-efficient outcome—i.e., there could be a good solution that law firms simply refuse to adopt, just as they refused to hire Jewish attorneys for no good reason. A skeptic might note that U.S. law firms in the late 1800s were much more homogeneous and parochial than today's multicultural, worldwide law firms.[200] What self-defeating cultural assumptions could be so widespread

199. *See* Eugene F. Fama, *Efficient Capital Markets: A Review of Theory and Empirical Work*, 25 J. FIN. 383, 416 (1970) (providing the modern formulation of the efficient markets hypothesis in which prices reflect all available information).

200. LAWRENCE M. FRIEDMAN, AMERICAN LAW IN THE TWENTIETH CENTURY 29 (2002).

in today's firms? The search for such cultural assumptions starts with examining commonalities across firms.

While such assumptions must, by their nature, be subtle, the one possibly flawed assumption I can identify that most law firms have in common is a devotion to the partner/associate dichotomy. The partner/associate setup creates a sharp discontinuity of incentives. Associates generally do not have a stake in the firm's profits until suddenly that is the entire source of their compensation when they become partner.[201] If instead law firms created more intermediate positions and transitioned more slowly to equity stakes, the gatekeeping between associate and partner would be less severe. Associates of color could slowly take on more of a client-facing role, and White associates would have less of an incentive to capitalize on their shared culture with White partners. Some firms have, perhaps unwittingly, been taking the first steps in the process of differentiating positions by creating more senior associates and non-equity partners.[202] These initiatives may be aimed at other goals, like retention of talented associates. They may even be pernicious in some cases, allowing the promotion of minority associates to non-partner roles that still count toward diversity statistics. However, filling out more of a spectrum of seniority at law firms would allow intermediate steps that could ultimately produce more racially diverse equity partners. To reach that end, firms would have to grant some equity to attorneys at these sub-partner tiers.

While it is difficult to know why firms have generally not attempted such a solution, we can speculate that it would require a major departure from the traditional firm model and potentially dilute the power or income of equity partners. There are ways to structure such a transition that could ease concerns—for example, doling out equity to the new sub-equity partner attorneys as partners retired rather than transferring it pro rata from existing equity partners. I suspect that, again, the benefits of the change are dissipated across the legal profession while the "costs" (e.g.,

201. Sharon Miki, *A Guide to Law Firm Partnership Structures*, CLIO, [https://perma.cc/CFE7-ZAHF] (last visited Jan. 8, 2023).

202. Lateral Link, *Evolving Partnership Economics: The Equity and Non-Equity Models Are Starting to Blur*, ABOVE THE L. (Aug. 13, 2021, 1:13 PM), [https://perma.cc/GV62-8QHM].

learning to manage a different system, concerns that clients will feel slighted at working with non-partner-level attorneys) are concentrated on the existing equity partners. Given that few firms are under any sort of pressure to do more than the *de rigeur* affinity groups and diversity committees, it is understandable that equity partners have not adopted anything that would significantly change the traditional firm model.

D. Coordination to Promote Racial Diversity

Coordination is the key to solving a multipolar trap.[203] By establishing rules limiting the terms of competition, the multipolar trap can be downgraded to productive competition instead of an all-consuming maelstrom. Coordination can be achieved either through an external authority's intervention or cooperation between the competing entities. Both options could plausibly promote racial diversity among the upper echelons of law firms, though external intervention seems more likely to succeed.

1. Coordination Through Cooperation

Law firms have already come together to create rule-enforcing entities—state- and national-level bar associations. Those associations recommend changes in laws governing lawyers to state governments and could theoretically do so in the case of racial diversity.[204] However, the rules usually enforced by those associations relate to relatively uncontroversial ethics and professionalism practices.[205] Bar associations enforcing rules relating to promoting racial diversity would be a vast expansion of their powers.

Some states are trying to improve racial diversity in the legal profession by requiring their state bars to take action, but those efforts are often hopelessly toothless and vague. For example,

---

203. *See* Alexander, *supra* note 23.

204. *See Governmental and Legislative Work*, AM. BAR ASS'N, [https://perma.cc/7974-5SW6] (last visited Jan. 9, 2023).

205. *See, e.g.*, *Actions on Rule Changes and Legal Ethics Opinions*, VA. STATE BAR, [https://perma.cc/8Y9D-XMSW] (last visited Jan. 9, 2023) (listing recent rule changes, all of which seem fairly uncontroversial).

California now requires its state bar association to produce a report every other year on how it is promoting racial diversity in the legal profession.[206] There have now been two reports submitted by the state bar association under this system. The first, issued on March 15, 2019, was laughably bureaucratic, passive, and wishy-washy.[207] The most prominent action highlighted was the adoption of a new mission statement by the state bar board of trustees, in which they declared "support for greater access to, and inclusion in, the legal system."[208] After that, the report indicates the state bar intended to "[s]erve as a data repository" on diversity-related information, "[c]onvene stakeholders to discuss emerging issues," and "[r]ecommend, incubate and/or pilot promising programs that are based on data and have the potential to scale throughout the state."[209]

Two years later, the 2021 report noted as an accomplishment that it had published a report card on the diversity of California's legal profession.[210] That report card does not grade firms. It contains the same findings the ABA has been putting out for at least a decade—poor racial diversity at law firms, particularly at the senior levels.[211] The state bar association effectively bought itself two years of inaction by reporting what everyone already knew. The 2021 report also noted that it had indeed followed through on its promise to launch "Diversity Summits" to discuss next steps.[212]

The 2021 report notes that the *next* two-year report will describe "aspirational" goals for law firms to "set and publicly commit to measurable diversity, equity, and inclusion goals."[213]

---

206. CAL. BUS. & PROF. CODE § 6001.3(c) (West 2020).

207. *See* OFF. OF ACCESS & INCLUSION, STATE BAR OF CAL., DIVERSITY & INCLUSION PLAN: 2019-2020 BIENNIAL REPORT TO THE LEGISLATURE 1 (2019), [https://perma.cc/4FV7-XTMV] (from the Executive Summary: "Pursuant to its Strategic Plan objectives the State Bar intends to: Serve as a data repository, research institution, and technical assistance provider" and "[c]onvene stakeholders to discuss emerging issues, best practices, and data collection").

208. *Id.*

209. *Id.*

210. *See* CAROLINA ALMARANTE ET AL., STATE BAR OF CAL., REPORT CARD ON THE DIVERSITY OF CALIFORNIA'S LEGAL PROFESSION (2020), [https://perma.cc/W27M-ZB3Z].

211. *Id.* at 13.

212. *See* OFF. OF ACCESS & INCLUSION, *supra* note 207, at 27.

213. *Id.* at 17.

In the "medium term, the State Bar intends to incorporate accountability measures to ensure that employers who are certified as [diversity, equity, and inclusion] leaders demonstrate results, not just intentions."[214] Tune in again in 2023 to see if the state bar has decided, after six years of study, how it *might* measure diversity at law firms! To be fair, the California diversity reports do note worthwhile programs on issues like attorney discipline and the bar exam, but they conspicuously fail to offer solutions about law firm diversity.[215] The reports generally seem to be the bare minimum the state bar association could get away with submitting to satisfy their statutory mandate—right down to the graphics clearly recycled between the diversity report card and the state-mandated report.[216]

The California experience suggests state bar associations will not deliver results, but it is useful to consider why. A simplistic, paranoid theory would be that firms dominate the legal profession's economics and consequently state bar associations do not want to cause problems for them. But this speculative theory does not hold up to scrutiny. First, while the trustees of the state bar are largely firm lawyers, the authors of the reports in

214. *Id.* at 5.

215. *See, e.g.*, *id.* at 15 n.3.

216. *Compare* OFF. OF ACCESS & INCLUSION, *supra* note 207, at 11, *with* ALMARANTE ET AL., *supra* note 210, at 8. The cynical observer would also note that the 2019 report was twenty pages long with twenty-eight pages of attachments. The 2021 report was forty-one pages long with 132 pages of attachments. A prediction for 2023: look for ever-expanding attachments and little change in content in the main body. The main body of the 2021 report recapitulated the findings of the diversity report card, barely avoiding outright plagiarism. For example, see this passage from page eight of the report card:

> Although the majority of attorneys, both [W]hite and of color, work in the private sector, [W]hite, Asian, Middle Eastern/North African, and attorneys categorized as "Other" are more likely to do so than Black/African American and Hispanic/Latino attorneys. Black/African American attorneys are less likely to work in law firms than all other racial/ethnic groups.

STATE BAR OF CAL., DIVERSITY, EQUITY, & INCLUSION PLAN: 2021-2022, BIENNIAL REPORT TO THE LEGISLATURE 8 (2021). Compare that to page eleven of the 2021 report:

> Although the majority of attorneys work in the private sector, [W]hite, Asian, Middle Eastern/North African, and attorneys categorized as "Other" are more likely to do so than Black and Latino attorneys. Black attorneys are less likely to work in law firms than all other racial/ethnic groups.

STATE BAR OF CAL., *supra*, at 11.

question are not.[217] More fundamentally, there is no reason to believe the major law firms differ tremendously in levels of diversity. All have similar compensation schemes and all are similarly stuck in the multipolar trap. Even if the firms secretly could influence the authors of these reports, what harm would come to any specific firm by calling for more specific action by *all* law firms?

There is no practical way to know the answer, but I speculate that the state bar associations, law firm management, and the authors of these reports have many interests in common, all of which militate against strong criticism of firm diversity. Legal non-profits rely on law firms and their employees for both financial support and pro-bono assistance.[218] While the authors of these reports may not be firm attorneys, leadership of state bar associations generally have strong personal ties to law firms.[219] Firm management would have little incentive to put pressure on their firm to change diversity practices even if the resulting criticism would apply to their competitors.

The foregoing discussion does not mean that firms could not solve the diversity problem through joint action, but it does suggest the bar associations will not be the medium through which they operate. A firm that had a particular passion about diversity could start an initiative to, say, disclose detailed statistics on the ability of associates of color to become partners at their firm—the sort of thing firms usually do not publish. Competitor firms would have no incentive to join that initiative. Even if there was an agreement that firms would only publish the data if a sufficient number of competitors also pledged to, the effort would be a logistical headache for little immediate gain. There would be significant incentive to renege on the agreement,

---

217. *See Board of Trustees*, STATE BAR OF CAL., [https://perma.cc/MK7Y-DFB2] (last visited Jan. 24, 2023); ALMARANTE ET AL., *supra* note 210, at 2. Without wishing to call attention to anyone in particular, a casual examination of California bar registrations reveals several authors work at non-profits. *See Attorney Search*, STATE BAR OF CAL., [https://perma.cc/LEF5-YUGC] (last visited Jan. 24, 2023).

218. *Cf.* Julianne Hill, *For More Good: Law Firms Find Other Ways to Provide Service to Society*, ABA J., Feb. 2017, at 34, 36 (describing financial donations from law firms to legal non-profits).

219. Again, without wishing to call attention to individuals, see *2020-2021 Board of Trustees*, STATE BAR OF CAL., [https://perma.cc/FJD9-2W6W] (last visited Jan. 9, 2023).

and firms would ultimately end up squabbling over whether a certain level of disclosure met their pledge of transparency. While effective self-policing by the firms is possible in theory, the fact that firms have not yet implemented anything like it suggests it will not be part of the ultimate solution.

2. Coordination Through Outside Intervention: Direct Regulation

The classic resolution of a multipolar trap is the imposition of order from outside, and it is easy to see why.[220] Outside entities, most notably government, can cut through the various incentive problems inhibiting firms from cooperating and simply mandate that competitors not sacrifice certain values. In a stereotypical multipolar trap, factories pollute because they cannot otherwise keep up with their competitors.[221] Government can intervene and put firm limits on pollution.[222] However, outside intervention is not a choice between direct government regulation and doing absolutely nothing.

Obviously, direct intervention by some level of government *could* solve the diversity problem in law firms through some sort of simple mandate: by some specific date, some specific percentage of equity partners at law firms with over $100 million in revenue must be non-White. How the Supreme Court—especially a six Republican-appointed Justice Court—would regard the constitutionality of such a requirement is difficult to say.[223] While state governments typically are the direct regulators of lawyers in their jurisdictions, the federal government could offer its authority under the Fourteenth Amendment as sufficient grounds to uphold such a requirement.[224] One could also envision

220. Alexander, *supra* note 23.

221. *Id.*

222. *Id.*

223. The most recent major affirmative action case, *Fisher v. Univ. of Tex.*, 579 U.S. 365 (2016), resulted in a 4-3 decision upholding the University of Texas's admissions policy. Since then, the Court has added three conservative Justices, and one liberal has left the Court. *See Supreme Court of the United States*, BALLOTPEDIA, [https://perma.cc/X76J-5SL6] (last visited Jan. 24, 2023). While this case regarded higher education admissions, it is potentially relevant to other race-conscious policies.

224. *See* U.S. CONST. art. XIV, § 5 (granting Congress power to enforce the provisions of the amendment).

slightly less direct policies that might make a difference in surviving a constitutional challenge—say, a significant tax credit for firms with a certain percentage of non-White equity partners.[225]

Two major problems complicate direct intervention by any level of government. The first is political reality, which interestingly mirrors the Supreme Court's own confused precedent on affirmative action.[226] The concept of affirmative action polls quite well—Gallup found in a 2021 poll that 62% of Americans favor affirmative action programs for racial minorities, up from 47% in 2001.[227] However, an even stronger majority—74%—said that companies and organizations should not "take a person's race and ethnicity into account, in addition to their qualifications, in order to increase diversity in the workplace."[228] It is difficult to logically square those two views—how else would an affirmative action program work other than to "take a person's race and ethnicity into account"? Nevertheless, a direct mandate on racial diversity would essentially require companies and organizations to do something that already polls extremely badly, suggesting that the public would not support such a plan. Such a program surviving long enough to be effective is unlikely.

That, in turn, raises the second major problem with direct intervention: reaction. One can easily imagine how cultural conservatives would react to such an initiative.[229] If they could

225. There are any number of federal programs that specifically allocate funds for non-White entities, ranging from Opportunity Zones to agricultural subsidies for minority farm owners. *See* Alexander Golding & Charlie Metzger, *Opportunity Zones Haven't Fully Reached Their Potential, but Don't Write Them Off Yet*, FORTUNE (Sept. 16, 2020, 2:00 PM), [https://perma.cc/2BAT-HWJM]; *Minority and Women Farmers and Ranchers*, U.S. DEP'T. OF ARGIC., [https://perma.cc/89YS-QXYV] (last visited Jan. 24, 2023). While beyond the scope of this Article, it seems unlikely that the Supreme Court would strike down all of these programs in order to also strike down an initiative to bolster firm diversity.

226. *See, e.g.*, Yuvraj Joshi, *Racial Indirection*, 52 U.C. DAVIS L. REV. 2495, 2497-98 (2019) (describing the Supreme Court's willingness to uphold affirmative action so long as the policy in question is filtered through some aspect not explicitly linked to race).

227. Lydia Saad, *Americans' Confidence in Racial Fairness Waning*, GALLUP (July 30, 2021), [https://perma.cc/8R7G-DA7L].

228. Frank Newport, *Affirmative Action and Public Opinion*, GALLUP (Aug. 7, 2020), [https://perma.cc/4FPP-KBFU].

229. My guess would be: not well. *See, e.g.*, Evan McMorris-Santoro et al., *Students Fight Back Against a Book Ban That Has a Pennsylvania Community Divided*, (Sept. 16,

harness the negative reaction from the 74%(!) of Americans who would likely disapprove of direct intervention, they could make the situation worse than the status quo by, say, banning any race-conscious diversity efforts.[230]

Any attempt at direct regulation can also spur creative methods to undermine the regulatory intent. For instance, a firm could start granting tiny blocks of equity to senior associates, then start counting them as "equity partners" for regulatory purposes. That particular scheme could be avoided by defining an equity partner as requiring some specified share of firm revenue, but it would be difficult to establish a satisfactory line that would fit the range of equity partner numbers across large firms. The point is not that it is impossible to identify and counter attempts at regulatory avoidance, but it would add to the complexity and difficulty of the undertaking, particularly if the political support for direct regulation is weak.

3. Coordination Through Outside Intervention: Transparency and Shaming

Outside intervention can defuse a multipolar trap in ways other than direct regulation. The key is to alter the individual firm's cost-benefit calculus regarding the sacrifice of the value in question.[231] Direct regulation achieves that end by inflicting a known, punishing cost to sacrifice of the value, but less direct methods can also increase costs or allocate benefits. A coordinated effort involving rigorous transparency and public shaming can simultaneously add a significant cost to sacrificing the value and a meaningful benefit to upholding it. In the classic pollution example, if some outside entity publishes data on

---

2021, 12:10 AM), [https://perma.cc/EE63-5JHB] (describing one Pennsylvania community that banned books teaching about the history of race in the United States in too negative a tone).

230. Those who think such a position far-fetched would do well to remember the conservative push to regulate companies they deem insufficiently conservative. *See* Eric Levitz, *Rubio Endorses Labor Unions (As a Punishment for 'Woke' Companies)*, INTELLIGENCER (Mar. 12, 2021), [https://perma.cc/3UU5-SNRL]. Banning all race-conscious policies administered by States is also Justice Clarence Thomas's view, as expressed in his dissent in *Fisher v. Univ. of Tex.*, 579 U.S. 365, 389 (2016) (Thomas, J., dissenting).

231. *See* Alexander, *supra* note 23.

pollution by each factory, and consumers are willing to pay even a small premium to punish particularly egregious offenders, rational factory owners will accord at least some attention to reducing pollution.

In the case of racial diversity, one can easily imagine how such an effort would work. Firms would be required to measure and disclose specific data relevant to measuring efforts to improve racial diversity. Obviously, a specific breakdown of the racial composition of equity partners would be important, but there are much more nuanced and potentially damning statistics out there. For example, firms could be required to measure and disclose face-to-face time between each partner and associate. They could be required by state governments to disclose the amount of time each associate actually spends with clients, how much time partners spend on mentorship with which associates, or numerical evaluations of the work produced by associates. These are just the tip of the iceberg of known or knowable data that would be relevant to evaluate a firm's commitment to racial diversity.

To be sure, various bar associations currently release interesting data on diversity in the legal profession, several examples of which have been cited in this Article.[232] The vital difference between that and what I am proposing is specific, firm-level data. One cannot shame a law firm with statistics about all law firms in the aggregate. But if one could point out that one firm is doing significantly worse than a competitor and ask why, the issue could quickly become a liability for the firm. Corporate clients could face pressure campaigns to hire from firms that are acting consistent with the corporation's (and firm's) stated support of racial diversity.

Because of the obvious potential for embarrassment, firms are not likely to embrace this sort of transparency unless they have little choice. Regulation to force the disclosure of those statistics may be necessary, though there are intriguing alternative possibilities. A sufficiently tenacious and dedicated non-profit could assemble data on racial diversity through publicly available information on most firms' websites, publicly available

232. *See supra* note 206 and accompanying text.

information on individual equity partners (e.g., social media), and potentially even leaks of internal firm data.[233]

The transparency-plus-shaming strategy has several compelling advantages to direct regulation. First, it is not as politically or legally sensitive as mandated affirmative action. Recall that polling shows broad support for affirmative action but broad opposition to directly taking race into account in promotion decisions.[234] Using client/consumer opinion as a cudgel against firms lagging on racial diversity could achieve results without crossing the political red lines evident in polling.

Second, transparency plus shaming could actually work faster than a direct mandate. The speed at which diversity is increased depends on when sanctions would actually start hurting a recalcitrant firm. In direct regulation, the sanctions are known to be sometime in the distant future.[235] With shaming, the sanctions are unpredictable. They may never come, or they may come immediately upon publication of the data in question. They could even come after a few years when some high-profile case causes the public to turn its attention to the issue. Not being able to predict when sanctions could arise puts a premium on acting quickly, particularly if the cost of increasing diversity is low compared to the threat of losing clients.

CONCLUSION

Why racial diversity in law firms lags behind similar professions (and even other parts of the legal profession) calls for explanation. Often, explanations for lack of racial diversity in a specific field fall into one of two camps. One explanation is individual bias, the conscious or subconscious tendency to discriminate against people of color.[236] Subconscious bias is very

233. The recipient of such leaks would be legally protected, though the person doing the leaking could be punished if detected. *See* N.Y. Times Co. v. United States, 403 U.S. 713, 730 (1971).

234. *See supra* note 227 and accompanying text.

235. *See supra* Section II.G.7.

236. *See, e.g.*, JOAN C. WILLIAMS ET AL., AM. BAR ASS'N, YOU CAN'T CHANGE WHAT YOU CAN'T SEE: INTERRUPTING RACIAL & GENDER BIAS IN THE LEGAL PROFESSION 7-10 (2018), [https://perma.cc/KY6G-3JYF] (describing systematic bias, both conscious and not, against attorneys of color).

well studied and has indeed been proven to exist in many different contexts.[237] But unless a particular industry triggers some specific stereotype or prejudice, bias does not explain why some industries do better than others at promoting racial diversity. Another explanation is the broad socioeconomic inequalities facing minorities long before they go to work at law firms.[238] Again, there is undeniable truth in this explanation, but it is not industry-specific.[239] If anything, lawyers are probably more aware of the history of discrimination than other professionals for the simple reason that it is impossible to understand constitutional law without it.

Understanding why law firms lag behind on racial diversity must start with broader comparison—what differences are there between law firms and, say, medical organizations? I have seen no evidence that law firm leadership is *more* prejudiced than, say, leaders of the medical community, but we can point to any number of differences of structure, evaluation, and compensation between law firms and medical organizations.[240] These differences give us a starting place for finding the cause of law firms' diversity problems. The specific incentive problems I have identified in this Article may not completely explain law firms' failure regarding racial diversity, but I strongly believe that law firm structure and incentives merit further, exhaustive study.

Focusing on incentives and structure has many benefits. First, it allows us to ignore utopian rhetoric. Lawyers often brag about the role we played in ending many forms of discrimination, but we are less profuse about the roles we play in continuing

237. *See, e.g.*, Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991, 992 (2004) (finding that "Black-sounding" applicants with identical resumes get fewer calls from employers).

238. *See* Debra Cassens Weiss, *Study Finds 'Lopsided' Concentration of Socioeconomic Elites at Law Schools*, ABA J. (Oct. 5, 2011, 11:00 AM), [https://perma.cc/597J-KGRX].

239. *See* Adi Gaskell, *Why Class Diversity Matters at Work*, FORBES (Jan. 22, 2019, 8:36 AM), [https://perma.cc/BK83-5FSK].

240. For example, the relative ease of evaluating doctor performance means that social hobnobbing akin to that of associates at a law firm would pay fewer dividends. *See* Erin A. Egan et al., *Comparing Ethics Education in Medicine and Law: Combing the Best of Both Worlds*, 13 ANNALS HEALTH L. 303, 319-22 (2004) (comparing professional socialization between the legal and medical professions).

discrimination. For every attorney who participated in the NAACP's legal battles against segregation, there was an attorney on the other side fighting just as hard to keep segregation alive. Ultimately, lawyers are just people, and people fall on a spectrum of goodness. Some will sacrifice anything in the name of justice. Others will sacrifice a little. Still others will do the right thing if it is equally lucrative as evil. And some will work for evil with a passion normally reserved for doing good. Consequently, we should expect a strong correlation between good results and good incentives. The more sacrifice required to do the right thing, the less it will be done.

Another benefit to examining incentives is falsifiability. Given data and economic theory, we can offer specific predictions about, say, the role of an eat-what-you-kill system in restraining racial diversity.[241] It is harder to measure the role of, for example, subconscious prejudice in the specific failure of law firms to achieve greater diversity. While that may indeed be an important factor, how law firms could fix subconscious prejudice is much harder to discern. Similarly, the broad socioeconomic inequalities that disadvantage people of color in the United States undoubtedly play some role in the diversity travails of law firms, but if we allow firms to do essentially nothing while society at large works on inequality, we will be eternally waiting.

Finally, we should note that some of the incentive problems at law firms that have stymied racial diversity could be profitable to fix for reasons beyond diversity. It is implausible that the world and the nature of legal work has changed so dramatically since the late nineteenth century, but law firm structure has remained largely static.[242] Rethinking how firms work—how associates get ahead, how partners interact with associates, how clients interact with firms—could lead to better provision of legal services. We should not mistake the lack of innovation in large law firms as suggesting the perfect structure has been reached. Rather, law

241. *See supra* Section I.D.4.

242. *See, e.g.*, Mary E. Vandenack, *Evolving Law Firm Models with Innovation: Using Lessons Learned from the Pandemic*, LAW PRAC. MAG., July-Aug. 2021, [https://perma.cc/9RZ6-B982] (providing as a dubious example of law firm innovation: "Add text reminders of appointments with directions to the office or links to videoconferencing.").

firms, protected from competition by bar requirements, have not faced a need to innovate.[243] If a dearth of racial diversity ultimately requires changes to the fundamental structure of firms, it should provide a rare opportunity to rethink the firm model. If the current model is most efficient, then it will survive. If it is not, then we should not hesitate to experiment with new organizational schemes.

243. *See generally* Casey Flaherty, *On Law Firm Marketing Bullshit*, 3 GEEKS & A L. BLOG (Sept. 10, 2017), [https://perma.cc/2FZL-U6YY] ("We've cultivated the illusion of innovation where constant chatter about innovation in and of itself has convinced partners that their firms are innovative . . . .").

SEARCHING FOR A COMPROMISE: A CASE FOR THE CRYPTO LIKE-KIND EXCHANGE

John Paul Boyter*

INTRODUCTION

In recent years, cryptocurrencies, cryptoassets, electronic coins, tokens, non-fungible tokens, and other various terms for electronic assets have gained prodigious attention in the financial world. From the spike (and subsequent drop) in value of Bitcoin,[1] to people spending millions of dollars on pixelated pictures of punks,[2] the market for these assets has been extremely active despite its ups and downs. However, in addition to potential financial success via crypto markets, the development of crypto technology has allowed for a transformation of how individuals and institutions think of currency, financial security, and access to information.[3] Demonstrative of this fact is that El Salvador became the first country to accept Bitcoin as legal tender in September 2021.[4] Despite volatility concerns and economists' criticisms, El Salvador's leaders felt that the advantages of adopting Bitcoin outweighed any disadvantages.[5] In defense of

---

* J.D. Candidate, 2023, University of Arkansas School of Law; Articles Editor, *Arkansas Law Review*. First, the author thanks Professor William Foster for providing his knowledge and guidance on a difficult issue. Second, the author thanks Professor Danielle Weatherby for her assistance and positive reinforcement during the writing process. Finally, the author thanks his wife, Caroline Boyter, for her constant support and affirmation.

1. *See* John Edwards, *Bitcoin's Price History*, INVESTOPEDIA (Dec. 20, 2022), [https://perma.cc/2JYN-7DYN].

2. *See* Stephen Graves et al., *The 15 Most Expensive NFTs Ever Sold*, DECRYPT (Feb. 21, 2022), [https://perma.cc/9B5C-PR6G].

3. *See* Serenity Gibbons, *10 Ways Cryptocurrency Will Make the World a Better Place*, DUE (Jan. 17, 2022), [https://perma.cc/5TQW-KYES].

4. Santiago Pérez, *IMF Urges El Salvador to Ditch Bitcoin's Legal Tender Status*, WALL ST. J. (Jan. 25, 2022, 5:03 PM), [https://perma.cc/3LW2-KJY9].

5. *See* Ciara Nugent, *El Salvador Is Betting on Bitcoin to Rebrand the Country—and Strengthen the President's Grip*, TIME (Oct. 1, 2021, 9:56 AM), [https://perma.cc/5GAL-NXZ9].

the decision, Salvadoran President Nayib Bukele stated that the country's adoption of Bitcoin has allowed for its citizens to enjoy financial freedom that was unavailable prior to Bitcoin's adoption.[6]

While countries like El Salvador are spearheading the way for the longevity of crypto, here in the United States, certain policies are restricting the growth and utility of cryptoassets. The regulatory nightmare and competition between federal agencies to control cryptoassets is just one example.[7] However, an exploration of the regulatory issues posed by cryptoassets is not the focus of this Comment; rather, the taxation policies that the United States has adopted regarding cryptoassets is one of the major restraints placed on their utility. This Comment advocates for a seemingly small but greatly important change to the current tax treatment of cryptoassets. Because the Internal Revenue Service ("IRS") classifies cryptoassets as property, when a taxpayer exchanges one crypto for another, that taxpayer must recognize any gain or loss due to the transaction.[8] The average taxpayer who invests in or uses crypto is likely unaware of this fact. Thus, this Comment advocates that the exchange of one cryptoasset for another should qualify as a "like-kind" exchange under § 1031 of the Internal Revenue Code (the "Code"). Succinctly, a like-kind exchange allows taxpayers to defer any gain or loss that results from the exchange of one piece of property for another that is like kind.[9]

This change to the current regime would allow for taxpayers to defer gains on crypto-for-crypto transactions until the exchanged-for asset is ultimately disposed of for a form of property that is not like kind.[10] This is by no means a small ask because the Tax Cuts and Jobs Act of 2017 ("TCJA") revised § 1031 to only allow real property exchanges as like-kind exchanges.[11] While it seems blatantly incongruous to only allow

---

6. *Id.*

7. *See* Sheelah Kolhatkar, *The Challenges of Regulating Cryptocurrency*, NEW YORKER (Oct. 6, 2021), [https://perma.cc/V4L2-5KZH].

8. *See* I.R.S. Notice 2014-21, 2014-16 I.R.B. 938.

9. I.R.C. § 1031.

10. *See id.*

11. *See* Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, § 13303(a), 131 Stat. 2054, 2123 (codified as amended at 26 U.S.C. § 1031).

real property and cryptoasset exchanges (perhaps the most dissimilar pair of assets in existence) to qualify as like-kind exchanges, adopting this change would better serve the interests of the tax system and stimulate the development, and in turn, the stability of crypto.

Part I of this Comment explains what a cryptoasset is, as well as the current tax regime applicable to them. Part II defines like-kind exchanges and provides the historical context for the nonrecognition event. It also considers the IRS's recent guidance pertaining to crypto like-kind exchanges. Part III puts forth this Comment's main arguments for allowing crypto-for-crypto exchanges to qualify as like-kind exchanges.

I. WHAT ARE CRYPTOS? HOW ARE THEY TAXED?

For simplicity, this Comment will use the term "cryptoasset" and "crypto" to refer to "cryptocurrency" as contemplated by the IRS.[12] Currently, there is a wide breadth of terminology in the cryptosphere, and even those well-versed in the area may use different terminologies than their peers.[13] Thus, cryptoasset and crypto are used in this Comment interchangeably to refer to those assets that share the denominational requirements to be considered a cryptocurrency under the IRS's definition of virtual currency: a cryptoasset is (1) "a type of virtual currency" (2) "that utilizes cryptography to secure transactions" (3) "that are digitally recorded on a distributed ledger, such as a blockchain."[14]

---

12. Moreover, this Comment largely focuses on the two most popular cryptos, Bitcoin and Ether. James Royal & Brian Baker, *12 Most Popular Types of Cryptocurrency*, BANKRATE (Jan. 13, 2022), [https://perma.cc/SN4S-XRJL]. This provides for a much more concise discussion of the claim because, as of the time of writing, there are at least 21,844 cryptos in existence. Josh Howarth, *How Many Cryptocurrencies Are There in 2022?*, EXPLODING TOPICS (Nov. 25, 2022), [https://perma.cc/N94J-G6Z4]. Further, it is not necessarily this Comment's claim that all cryptoassets are created equally, nor should *every single* crypto be subject to similar treatment. However, because the technology is in its infancy, a taxation policy should be drawn in favor of promoting its development.

13. *Compare Cryptoassets Manual*, GOV.UK (Nov. 3, 2022), [https://perma.cc/67RF-MAM9], *with* Piyali Chatterjee & Eric Essian, *Demystifying Blockchain, Cryptoassets and Tokenization*, RICHEY MAY & CO. (Feb. 2019), [https://perma.cc/9PMB-S3JZ], *and* Aashish Pahwa, *What Is a Cryptoasset? Types of Cryptoassets [Ultimate Guide]*, FEEDOUGH (Jan. 15, 2022), [https://perma.cc/EH9D-PUX3].

14. Rev. Rul. 2019-24, 2019-44 I.R.B. 1004.

A. What Are Cryptocurrencies and Why Should We Care?

Before discussing the tax ramifications of crypto transactions, it is necessary to first define both virtual currency and cryptocurrency. The IRS defines virtual currency as "a digital representation of value that functions as a medium of exchange, a unit of account, and a store of value."[15] Underneath the umbrella of virtual currency, the IRS defines cryptocurrency as "a type of virtual currency that utilizes cryptography to secure transactions that are digitally recorded on a distributed ledger, such as a blockchain."[16] In the same ruling, the IRS felt that it was necessary to contrast these two definitions with that of foreign currency: "Foreign currency is the coin and paper money of a country other than the United States that is designated as legal tender, circulates, and is customarily used and accepted as a medium of exchange in the country of issuance."[17] This last definition is particularly interesting considering El Salvador's adoption of Bitcoin as legal tender.[18]

It is important to understand the value of cryptoassets before one can fully realize why the technology should be encouraged and why it is necessary to discover a more efficient tax treatment for it, at the very least, by allowing crypto-for-crypto exchanges to qualify as like kind under § 1031 of the Code. According to Satoshi Nakamoto, the person and/or group responsible for Bitcoin's creation, an electronic coin is "a chain of digital signatures."[19] This is a very different, yet similar, way of describing a traditional currency. Traditional currencies have generally either been considered a fiat currency or a commodity currency.[20] The former is a currency that is backed by a sovereign, and the value of such is determined by the stability and

15. *Id.*

16. *Id.*

17. *Id.*

18. *Countries Which Allow Cryptocurrency as Legal Tender*, COINMARKETCAP, [https://perma.cc/2ERD-B9LX] (last visited Feb. 7, 2023).

19. SATOSHI NAKAMOTO, BITCOIN: A PEER-TO-PEER ELECTRONIC CASH SYSTEM 2 (2008), [https://perma.cc/MT5Y-DS7K]; *Who Created Bitcoin?*, BITCOIN MAG. (Oct. 7, 2020), [https://perma.cc/Z6G9-GUQU].

20. Dave Ahern, *Fiat Money vs. Commodity Money: A Breakdown of the Pros and Cons*, EINVESTING FOR BEGINNERS (May 21, 2022), [https://perma.cc/A5P5-9GMR].

reliability of the sovereign that backs the currency.[21] In contrast, commodity currencies are those that are backed by a physical resource, such as gold or silver.[22] The significance of Nakamoto's definition of an electronic coin is that the value is determined and controlled *solely* by the individuals that use Bitcoin.[23] While this concept is not necessarily novel in the sense that all currencies are in some way affected by the value individuals place in traditional currencies—whether in the sovereign or the physical resource that backs the currency—it *is* novel in the sense that there is literally no other value tied to electronic coins other than what individuals decide they are worth.[24] However, this is not to say that Bitcoin and other cryptoassets are valueless. The true value of cryptoassets lies in the process that is used in creating, transferring, and verifying transactions involving cryptoassets. This process is encapsulated in the "blockchain."[25]

For brevity, this background will focus on Bitcoin's use of blockchain. The blockchain acts as a public record that tracks the creation and transfer of every Bitcoin.[26] Each time a Bitcoin user makes a transaction with her Bitcoin, it is recorded on the blockchain.[27] This transaction is then verified by the "private key" that is attached to the coin and/or "[b]itcoin wallet" and again by the process of "mining."[28] Mining is the combined action of people across the world that confirms pending transactions by solving mathematical equations that prove that the Bitcoin that was transferred or received was initially created on the original public ledger, or blockchain.[29] This is truly remarkable because no one group or individual is capable of

---

21. James Chen, *Fiat Money: What It Is, How It Works, Example, Pros & Cons*, INVESTOPEDIA (Apr. 19, 2022), [https://perma.cc/JKN9-C23Y].
22. *Id.*
23. *See* Eric Rosenberg, *How Is Bitcoin Valued?*, THE BALANCE (May 22, 2022), [https://perma.cc/EM9H-YAAE].
24. *See* Carla Tardi, *Understanding the Different Types of Cryptocurrency*, SOFI (July 12, 2022), [https://perma.cc/8ND6-5XHA].
25. *How Does Bitcoin Work?*, BITCOIN, [https://perma.cc/7KC3-KTQV] (last visited Feb. 8, 2023).
26. *Id.*
27. *Id.*
28. *Id.*
29. *Id.*

disrupting the blockchain because each transaction is confirmed or rejected by multiple miners around the globe that are incentivized to verify transactions by receiving Bitcoin as payment.[30] Further, each "block" in the blockchain is coded using extremely complex cryptography that disallows transactions that do not fit in seamlessly.[31] Every Bitcoin ever created can be traced back through each transaction it went through all the way to its inception.[32] Essentially, the blockchain and the process of mining have stepped in the shoes of the traditional bank that normally would verify transactions.[33] This fact is what gives Bitcoin and other similar cryptoassets their decentralized nature.

However, keep in mind that coins are just one type of cryptoasset.[34] To clarify, "coins" are cryptoassets that are intended to be used as a medium of exchange, and they are created and can only be used on the coin's "native blockchain."[35] However, "tokens" are digital representations of assets, meaning that the specific token can have different uses.[36] Similar to coins, tokens can be used for investment purposes or to make purchases; however, to illustrate the difference, holders of tokens may have additional privileges on certain platforms, such as staking tokens and earning interest, being entitled to view streaming content, and earning loyalty points for retailers that are managed on the same blockchain.[37]

Another example of a token is the Non-Fungible Token ("NFT"). NFTs are digital representations of, well, anything.[38]

---

30. *How Does Bitcoin Work?*, *supra* note 25; Euny Hong, *How Does Bitcoin Mining Work?*, INVESTOPEDIA (May 5, 2022), [https://perma.cc/FE5E-23TP].

31. *How Does Bitcoin Work?*, *supra* note 25.

32. *See id.*

33. *See* Aaron Hsieh, Note, *The Faceless Coin: Achieving a Modern Tax Policy in the Changing Landscape of Cryptocurrency*, 2019 U. ILL. L. REV. 1079, 1085.

34. Madana Prathap, *Know Your Cryptocurrency Lingo—Crypto Coins and Tokens Are Not the Same Thing*, BUS. INSIDER INDIA (Dec. 24, 2021), [https://perma.cc/2BGB-4YA4].

35. *Id.*

36. *Id.*

37. Jake Frankenfield, *What Are Crypto Tokens, and How Do They Work?*, INVESTOPEDIA (May 20, 2022), [https://perma.cc/9NAL-84SB]; Lyle Daly, *What Is Staking in Crypto?*, THE MOTLEY FOOL (Nov. 2, 2022, 10:58 AM), [https://perma.cc/2YQK-66WG].

38. Joshua Caswell & Leigh E. Furtado, *NFTs for Estate Planners: Not Just a Token Concern*, PROB. & PROP., Sept./Oct. 2021, at 10, 12.

Literally anything that is digital can be turned into an NFT, including pictures, art, videos, tweets, and even virtual real estate—the oxymoron of the century.[39] NFTs are "[n]on-[f]ungible" because each NFT cannot be replaced with anything else—each NFT is individually unique and cannot be substituted or replicated, comparable to a one-of-a-kind baseball card.[40] The ownership of an NFT is tracked the same way as coins and other tokens; each purchase, sale, or trade is tracked on a blockchain.[41] The mockery of NFTs can easily be put to rest by the massive amounts of money that have been invested in them.[42] Millions of dollars are being poured into this juvenile market, and unsurprisingly, most people are unaware that *each* transaction an individual makes with a cryptoasset, except for the initial purchase of crypto with fiat currency (at least on the buyer's side), is taxable.

B. Current Tax Treatment

In the seminal case of *Commissioner v. Glenshaw Glass*, the United States Supreme Court held that a taxpayer is in receipt of income, and subject to taxation on that income, when the taxpayer has an "instance[] of undeniable accession[] to wealth, clearly realized, and over which the taxpayer[] ha[s] complete dominion."[43] Within the Code, gross income is defined as "all income from whatever source derived, including . . . [g]ains derived from dealings in property."[44] Cryptoassets are treated as property for federal tax purposes, meaning that any gain that results from "dealing[]" in crypto will be included in the taxpayer's gross income.[45] Under § 1001, typical property transactions require that a taxpayer recognize a gain or loss upon

39. *Id.* at 11-12; *Virtual Real Estate NFT | Predecessor of Metaverse*, FINEXTRA (Dec. 31, 2021), [https://perma.cc/8LK4-FFL2].

40. Gary P. Kohn, *NFTs and the Law*, L.A. LAW., November 2021, at 19, 20.

41. Caswell & Furtado, *supra* note 38, at 12.

42. *See, e.g.*, Jacob Hale, *Top 10 Most Expensive NFTs Ever Sold*, DEXERTO (Dec. 8, 2022), [https://perma.cc/WHX2-WYEE].

43. Comm'r v. Glenshaw Glass Co., 348 U.S. 426, 431 (1955).

44. I.R.C. § 61(a).

45. Rev. Rul. 2019-24, 2019-44 I.R.B. 1004; I.R.S. Notice 2014-21, 2014-16 I.R.B. 938.

the exchange of property for either cash, goods, services, or another form of property.[46]

The interpretative regulation of § 1001 provides, "[T]he gain or loss realized from the conversion of property into cash, or from the exchange of property for other property *differing materially* either in kind or in extent, is treated as income or as loss sustained."[47] To determine whether a taxpayer has realized a gain or loss, the taxpayer must first determine her amount realized.[48] § 1001 of the Code states, "The amount realized from the sale or other disposition of property shall be the sum of any money received plus the fair market value of the property (other than money) received."[49] Then, to determine whether one has realized a gain or loss upon the sale or disposition of her property, one must determine her basis and/or adjusted basis in the property that she exchanged.[50]

Generally, and in the case of cryptoassets, one's basis in property is the cost of that property at the time the taxpayer acquired it including any transactional fees.[51] Further, because cryptoassets are treated as property, sales or dispositions of cryptoassets will result in either short-term or long-term gains or losses.[52] Short-term gains or losses are those that are realized after holding the property for less than one year.[53] Long-term gains or losses are those that are realized after holding the property for more than one year.[54] Short-term gains are taxed at the taxpayer's ordinary income tax rate while long-term gains are taxed at the more taxpayer-friendly capital gains tax rate.[55]

For example, if *A* purchases a Bitcoin for $50,000 on June 1, 2021, her basis in the Bitcoin is the cost—$50,000. Then, on August 1, 2021, *A* decides to sell her Bitcoin for $60,000. The

46. I.R.C. § 1001.
47. Treas. Reg. § 1.1001-1 (2017) (emphasis added).
48. I.R.C. § 1001.
49. I.R.C. § 1001.
50. I.R.C. § 1011; I.R.C. § 1012.
51. I.R.C. § 1012.
52. I.R.S. Notice 2014-21, 2014-16 I.R.B. 938-39.
53. I.R.C. § 1222.
54. I.R.C. § 1222.
55. *See Topic No. 409 Capital Gains and Losses*, IRS (Jan. 26, 2023), [https://perma.cc/UAH2-6RKS].

difference between the amount realized—$60,000—and *A*'s cost basis—$50,000—results in a $10,000 gain. Further, because this transaction took place within one year after *A* acquired the Bitcoin, the gain will be subject to *A*'s ordinary income tax rate, meaning *A* realized a short-term gain of $10,000.

Importantly, however, if *A* were to trade her Bitcoin for either property or services, she must determine whether she realized a gain or loss.[56] Now is when a plethora of issues arise due to the dichotomy that is the IRS's treatment of cryptoassets as property and the taxpayers' treatment of cryptoassets as currency. For example, an employer, *X*, who compensates *A* for her services with Bitcoin must realize a gain or loss at the time the Bitcoin is transferred. *X* must determine its basis in the Bitcoin transferred and realize a short-term or long-term gain or loss that results from the fair market value of the Bitcoin at the time it is transferred to *A* (or the fair market value of *A*'s services) less *X*'s basis in the Bitcoin. While this may be a typical transaction that results in run-of-the-mill tax considerations for *X*, *A* is likely unaware that she needs to track her basis in the Bitcoin if she ever wants to put it to use.

In a different context, consider if *A* purchased one Bitcoin for $30,000, and later she purchased another Bitcoin for $50,000. Then, *A* trades half of a Bitcoin for three Ether.[57] Which cost basis must *A* use to determine her tax liability? Pursuant to its guidance, the IRS allows taxpayers to specifically identify the units being sold or exchanged,[58] meaning that if *A* would prefer to have a higher cost basis for the transaction and she can specifically identify which units of Bitcoin she exchanged, she could choose to use the $50,000 basis. Specific identification requires the taxpayer to show:

> (1) the date and time each unit was acquired, (2) [the taxpayer's] basis and the fair market value of each unit at the

---

56. I.R.S. Notice 2014-21, 2014-16 I.R.B. 938-39.

57. Ether, or "Eth," is the form of currency that is used on Ethereum's blockchain. VITALIK BUTERIN, ETHEREUM: A NEXT-GENERATION SMART CONTRACT AND DECENTRALIZED APPLICATION PLATFORM 13 (2014), [https://perma.cc/CF7D-EVV8]. The distinction between Ethereum, the platform, and Ether, the cryptoasset, is an important one that is discussed further below. *See infra* notes 139-41 and accompanying text.

58. *Frequently Asked Questions on Virtual Currency Transactions*, IRS (Jan. 13, 2023), [https://perma.cc/AEC3-H7JR].

> time it was acquired, (3) the date and time each unit was sold, exchanged, or otherwise disposed of, and (4) the fair market value of each unit when sold, exchanged, or disposed of, and the amount of money or the value of property received for each unit.[59]

On the other hand, if *A* cannot provide the requisite information for specific identification, she must use the first in, first out ("FIFO") method.[60] If a taxpayer uses FIFO, the units disposed of are deemed to be disposed of in chronological order,[61] meaning that *A* would use her cost basis in the Bitcoin she purchased first—$30,000.

Ignoring specific identification and FIFO, just consider that *A* decides she wants to trade some of her Bitcoin for Ether. For example, say *A* trades one of her Bitcoins (with a basis of $50,000) for five Ether (valued at $60,000). Likely unbeknownst to *A*, this is a taxable event. In this case, *A* would realize and recognize a $10,000 gain taxed at either her ordinary income tax rate or capital gains tax rate depending on how long she held the Bitcoin. Further, the only reason that *A* wanted Ether is so she could purchase an NFT of her favorite pixelated picture, which is valued at four Ether. However, between the time *A* received her Ether and purchased her NFT, her five Ether increased in value from $60,000 to $80,000.[62] Now, one Ether is worth $16,000 ($80,000 divided by five Ether); however, when she purchased her Ether, her basis in one Ether was $12,000 ($60,000 divided by five Ether). Thus, the purchase of the NFT for four Ether is valued at $64,000 compared to her basis in those four Ether being $48,000. *A*'s amount realized—$64,000—less her cost basis—$48,000—results in a $16,000 gain.

To someone untrained in tax matters—specifically the approximate 27 million Americans that own some form crypto[63]—this might seem preposterous. *A* has to pay tax two

---

59. *Id.*
60. *Id.*
61. *Id.*
62. This is not as farfetched as it may seem. Compare Bitcoin's value in January 2021—$29,405.12—to October 2021—$66,008.47. *Bitcoin*, COINDESK, [https://perma.cc/BY8Z-K6DS] (last visited Jan. 22, 2023).
63. *Cryptocurrency Information About United States of America*, TRIPLEA, [https://perma.cc/XXE7-NY83] (last visited Jan. 22, 2023).

different times without ever receiving cash, services, or other forms of property. The justification for having to pay tax on exchanges of property that is not like kind stems from a concept deemed the "realization event."[64]

In *Cottage Savings Ass'n v. Commissioner*, the United States Supreme Court agreed "with the Commissioner that an exchange of property gives rise to a realization event under § 1001(a) only if the properties exchanged are 'materially different.'"[65] It is necessary to briefly mention the facts of the case to demonstrate what exactly a "realization event" means. In *Cottage Savings Ass'n*, Cottage Savings exchanged with various savings and loan associations its mortgage interests in 252 single-family homes for separate mortgage interests in 305 single-family homes.[66] At the time, the Federal Home Loan Bank Board ("FHLBB") allowed savings and loan associations to abstain from reporting losses "associated with mortgages that are exchanged for 'substantially identical' mortgages held by other lenders."[67] The purpose of this accounting regulation was to allow for transactions that would generate tax losses while not affecting the economic position of the entity in any substantial way.[68]

In the transaction, the face value of the mortgage interests Cottage Savings exchanged was approximately $6,900,000, and the fair market value of the mortgage interests Cottage Savings received was approximately $4,500,000.[69] Because the mortgage interests that Cottage Savings exchanged had decreased in value by nearly $2,500,000, Cottage Savings attempted to deduct $2,447,091 from its income as a loss, representing the difference between the face value of the mortgage interests relinquished and the fair market value of the mortgage interests received.[70]

The Commissioner disallowed the deduction on the ground that the properties exchanged were not "materially different,"—in other words, the properties were like kind—as evidenced by

---

64. Cottage Sav. Ass'n v. Comm'r, 499 U.S. 554, 559-60 (1991).
65. *Id.* at 560; *see also* Treas. Reg. § 1.1001-1 (2017).
66. *Cottage Sav. Ass'n*, 499 U.S. at 557-58.
67. *Id.* at 557.
68. *Id.*
69. *Id.* at 558.
70. *Id.*

the fact that Cottage Savings did not have to report losses on the exchange under the FHLBB accounting regulations because the exchange was for "substantially identical" mortgages.[71] Cottage Savings challenged the Commissioner's position in the Tax Court, which held that Cottage Savings was entitled to take the deduction because the properties exchanged were "materially different."[72] The Sixth Circuit reversed, holding that Cottage Savings did not actually sustain losses, rather than focusing on the "materially different" requirement of § 1001(a).[73]

Thus, before the Supreme Court, the relevant question was whether the properties exchanged were "materially different" under § 1001(a) and Treasury Regulation § 1.1001-1.[74] In interpreting the "materially different" requirement for recognition of gain or loss, the Supreme Court held that "properties are 'different' in the sense that is 'material' to the Internal Revenue Code so long as their respective possessors enjoy legal entitlements that are different in kind or extent."[75] The Supreme Court agreed with Cottage Savings and held that the mortgages exchanged were materially different because the loan interests possessed distinct legal entitlements—i.e., the loans were secured by different properties and made to different obligors.[76] All this to say, properties are materially different—or are not like kind—and require recognition of gain or loss upon exchange "so long as they embody legally distinct entitlements."[77]

II. DISCUSSION OF § 1031

A. History and Purpose of the § 1031 Like-Kind Exchange

It is important to understand the purpose and development of § 1031 exchanges to see why they should be available for

71. Cottage Sav. Ass'n v. Comm'r, 890 F.2d 848, 848, 850 (6th Cir. 1989), *rev'd*, 499 U.S. 554 (1991).

72. Cottage Sav. Ass'n v. Comm'r, 90 T.C. 372, 402 (1988), *rev'd*, 890 F.2d 848 (6th Cir. 1989), *rev'd*, 499 U.S. 554 (1991).

73. *Cottage Sav. Ass'n*, 890 F.2d at 855.

74. *Cottage Savings Ass'n*, 499 U.S. at 560.

75. *Id.* at 565.

76. *Id.* at 566.

77. *Id.*

cryptocurrency transactions. Ultimately, the purpose of a like-kind exchange is to allow the taxpayer to defer any gain that would otherwise be recognized upon the initial exchange of like-kind properties to a point in time where the taxpayer exchanges the second property for another that is not like kind.[78] This allows the taxpayer to simply transfer her basis in the exchanged property to the property exchanged for. For example, if *A* exchanges Blackacre with a cost basis of $100,000 to *X* for Greenacre, *A*'s basis in Greenacre will be $100,000. This means that if *A* later exchanges Greenacre for $200,000 in cash or for $200,000 worth of stock in a company, *A* will recognize a $100,000 gain. It is important to realize that *A* does not avoid her tax liability. She merely defers the (potential) tax liability to the point in time in which she disposes of the exchanged-for property for property that is not like kind.

Currently, § 1031 provides:

> No gain or loss shall be recognized on the exchange of *real property* held for productive use in a trade or business or for investment if such *real property* is exchanged solely for real property of like kind which is to be held either for productive use in a trade or business or for investment.[79]

However, when the concept was first introduced, section 202(c)(1) of the Revenue Act of 1921 provided that

> an exchange of *property, real, personal or mixed*, for any other such property, no gain or loss shall be recognized unless the property received in exchange has a readily realizable market value; but even if the property received in exchange has a readily realizable market value, no gain or loss shall be recognized [if the transaction meets certain conditions.][80]

As discerned from its plain language, if a taxpayer received property, that taxpayer was required to recognize gain or loss if "the property received in exchange ha[d] a readily realizable

78. Julia Kagan, *Like-Kind Exchange: Definition, Example, Pros & Cons*, INVESTOPEDIA (Nov. 30, 2021), [https://perma.cc/UXU6-UBLF].

79. I.R.C. § 1031(a)(1) (emphasis added).

80. Revenue Act of 1921, Pub. L. No. 67-98, § 202(c), 42 Stat. 227, 230 (emphasis added).

market value."[81] However, even if a taxpayer received property with a readily realizable market value, that taxpayer would not have to recognize gain (immediately) if the transaction consisted of like-kind properties and met other requirements.[82] Initially, in accordance with this old rule, taxpayers were able to enjoy the nonrecognition benefits when trading stocks, bonds, and other similar instruments.[83] Congress and others saw this as an abuse of the Code and subsequently amended the provision to exclude stocks, bonds, and similar instruments from the nonrecognition protection.[84] Further, Congress decided to tax any gain due to the receipt of property other than what was considered like kind (i.e., cash).[85] There are still passionate debates about whether holders of stocks, bonds, and similar instruments are treated unfairly compared to holders of real estate under the Code.[86] Nonetheless, Congress made it clear that taxpayers should not be able to use like-kind exchanges to skirt truly realized gain recognition.

When Congress adopted 1921's version of the like-kind exchange, the main two justifications for it were (1) the avoidance of taxing theoretical gains and losses and (2) the Continuity of Investment Theory.[87]

First, Congress recognized that the purpose of a like-kind exchange is to avoid taxing "paper" gains and losses when merely the form, rather than the substance, of the investment has changed.[88] Congress stated:

---

81. § 202(c), 42 Stat. at 230.

82. § 202(c), 42 Stat. at 230; *see also* Marjorie E. Kornhauser, *Section 1031: We Don't Need Another Hero*, 60 S. CAL. L. REV. 397, 403-04 (1987).

83. Kornhauser, *supra* note 82*,* at 403.

84. *Id.*

85. *Id.* at 404.

86. *See* Albert B. Crenshaw, *EPIC Built Empire on Foundation of U.S. Tax Code*, WASH. POST (Sept. 1, 1985), [https://perma.cc/5ZQ2-RP9Y].

87. Another congressional purpose for the like kind exchange at the time was administrative convenience. LEGISLATIVE HISTORY OF TAX POLICIES SUPPORTING IRC SECTION 1031, FED'N OF EXCH. ACCOMMODATORS [hereinafter FEA LEGISLATIVE HISTORY OF 1031 WHITEPAPER], [https://perma.cc/Q5HT-AW9F] (last visited Feb. 11, 2023). This was due to the difficulty of determining the values of the properties exchanged. *Id.* However, this purpose was abandoned only three years later in a 1924 amendment to the statute. *Id.* Although, perhaps this justification holds much greater weight in the context of crypto exchanges. *See* discussion *infra* Section III.D.

88. FEA LEGISLATIVE HISTORY OF 1031 WHITEPAPER, *supra* note 87.

> In other words, profit or loss is recognized in the case of exchanges of notes or securities, which are essentially like money; or in the case of stock in trade; or in case the taxpayer exchanges the property comprising his original investment for a different kind of property; but if the taxpayer's money is still tied up in the same kind of property as that in which it was originally invested, he is not allowed to compute and deduct his theoretical loss on the exchange, nor is he charged with a tax upon his theoretical profit. The calculation of the profit or loss is deferred until it is realized in cash, marketable securities, or other property not of the same kind having a fair market value.[89]

In interpreting the legislative history of § 1031, the Second Circuit acknowledged that "Congress was primarily concerned with the inequity, in the case of an exchange, of forcing a taxpayer to recognize a paper gain which was still tied up in a continuing investment of the same sort."[90] The same principle holds true today; however, perhaps this principle should have broader application for other forms of investments, such as crypto.

Admittedly, the terms "essentially like money," "marketable securities," and "cash" ring an alarm that leads to the reasonable conclusion that crypto is often treated like currency or a cash equivalent, so why should the IRS allow for crypto like-kind exchanges?[91] The short answer is that by classifying crypto as property, it is simply not cash nor essentially like money because an obstacle stands in between the taxpayer and the utility of the crypto—a taxable event.[92]

Second, recognized as one of the more prominent justifications for § 1031 exchanges is the Continuity of Investment Theory.[93] Essentially, this theory holds that after a like-kind exchange, the taxpayer is practically in the same, or very close to the same, position that the taxpayer was in prior to the

---

89. H.R. REP. NO. 73-704, at 13 (1934).

90. Jordan Marsh Co. v. Comm'r, 269 F.2d 453, 456 (2d Cir. 1959). Similarly, the Ninth Circuit found that the legislative history of § 1031 showed that the provision "was designed to avoid the imposition of a tax on those who do not 'cash in' on their investments in trade or business property." Starker v. United States, 602 F.2d 1341, 1352 (9th Cir. 1979).

91. *Cf.* H.R. REP. NO. 73-704, at 13 (1934).

92. See *infra* Part III for the long answer and a comparison of stocks and crypto.

93. Erik M. Jensen, *The Uneasy Justification for Special Treatment of Like-Kind Exchanges*, 4 AM. J. TAX POL'Y 193, 199 (1985).

exchange.[94] While this theory somewhat relies on the illiquidity of the property exchanged,[95] the essence of the theory is that § 1031 provides a solution for the taxpayer "who is unable or unwilling to sell investment property because of the burden that capital gains and recapture taxes would place on the taxpayer's cash flow and net worth."[96] This justification advances the country's and individuals' interests by incentivizing investment efforts and transactional activity.[97]

Both justifications may be read along with each other and are still relevant today.[98] The synthesis of the two presents a policy that promotes the transferability of the substance of an investment without the hindrance or interruption of a resulting tax liability. However, in 2017, Congress modified § 1031 through the TCJA and narrowed the scope of nonrecognition by limiting like-kind exchanges to only those that are exchanges of real property.[99]

B. Definition of Like Kind

If the tax benefits of § 1031 are only available if the exchanged properties are like kind, what exactly does that mean? Prior to the TCJA, it was sometimes difficult to determine whether property was like kind because exchanges of personal property could potentially seek refuge under § 1031.[100] Prior to the TCJA and now, the IRS considers properties to be like kind "if they're of the same nature or character, even if they differ in grade or quality."[101] While only real properties may be like kind

---

94. *Id.*; *see also* H.R. REP. NO. 73-704, at 13 (1934).

95. Jensen, *supra* note 93, at 199 n.29.

96. FEA LEGISLATIVE HISTORY OF 1031 WHITEPAPER, *supra* note 87.

97. *Id.*

98. *Id.*

99. Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, § 13303, 131 Stat. 2054, 2123 (codified as amended at I.R.C. § 1031); INTERNAL REVENUE SERV., U.S. DEP'T OF THE TREASURY, PUBLICATION 544: SALES AND OTHER DISPOSITIONS OF ASSETS 12 (2022), [https://perma.cc/KZB8-TJY4].

100. *See Like-Kind Exchanges—Real Estate Tax Tips*, IRS (Nov. 10, 2022), [https://perma.cc/P3DB-E9JW].

101. *Id.* Now, however, the inquiry is much clearer. In the same guidance, the IRS states, "Real properties generally are of like-kind, regardless of whether they're improved or unimproved." *Id.*

under the current regime, it is helpful to explore the IRS's past guidance and allowances of like-kind exchanges involving personal property.

The relevant portion of the definition of like kind is "same nature or character."[102] In a 2008 IRS Fact Sheet, the guidance attempted to define what "nature, character or class" meant but failed to do so in any meaningful way.[103] However, the guidance states, "In personal property exchanges, the rules pertaining to what qualifies as like-kind are more restrictive than the rules pertaining to real property."[104]

Examples of personal property like-kind exchanges allowed in the past include an FCC Radio License for an FCC Television License, baseball player contracts, fishing permits, and coins of the same material.[105] These examples illustrate that the Code allowed for certain exchanges of properties to qualify as like kind even though they possessed arguably much greater differences than two separate cryptoassets. For example, regarding the fishing licenses, the IRS stated, "The exchange of a fishing permit/license for another fishing permit/license qualifies for nontaxable exchange treatment under IRC § 1031, regardless of whether the permit is for a different fishery, a different species of fish, or a different type of fishing gear."[106] To analogize, the exchange of one crypto for another should qualify for § 1031 treatment regardless of whether the crypto is used on a different platform, to purchase an NFT, or to stake an investment.

C. Held for Investment Requirement of § 1031

Along with being like kind, the exchanged property and the exchanged-for property must both "be held either for productive use in a trade or business or for investment."[107] For simplicity and relevance, this Comment will focus on property held for

102. *Id.*
103. I.R.S. Fact Sheet FS-2008-18 (Feb. 2008), [https://perma.cc/CB5Q-SY6T].
104. *Id.*
105. Eli Cole, *Cryptocurrency and the § 1031 Like Kind Exchange*, 10 HASTINGS SCI. & TECH. L.J. 75, 93-95 (2019).
106. I.R.S. Fishing Audit Technique Guide (Aug. 2011), [https://perma.cc/X72D-UHK3].
107. I.R.C. § 1031(a)(1).

investment. Whether property is held for investment depends on the intent of the property holder.[108] The IRS has recognized that cryptos may be held for investment.[109]

Notably missing from § 1031 is personal use property.[110] This is an important distinction because if a crypto holder is found to hold the crypto for personal use, it will not be subject to this Comment's proposed rule. Personal use property is anything a taxpayer owns and uses for that taxpayer's own benefit and enjoyment.[111] In contrast, property held for investment is generally understood as property that produces income or appreciates in value.[112] While a taxpayer might hold crypto for both investment and personal use, the primary intent of the taxpayer will control.[113] Thus, only those taxpayers whose *primary intent* for holding crypto is for investment purposes will be subject to this Comment's proposed rule.

D. IRS Memorandum Considering Whether Crypto Swaps Prior to the TCJA Qualified as Like Kind

To further complicate achieving this Comment's objective, in a 2021 Memorandum, the IRS plainly stated that crypto swaps made even prior to the TCJA did not qualify for § 1031 treatment.[114] The IRS first considered exchanges of Litecoin for Bitcoin or Litecoin for Ether.[115] In holding that neither of these transactions were like kind, the IRS stated that Bitcoin and Ether held "special position[s]" compared to other cryptos.[116] This position was the "on and off-ramp" for other cryptos, meaning that if a taxpayer wanted to either purchase or sell Litecoin, in the event of a purchase, that taxpayer would have to buy Bitcoin (or Ether) first and use that Bitcoin to purchase Litecoin, or in the

108. *See* Bolker v. Comm'r, 81 T.C. 782, 804 (1983), *aff'd*, 760 F.2d 1039 (9th Cir. 1985).

109. I.R.S. Notice 2014-21, 2014-16 I.R.B. 938.

110. I.R.C. § 1031.

111. Carissa Rawson, *What Is Personal Use Property?*, THE BALANCE (Apr. 27, 2022), [https://perma.cc/ND5W-L8YD].

112. *See* I.R.S. Form 4952 (2022), [https://perma.cc/U59S-HL62].

113. *See Bolker*, 81 T.C. at 804.

114. I.R.S. Mem. 202124008, at 1 (June 8, 2021), [https://perma.cc/U4EW-RPFB].

115. *Id.*

116. *Id.* at 3.

event of a sale, trade Litecoin for Bitcoin and then ultimately dispose of the Bitcoin for cash.[117] The IRS came to the conclusion that this special position that Bitcoin and Ether held "played a fundamentally different role from other cryptocurrencies within the broader cryptocurrency market."[118]

Next, the Memorandum contemplated whether a trade of Bitcoin for Ether or vice versa qualified as like kind.[119] Again, the IRS concluded that the cryptos were not like kind.[120] In so concluding, the IRS stated that "while both cryptocurrencies share similar qualities and uses, they are also fundamentally different from each other because of the difference in overall design, intended use, and actual use."[121] The Memorandum acknowledged that both Bitcoin and Ethereum are intended to be used as "payment network[s]" but found that Ethereum's additional functionality of operating as a platform for smart contracts and applications made it *fundamentally* different from Bitcoin.[122]

III. WHY CRYPTO-FOR-CRYPTO SHOULD BE SUBJECT TO § 1031 TREATMENT

A. The (Non)realization Event

The first argument for allowing crypto like-kind exchanges is that swapping one crypto, say Bitcoin, for another, say Ether, does not constitute a realization event as understood in *Cottage Savings Ass'n*. This is so because, contrary to the IRS's opinion, the cryptos are not "materially different" from one another. As stated previously, properties are materially different if they embody legally distinct entitlements.[123] Inescapably, certain cryptoassets have different properties than others. For example, Bitcoin operates as an electronic form of payment intended to be

117. *Id.*
118. *Id.*
119. I.R.S. Mem. 202124008, *supra* note 114, at 3.
120. *Id.*
121. *Id.*
122. *Id.*
123. Cottage Sav. Ass'n v. Comm'r, 499 U.S. 554, 566 (1991).

used as a currency on Bitcoin's network,[124] whereas Ether operates as a currency on Ethereum's platform that enables the use of smart contracts and decentralized applications.[125] However, while cryptoassets may have different properties, they are *sufficiently* like kind to justify the deferral of any gain realized in a crypto-for-crypto transaction because no gain is truly realized until the cryptoasset is exchanged for a completely different form of property, for services, or for cash.

To state it plainly, Bitcoin and Ethereum do not possess legally distinct entitlements as understood in *Cottage Savings Ass'n*. There, the Court considered that the mortgages exchanged and the mortgages exchanged for derived from loans made to different obligors that were secured by different properties.[126] The key distinction between the assets considered in *Cottage Savings Ass'n* and cryptoassets is that, in the former situation, the taxpayer was the recipient of a completely different set of rights and a *new* set of entitlements.[127] However, in the case of crypto, the only difference between holders of Bitcoin and holders of Ether is that the holders of Ether have the *ability* to spend Ether on Ethereum's network to operate smart contracts or applications.[128] Importantly, this ability does not come with a new set of rights, such as those contemplated in *Cottage Savings Ass'n* or those that accompany the purchase of stock.[129] Because Ether is simply the "fuel" used on Ethereum's network, when taken out of context of the network, it is just another form of investment nearly identical to that of Bitcoin.[130]

To illuminate the likeness of cryptos, it is helpful to compare legal entitlements that accompany stocks to legal entitlements that accompany decentralized cryptos. In enacting § 1031, Congress clearly did not intend for marketable securities, such as stocks, to

---

124. NAKAMOTO, *supra* note 19, at 1.

125. BUTERIN, *supra* note 57, at 13.

126. *Cottage Sav. Ass'n*, 499 U.S. at 566.

127. *See id.*

128. BUTERIN, *supra* note 57, at 13.

129. *See Introduction to Ethereum Governance*, ETHEREUM (Feb. 10, 2023), [https://perma.cc/675J-WXWV].

130. Nathan Reiff, *What Is Ether (ETH), the Cryptocurrency of Ethereum Apps?*, INVESTOPEDIA (Nov. 17, 2022), [https://perma.cc/46ML-K5CB].

benefit from like-kind exchanges.[131] However, it is important to understand the reasons why and how those reasons are absent when discussing crypto. To start, a common stockholder of a corporation does not just receive equity in the corporation; rather, the stockholder receives a bundle of rights along with the shares.[132] These rights include voting power and the right to sue for certain wrongful acts, among others.[133] These rights fit neatly within the Supreme Court's definition of "materially different" properties in *Cottage Savings Ass'n*.[134] For example, if *A* owns stock in corporation *X*, *A* may participate in shareholder voting for *X* or in lawsuits against *X* and its directors. However, if *A* trades her shares in *X* for shares in corporation *Y*, *A* may no longer participate in shareholder voting for *X*. It is not hard to see why *A*'s investment has changed in substance—she has a completely new and different set of legal entitlements because she can no longer vote in corporate elections for *X*, nor can she sue *X*'s directors for breaches of fiduciary duties.

While certain blockchains may possess some form of governance,[135] opportunities to vote on the course that a specific crypto technology takes pale in comparison to shareholder voting.[136] Furthermore, and perhaps more importantly, crypto holders are not entitled to any causes of action against the creators or developers of the technology for mismanagement or breaches of fiduciary duties.[137] Thus, crypto investors simply have an

131. H.R. REP. NO. 73-704, at 13 (1934).

132. Basia Hellwig, *Know Your Shareholder Rights*, INVESTOPEDIA (May 31, 2022), [https://perma.cc/2HP9-PR3T].

133. *Id.*

134. Cottage Sav. Ass'n v. Comm'r, 499 U.S. 554, 566 (1991).

135. *See* Jake Frankenfield, *On-Chain Governance: Definition, Types, Vs. Off-Chain*, INVESTOPEDIA (Oct. 25, 2021), [https://perma.cc/VZF5-YCEJ].

136. *Compare* Adam Hayes, *What Are Stockholder Voting Rights, and Who Gets a Vote?*, INVESTOPEDIA (Mar. 21, 2021), [https://perma.cc/ASQ2-TUMA] (discussing how shareholders vote on certain corporate governance issues *annually* for all public corporations), *with* Frankenfield, *supra* note 135 (discussing how individual crypto holders may participate in scarce, informal governance matters; however, this governance is claimed to be centralized among developers, and the consequences of such centralization led to the hard forks that both Ethereum and Bitcoin went through).

137. *See* Andrea Tinianow, *When Blockchains Crash, Who Can You Sue?*, FORBES (Feb. 7, 2019, 2:32 PM), [https://perma.cc/4ZN6-JA6K] (discussing how tort law is likely the only remedy for damaged investors and how crypto developers are not fiduciaries because they simply suggest code enhancements that are either accepted or rejected by miners and investors).

economic stake in whatever crypto they own without the participation and legal rights that shareholders of a corporation possess. Because a crypto investor does not avail herself of a separate or new set of rights upon the exchange of one crypto for another, the exchange is not a realization event and thus should not be a taxable event.

B. Rebutting the IRS's Conclusion in Its 2021 Memorandum

The IRS concluded in its 2021 Memorandum that because Ethereum allows for the operation of a payment network as well as the operation of applications and smart contracts on its platform, it cannot be like kind to Bitcoin, whose platform is meant to be used purely as a payment network.[138] This reasoning is unpersuasive. While the Ethereum *platform* allows for the operation of smart contracts and applications, the actual crypto in dispute, Ether, is the currency that is used on the platform.[139] Because Ether is simply the crypto that is used on Ethereum's *platform*, the only difference between an Ether and a Bitcoin—other than the underlying ones and zeroes—is the intended use the creators envisioned.[140] However, the overwhelming conduct of those involved in crypto is investment.[141]

To analogize, consider that an exchange of farmland in central Kansas for an apartment complex in New York City would qualify as like kind.[142] The owners of such properties may have completely different intended uses for the properties; however, as long as the owners hold the properties for *investment*, the inquiry comes to an end, and the owners may defer gain under § 1031.[143] The point is that as long as the holders of Bitcoin and Ether hold the cryptos for investment purposes, if they exchange one for the

138. I.R.S. Mem. 202124008, *supra* note 114, at 3.

139. Reiff, *supra* note 130.

140. Nathan Reiff, *Bitcoin vs. Ethereum: What's the Difference?*, INVESTOPEDIA (Oct. 4, 2022), [https://perma.cc/E5KR-2B6W].

141. *See* Nathan Reiff, *Why Should Anyone Invest in Crypto?*, INVESTOPEDIA (Aug. 24, 2021), [https://perma.cc/6BCN-QDEF].

142. *Like-Kind Property: What Qualifies and What Doesn't*, 1031GATEWAY, [https://perma.cc/BTU6-ARJD] (last visited Feb. 11, 2023).

143. Provided the other requirements are met. *See* I.R.C. § 1031; *Like-Kind Property: What Qualifies and What Doesn't*, *supra* note 142.

other, the nonrecognition allowed in § 1031 should apply. To be clear, this Comment is not advocating for a change in policy regarding real estate like-kind exchanges. Rather, the point is that Congress and the IRS have made a necessary concession in allowing for the deferral of gain in land transactions, even when the pieces of property are drastically different, for good reason. However, going one step further to allow for similar treatment in crypto transactions would only further advance the purposes Congress adopted when it enacted § 1031.[144]

Even if the reader does not buy the argument that Bitcoin and Ether are sufficiently similar and agrees with the IRS that, because the two cryptos have different intended uses, they cannot be like kind, the reader should consider the case of swapping Bitcoin for Litecoin. Both Bitcoin's and Litecoin's platforms are intended to be used as payment networks.[145] Now is when the reasoning of the IRS Memorandum appears unavoidably flawed. If *A* swaps one Bitcoin for 400 Litecoins, how has her position changed? Simply put, *A* is in a nearly identical position before and after the exchange even if the value of the 400 Litecoins exceeds *A*'s cost basis in her Bitcoin. This is so because, assuming an arm's length transaction, the one Bitcoin and the 400 Litecoins had equal fair market values. The only difference is the form of crypto that *A* has an economic interest in—*A* has not realized income as defined in *Glenshaw Glass*,[146] nor has she received a new set of legal entitlements as understood in *Cottage Savings Ass'n*.[147]

---

144. *See infra* Section III.C.

145. *What Is Litecoin?*, LITECOIN, [https://perma.cc/GJ7S-5HPW] (last visited Apr. 18, 2022). Perhaps an easier bright-line rule would be to only allow crypto like-kind exchanges for cryptos that are not classified as securities. However, whether a certain crypto qualifies as a security is not an easy question to answer. For example, in 2018, then SEC chairman Jay Clayton stated plainly, "Cryptocurrencies: These are replacements for sovereign currencies, replace the dollar, the euro, the yen with bitcoin . . . . That type of currency is not a security." Kate Rooney, *SEC Chief Says Agency Won't Change Securities Laws to Cater to Cryptocurrencies*, CNBC (June 11, 2018, 9:35 AM), [https://perma.cc/JU72-ZZ7F]. However, when discussing tokens such as Ethereum, Clayton said, "A token, a digital asset, where I give you my money and you go off and make a venture, and in return for giving you my money I say 'you can get a return' that is a security and we regulate that." *Id.*

146. *See* Comm'r v. Glenshaw Glass Co., 348 U.S. 426, 429 (1955).

147. *See* Cottage Sav. Ass'n v. Comm'r, 499 U.S. 554, 555 (1991).

C. Applying Congress's Justifications for § 1031 to Crypto

First, in light of the "paper" gain justification, Congress recognized the inherent unfairness of taxing individuals whose investment merely changed form rather than substance, even though the individual received property with greater value than the individual's basis in the exchanged property.[148] In other words, Congress intended to avoid "speculative taxation in the middle of an ongoing investment" when "[a] duplex investment property becomes a six flat, which becomes a strip mall, then a shopping center."[149]

The same reasoning is effortlessly applied in the context of crypto exchanges. When *A* trades her one Bitcoin with a basis of $50,000 for five Ether with a fair market value of $60,000, *A* merely has a paper gain. She has simply changed the form of crypto that her financial interest presides in. By requiring *A* to pay tax on the exchange, the IRS is "forcing a taxpayer to recognize a paper gain which was still tied up in a continuing investment of the same sort."[150]

In the same vein, the Continuity of Investment Theory establishes a policy that encourages property exchanges.[151] Inherent in this justification is the notion that Congress does not want to prevent a taxpayer from changing the form of her investment because of a potential tax liability. This allows investors to look for more lucrative opportunities in which to invest, which promotes economic stimulus and transactional activity.[152] In real estate transactions, taxpayers commonly trade parcels of real estate because the property the taxpayer is receiving either looks more promising from an economic standpoint or the property simply meets the taxpayer's desires.

When a taxpayer chooses to invest in one crypto over another, reasons why may include comparative volatility and risk,

---

148. H.R. REP. NO. 73-704, at 13 (1934).
149. FEA LEGISLATIVE HISTORY OF 1031 WHITEPAPER, *supra* note 87.
150. Jordan Marsh Co. v. Comm'r, 269 F.2d 453, 456 (2d Cir. 1959).
151. FEA LEGISLATIVE HISTORY OF 1031 WHITEPAPER, *supra* note 87.
152. *Id.* at 1; Kagan, supra note 78.

faith in one technology over another, the ability to use the crypto on a certain network, or mere speculation.[153] Just as in real estate exchanges, a taxpayer might lose faith in the potential success of her current investment or may believe there is greater potential for success in a different investment. Thus, a logical conclusion follows: taxpayers should be free to swap one crypto for another without incurring tax liability because it would promote transactional activity and economic stimulus.

D. Practical Considerations

Perhaps the most persuasive arguments for allowing a § 1031 crypto exchange are everyday, practical considerations: it would dramatically increase the ease with which a taxpayer assesses his or her tax liability, allow for administrative ease in revenue collection, and stimulate the growth and stability of crypto.

First, a taxpayer who is unaware of the tax classification of crypto is due for a rude awakening come Tax Day if the taxpayer made any transactions with the crypto. In response to this fact, software—e.g., CoinLedger—has been created specifically to track crypto-tax events.[154] However, even so, a survey found that only 54.8% of crypto users reported cryptocurrency on their taxes.[155] This is partly due to tax reporting platforms, such as TurboTax, not having the ability to track crypto transactions across multiple platforms.[156] Further, some accountants lack the knowledge required to advise clients about their crypto transactions.[157] However, even if an accountant possesses the requisite knowledge, it is nonetheless difficult and costly to

153. *See, e.g.*, John Divine, *6 of the Best Cryptocurrencies to Buy Now*, U.S. NEWS (Feb. 3, 2023, 1:38 PM), [https://perma.cc/5Z8D-56AF]; Chris Davis et al., *How to Buy Dogecoin (DOGE)*, NERDWALLET (Dec. 20, 2022), [https://perma.cc/XL7Q-TWSE].

154. David Yaffe-Bellany & Ron Lieber, *Trade Your Crypto. Buy Your NFTs. And Pay Your Taxes.*, N.Y. TIMES (Mar. 26, 2022), [https://perma.cc/5FNP-QPF9].

155. CRYPTOTRADER.TAX, THE STATE OF CRYPTOCURRENCY TAX REPORTING IN 2022, at 3 (2022), [https://perma.cc/Z2HD-RK8D].

156. *Id.* at 10.

157. *Id.*

discern a crypto-trader's tax liability if that taxpayer made several transactions with crypto.[158]

Consider the hypothetical mentioned previously.[159] *A* must track her basis each time she purchases crypto or receives it for services,[160] discern which specific units of crypto she exchanged or use the FIFO method, find the relevant fair market value of the crypto she is receiving, and finally, determine the difference between the amount realized and her basis.[161] As onerous as it seems, this is a relatively simple situation, and this quagmire becomes exceedingly complex when there are multiple exchanges or wallets involved.[162] After suffering a justified headache, *A* might owe a nominal tax.

Second, if taxpayers investing in crypto find it difficult to assess their tax liabilities, it is no surprise that the IRS is struggling similarly in enforcing tax compliance and revenue collection regarding crypto transactions.[163] If the IRS allowed for crypto like-kind exchanges, it would dramatically decrease the difficulty the agency has faced by removing copious amounts of transactions from the IRS's plate and allow the agency to focus on less transactions that will ultimately result in nearly the same amount of tax being paid. Not only would this increase the IRS's ability to scout for true tax evaders that use crypto, but it would also provide a bright-line rule for taxpayers that is easy to follow.

Finally, if taxpayers are free to exchange cryptos without incurring tax liability, they will inevitably be more likely to do so.

---

158. *Id.* at 9.

159. *See supra* text accompanying notes 55-64.

160. Remember, even if *A* has only purchased one crypto, Bitcoin, but she has done so multiple times, she must track her basis *with each purchase*. Similarly, if *A* is compensated with Bitcoin multiple times, she must track her basis in each separate receipt of Bitcoin.

161. *See Frequently Asked Questions on Virtual Currency Transactions, supra* note 58.

162. *See* CRYPTOTRADER.TAX, *supra* note 155, at 9.

163. Lynn Mucenski Keck, *How the IRS Is Looking for Its Share of Cryptocurrency and NFT Growth*, FORBES (Feb. 22, 2022, 9:00 AM), [https://perma.cc/5TCT-QMVN] (discussing IRS enforcement actions and "John Doe" summonses and how the agency has sent thousands of compliance letters to taxpayers for failure to report crypto transactions). Also, consider that if a taxpayer purchases a cup of coffee from Starbucks with Bitcoin, that taxpayer must determine if she has gain or loss on the transaction because it is an exchange of property for goods. *See* Lionel Laurent, *Bitcoin at Starbucks Is More Meme Than Money*, BQ PRIME (Nov. 17, 2021, 7:40 AM), [https://perma.cc/3JUB-VYQT]. This is just another example of how difficult and tedious it is to track crypto transactions.

This will foster the evolution and development of a technology that has the potential to revolutionize the way sovereigns and institutions contemplate currency. If security, privacy, speed, access to information, globalization, and more are the fruits of this technology, U.S. tax policies should not stunt the growth of crypto when it is in its infancy. Naturally, the taxation policy in favor of hindering crypto transactions is preventing further adoption of this technology by institutions, cautious citizens, and taxpayers that do not believe it is worth the frustration that comes with assessing tax liability.[164]

CONCLUSION

By adopting crypto-friendly policies, the government could modernize the way individuals and institutions think of currency. The financial security, privacy, and freedom that crypto advocates champion is only the start to what this technology could provide to American citizens, the Nation itself, and the world as a whole. By classifying crypto as property,[165] the IRS has created a complex scheme that is the cause of tax attorneys' and accountants' nightmares. Thus, at the very least, Congress and the IRS should allow crypto swaps to qualify as § 1031 exchanges. While this may not be a perfect nor permanent solution, it adequately balances the interests between revenue collection and taxpayer freedom.

First, this policy would better align with the Supreme Court's definition of what constitutes materially different properties. Second, the congressional justifications for like-kind exchanges are equally present in the context of crypto as they are in real estate transactions. Third, allowing crypto like-kind exchanges would relieve the administrative burden in enforcing compliance and the burden taxpayers face come Tax Day. Finally, investors would not be as skeptical of crypto, and as a consequence, the technology would more quickly develop to a

---

164. *See* Rob Garver, *Crypto Tax Compliance Remains Minefield as IRS Leaves Key Questions Unresolved*, COINDESK (Feb. 24, 2022, 9:38 AM), [https://perma.cc/JD85-2CPS].

165. To be clear, there may not be a better way to classify crypto. Maybe, in the years to come, a whole section of the Code could be dedicated to such assets to take into account their uniqueness.

point where crypto's ultimate goals are met: the streamlining of existing financial architecture[166] and "putting the power and responsibility in the currency holders' hands."[167]

166. Jake Frankenfield, *Cryptocurrency Explained with Pros and Cons for Investment*, INVESTOPEDIA (Feb. 4, 2023), [https://perma.cc/FT3P-JRUU].

167. *What Is Cryptocurrency and Blockchain?*, AFS FIN. GRP. (July 30, 2021), [https://perma.cc/F8KD-RECS].

THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS: PAST, PRESENT, AND FUTURE

Samantha Doss*

I. INTRODUCTION

In 2018, the United States Department of Agriculture (USDA) proposed replacing much of the Supplemental Nutrition Assistance Program (SNAP) with "America's Harvest Box," a program that would directly distribute a package of non-perishable food items to low-income families.[1] The proposal was met with intense controversy. Many hunger advocates, grocery retailers, and former government officials spurned the idea, citing logistics challenges, nutrition concerns, and stigmatization associated with a direct distribution system.[2] However, a few Indigenous advocates were quick to point out that a direct commodity distribution system has been in place in the United States for generations, often overlooked due to its singular audience: Native Americans living on reservations.[3]

---

* J.D. and M.P.A. Candidate, 2024; Note & Comment Editor, *Arkansas Law Review*. My deepest gratitude to Erin Parker for her guidance, time, and wealth of knowledge. I also owe many thanks to the *Arkansas Law Review* staff, who made this Comment happen.

1. Helena Bottemiller Evich, *Trump Pitches Plan to Replace Food Stamps with Food Boxes*, POLITICO (Feb. 2, 2018, 9:32 PM), [https://perma.cc/KZR8-F9P7].

2. Catherine Boudreau & Helena Bottemiller Evich, *Trump's 'Harvest Box' Plan Met with Boos*, POLITICO (Feb. 27, 2018, 8:00 AM), [https://perma.cc/4RLT-APGQ]; Glenn Thrush, *Trump's 'Harvest Box' Isn't Viable in SNAP Overhaul, Officials Say*, N.Y. TIMES (Feb. 13, 2018), [https://perma.cc/P62V-4A9B].

3. *See* Tristan Picotte, *SNAP/EBT or America's Harvest Box?*, P'SHIP WITH NATIVE AMS. (May 15, 2018), [https://perma.cc/9H6E-6QAX]. For this Comment, language and term usage follows the guidance outlined by the Native American Journalists Association, IllumiNative's Reclaiming Native Truth Allies Guide, and the National Congress of American Indians. *Reporting and Indigenous Terminology*, NATIVE AM. JOURNALISTS ASS'N, [https://perma.cc/X7HF-TVJ5] (last visited Nov. 17, 2022); FIRST NATIONS DEV. INST. & ECHO HAWK CONSULTING, CHANGING THE NARRATIVE ABOUT NATIVE AMERICANS: A GUIDE FOR ALLIES (2018), [https://perma.cc/UA3R-C9LJ]; NAT'L CONG. OF AM. INDIANS, TRIBAL NATIONS AND THE UNITED STATES 11 (2020), [https://perma.cc/H6L9-7PRC]. Following this guidance, the term Native American is used

Often colloquially referred to as "commods,"[4] the Food Distribution Program on Indian Reservations (FDPIR) is a commodity food program that directly distributes monthly packages of food to low-income Native Americans.[5] FDPIR is unique because it is only available to people who live on or near a reservation.[6] Although at first glance this appears to be a narrow population, the Native American communities FDPIR serves are as diverse as they are numerous. There are currently 574 federally recognized Tribal entities in the United States.[7] The reservations and trust land areas that make up Indian Country are approximately 56.2 million acres.[8] The Navajo Nation alone would be the fortieth largest state in the country.[9] Indian Country is also growing rapidly. Between 2000 and 2010, the Native American population grew at almost twice the rate of the rest of the country.[10] It is critical to understand that Indian Country is not a monolith, however. Tribal nations are as sovereign and diverse as any other nation. Yet, the collective history of

---

to refer to all Native people of the United States and its territories. The term Indian Country is used for both its legal definition under Title 18 and its popular definition to describe areas with Tribal jurisdiction and Native American populations. 18 U.S.C. § 1151. Any variance from this guidance will be out of respect for how individuals and groups refer to themselves.

4. Megan Mucioki et al., *Thinking Inside and Outside the Box: Local and National Considerations of the Food Distribution Program on Indian Reservations (FDPIR)*, 57 J. RURAL STUD. 88, 89 (2018).

5. U.S. DEP'T OF AGRIC., FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS 1 (2020), [https://perma.cc/79FM-EMBN].

6. KENNETH FINEGOLD ET AL., TRIBAL FOOD ASSISTANCE: A COMPARISON OF THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) AND THE SUPPLEMENTAL NUTRITION ASSISTANCE PROGRAM (SNAP) 4 (2009), [https://perma.cc/SW4M-U8EX]. FDPIR is also available to approved service areas in Oklahoma, as traditionally reservation boundaries were not recognized in the state. *Id.* The recent Supreme Court decision in *McGirt v. Oklahoma*, recognizing Oklahoma reservation boundaries, could impact FDPIR, but this is beyond the scope of this Comment. 140 S. Ct. 2452, 2462 (2020).

7. Indian Entities Recognized by and Eligible to Receive Services from the United States Bureau of Indian Affairs, 86 Fed. Reg. 7554 (Jan. 29, 2021) (notice).

8. CRYSTAL ECHO HAWK, JANIE HIPP & WILSON PIPESTEM, FEEDING OURSELVES: FOOD ACCESS, HEALTH DISPARITIES, AND THE PATHWAYS TO HEALTHY NATIVE AMERICAN COMMUNITIES 21 (2015), [https://perma.cc/QX4Z-FXD4].

9. NAT'L CONG. OF AM. INDIANS, *supra* note 3, at 10.

10. KATHRYN L.S. PETTIT ET AL., U.S. DEP'T OF HOUS. & URB. DEV., CONTINUITY AND CHANGE: DEMOGRAPHIC, SOCIOECONOMIC, AND HOUSING CONDITIONS OF AMERICAN INDIANS AND ALASKA NATIVES, at ix (2014), [https://perma.cc/PST7-YG5D].

colonialism and displacement creates contemporary challenges that many Tribal nations share.

Because of the unique scope of the program, FDPIR has potential as an effective tool to address health disparities, alleviate rural food deserts, and increase Tribal sovereignty. Largely a result of colonialism and the disruption of traditional foodways, nutrition-related health problems are twice as prevalent in Native American populations when compared to the rest of the country.[11] This challenge is now so severe that one social scientist from the Osage Nation considers diabetes the final genocide of Native Americans.[12] Perhaps unsurprisingly, the U.S. Commission on Civil Rights found that out of all of the groups in the United States, Native Americans have the most severe unmet health needs.[13]

High levels of food insecurity and lack of food access contribute to these unmet health needs.[14] For example, lack of a vehicle is a significant barrier to accessing healthy food in Indian Country, as only 25% of people living in Tribal areas in 2014 were within walking distance of a supermarket, compared to 58% of all Americans.[15] This is due in part to the low population density and low resident incomes of many reservations, which can be unattractive to large food retailers.[16] As a result, most, if not all, of Indian Country is a retail food desert.[17] This can lead to smaller convenience or fast-food retailers becoming the primary food source for many rural Native American communities.[18]

---

11. PHILLIP KAUFMAN ET AL., U.S. DEP'T OF AGRIC., MEASURING ACCESS TO HEALTHFUL, AFFORDABLE FOOD IN AMERICAN INDIAN AND ALASKA NATIVE TRIBAL AREAS, at iii (2014), [https://perma.cc/62ZV-4KGR]; *see also* ECHO HAWK ET AL., *supra* note 8, at 18-20, 25, 30-31.

12. ECHO HAWK ET AL., *supra* note 8, at 31-32.

13. *Id.* at 25.

14. Food security and food access are closely linked. The USDA defines food security as everyone in a household always having enough food to maintain an active, healthy lifestyle. *Food Security in the U.S.*, U.S. DEP'T OF AGRIC. ECON. RSCH. SERV. (Oct. 17, 2022), [https://perma.cc/37XX-MWDJ]. In contrast to food security, food access is primarily a question of healthy food accessibility, including individual and neighborhood resources affecting accessibility. *Food Access Research Atlas*, U.S. DEP'T OF AGRIC. ECON. RSCH. SERV. (Oct. 20, 2022), [https://perma.cc/VS56-WTG3].

15. KAUFMAN ET AL., *supra* note 11, at i, 27.

16. *Id.* at 5.

17. ECHO HAWK ET AL., *supra* note 8, at 33.

18. KAUFMAN ET AL., *supra* note 11, at 27.

Reliance on these retailers in turn can lead to poorer health outcomes and greater food insecurity.[19] COVID-19 both exposed and exacerbated food insecurity and the presence of food deserts in Indian Country.[20]

Despite FDPIR's potential to address many of the challenges facing Indian Country, both the program and the communities it serves are often left out of larger discussions of nutrition services, food security, and agriculture.[21] This is in part due to the pervasive invisibility and underreporting that Indian Country continues to face across multiple sectors.[22] The 2018 Reclaiming Native Truth study, the largest public opinion research project undertaken in Indian Country to date, found invisibility to be both "the modern form of bias against Native Americans" and "one of the biggest barriers Native peoples face in advocating for [T]ribal sovereignty, equity[, and] social justice."[23] The invisibility Native American communities face extends to legal academia and, more critically, to Congress, as "most [Congress] members have little knowledge of Native issues and rely heavily on peers

---

19. ECHO HAWK ET AL., *supra* note 8, at 36-37.

20. Kat Eschner, *COVID-19 is Exposing the Food Deserts Around Native American Reservations*, POPULAR SCI. (July 13, 2020, 2:00 PM), [https://perma.cc/C9P2-EM8B].

21. Native American populations are not regularly included in the USDA's annual food security analysis, which informs federal decision-making around food security. Valarie Blue Bird Jernigan et al., *Food Insecurity Among American Indians and Alaska Natives: A National Profile Using the Current Population Survey-Food Security Supplement*, 12 J. HUNGER & ENV'T NUTRITION 1, 7 (2017). *See generally* ALISHA COLEMAN-JENSEN ET AL., U.S. DEP'T OF AGRIC., HOUSEHOLD FOOD SECURITY IN THE UNITED STATES IN 2020 (2021), [https://perma.cc/CL6B-ZYPX] (tracking food security for some racial and ethnic groups but not Native American populations); ECHO HAWK ET AL., *supra* note 8, at 38-39 (noting the lack of good data on food systems in Indian Country). This is particularly striking because food insecurity is also linked to the availability and effectiveness of federal programs. Jernigan et al., *supra*, at 1, 2, 5-7 (suggesting the lack of tribal-specific resources as a contributing factor to higher rates of urban Native American food insecurity).

22. *See, e.g.*, Lizzie Wade, *COVID-19 Data on Native Americans Is 'a National Disgrace.' This Scientist Is Fighting to Be Counted*, SCI. (Sept. 24, 2020), [https://perma.cc/6CMM-RKHS]; Graham Lee Brewer, *Native Americans Are Under-Reported in Opioid Overdose Data*, HIGH COUNTRY NEWS (Dec. 21, 2018), [https://perma.cc/CEX8-C9EH]; Cynthia-Lou Coleman & Jackleen de La Harpe, *It's Not Easy to Cover Indian Country. Here's Why You Must.*, POYNTER (Sept. 22, 2021), [https://perma.cc/75LR-9VCQ].

23. *United South and Eastern Tribes Celebrates Native American Heritage* Month, NAT'L INDIAN HEALTH BD., [https://perma.cc/W6Z2-TK49] (last visited Nov. 19, 2022); *see also* FIRST NATIONS DEV. INST. & ECHO HAWK CONSULTING, RECLAIMING NATIVE TRUTH: RESEARCH FINDINGS (2018), [https://perma.cc/8A77-ZTS7].

with greater interest and expertise when casting votes."[24] This lack of knowledge has shaped federal policy responses in Indian Country for generations.

Many of the problems facing FDPIR today are rooted in the fraught history of government nutrition assistance and intervention in Indian Country over the last century. While FDPIR is the program's modern name, the roots of this system, and many of the challenges the program continues to face, date back to colonization. This Comment traces these roots, looking at the history of commodity distribution in Indian Country to better understand its present, and perhaps create a better future for the thousands of participants who rely on it every day.

II. PAST: FROM RATIONS TO REGULATION

The history of commodity foods in Indian Country is rooted in colonization. The violent displacement of Native communities not only disconnected people from their homelands, it disconnected people from traditional food systems.[25] This practice was intentional and violent.[26] Early writings from the foundational years of the United States, including from the first President, show the separation of Native American communities from their food systems was part of a broader plan of assimilation, if not outright eradication.[27] In the absence of traditional food systems, and often any food systems at all, people were forced to rely on rations from the federal government for survival.[28] In the mid-1800s, the new Office of Indian Affairs, no longer housed under the War Department, became responsible for distributing these rations to newly created reservations to prevent widespread starvation.[29] While these rations were originally intended to be temporary, the barren landscape of many relocation areas and lack of other governmental assistance led to their permanence in both

24. FIRST NATIONS DEV. INST. & ECHO HAWK CONSULTING, *supra* note 23, at 8.

25. ECHO HAWK ET AL., *supra* note 8, at 30.

26. Andrea Freeman, *Unconstitutional Food Inequality*, 55 HARV. C.R.-C.L. L. REV. 840, 858-60 (2020).

27. *Id.* at 858 (noting President Washington's plan to "ruin their crops on the ground and prevent them planting more").

28. ECHO HAWK ET AL., *supra* note 8, at 30.

29. Freeman, *supra* note 26, at 859.

Native American policy and culture by the end of the nineteenth century.[30] Thus, although it was housed in different departments and underwent policy changes over the next two centuries, FDPIR is rooted in this bedrock of rations, violent colonization, and displacement.

In the early twentieth century, the United States began to experience agricultural surpluses, which the dramatic economic changes of the 1930s only increased.[31] In 1935, Congress amended the Agricultural Adjustment Act of 1933 to address this growing surplus.[32] Section 32 of the amendments appropriated 30% of customs receipts for the Secretary of Agriculture to use to "encourage the domestic consumption of such commodities or products . . . by the payment of benefits or indemnities or by other means."[33] This amendment paved the way for commodity feeding programs to grow as part of the USDA, and grow they did. By 1938, more than $54 million in surplus food was being distributed to low-income American families each year by the Federal Surplus Commodities Corporation, a non-profit organization charged with domestic distribution of surplus food purchased with Section 32 money.[34]

Already, commodity distribution was experiencing many of the logistical challenges that FDPIR still experiences today. In a 1939 overview of the program, then Secretary of Agriculture Henry Wallace noted that the Department "had complaints of waste because families suddenly received more food than could be kept without spoiling."[35] Other difficulties included unpredictability, the creation of informal markets, producer complaints about buying prices, and concerns about "morale" and social harm to recipients.[36]

---

30. Dana Vantrease, *Commod Bods and Frybread Power: Government Food Aid in American Indian Culture*, 126 J. AM. FOLKLORE 55, 56-57 (2013). *See generally* William Least Heat-Moon, *A Stark Reminder of How the U.S. Forced American Indians into a New Way of Life*, SMITHSONIAN MAG. (Nov. 2013), [https://perma.cc/GT3A-GU74] (providing an anecdote about a ration ticket in the Smithsonian collection).

31. H.R. REP. NO. 95-464, at 415 (1977).

32. Act of Aug. 24, 1935, Pub. L. No. 74-320, 48 Stat. 750.

33. § 32, 48 Stat. at 774.

34. H.R. REP. NO. 95-464, at 713-14.

35. *Id.* at 714-15.

36. *Id.* at 715.

Despite these difficulties, the commodity distribution program was formally extended to reservations through the Agricultural Act of 1949. Section 416 of the Act authorized the Secretary of Agriculture to make commodity foods available at no cost to select groups, prioritizing delivery to the Bureau of Indian Affairs and other "local public welfare organizations for the assistance of needy Indians and other needy persons."[37] Section 416 paved the way for FDPIR. It was the first statutory authorization of a USDA commodity food program that explicitly targeted Native Americans.[38] Although far from acknowledging the history of colonization and targeted destruction of traditional foods, Section 416's implicit acknowledgement of the unique challenges and needs of Indian Country set the stage for the creation of FDPIR in the coming decades.

Food stamps, the precursor to what is today known as SNAP, have been intricately connected to FDPIR since the mid-twentieth century. As part of President Johnson's "war on poverty," the Food Stamp Act of 1964 established food stamps as an official alternative to direct distribution.[39] Although the Act made no mention of Tribal nations or Tribal administration,[40] Native Americans could participate in the food stamp program if it was available to them, just like any other citizen.[41] However, disparities in Indian Country's food stamp access dated back to the earliest pilots of the program. For example, a December 1941 map of Food Stamp Plan Areas produced by the USDA shows a marked absence of the food stamp program in Oklahoma when compared to its surrounding states.[42] Despite these disparities, the Agriculture and Consumer Protection Act of 1973 mandated every political jurisdiction in the country adopt the food stamp

37. Agricultural Act of 1949, ch. 792, sec. 416, 63 Stat. 1051, 1058 (codified as amended at 7 U.S.C. § 1431(a)).

38. The term "Indian Country" appears to be absent from the era's farm bills and formal food commodity laws.

39. H.R. REP. NO. 95-464, at 817.

40. Food Stamp Act of 1964, Pub. L. No. 88-525, § 3, 78 Stat. 703, 703 (defining "State" as "the fifty States and the District of Columbia").

41. H.R. REP. NO. 95-464, at 130.

42. *Id.* at 745.

program, including Indian Country.[43] Because the food stamp program precluded direct distribution programs, this would have meant the end of commodities on reservations.

However, less than a year later, Congress amended the 1973 Act to continue Section 32 commodity distribution, including "the family commodity distribution program on Indian reservations not requesting a food stamp program."[44] The amendment further allowed reservations that had already entered the food stamp program to reinstate commodity distribution and excused reservations as a political jurisdiction from the food stamp adoption mandate until 1977.[45] The 1973 Act marked the first congressional recognition of the challenges to food stamp access that many reservations faced, and it was an important precursor to the establishment of FDPIR.[46] By the end of the decade, out of 284 total reservations, 249 were participating in the food stamp program, while 35 chose to continue participating in a commodity distribution program.[47]

The 1977 Food and Agriculture Act finally resolved the food stamp commodity distribution conflict, making specific provisions for reservations. Under the Act, direct distribution was on the way out, permitted in only limited circumstances like disaster relief.[48] However, an exception was made for "[d]istribution of commodities, with or without the food stamp program," when requested by a Tribal organization.[49] Thus, Tribes became the only political entity permitted to run food stamp and distribution programs concurrently. While the 1977 Food and Agriculture Act is often credited as the creation of FDPIR, in reality it only prevented the new food stamp program from eclipsing an existing commodity distribution system under

43. *A Short History of SNAP*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV., [https://perma.cc/NVU5-P26P] (last visited Nov. 22, 2022).

44. Act of July 12, 1974, Pub. L. No. 93-347, 88 Stat. 340, 340-41.

45. § 4, 88 Stat. at 341; H.R. REP. NO. 95-464, at 131.

46. U.S. DEP'T OF AGRIC., FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS 1 (2020), [https://perma.cc/7RAU-VTPK].

47. H.R. REP. NO. 95-464, at 131.

48. Food and Agriculture Act of 1977, Pub. L. No. 95-113, sec. 1301, § 4, 91 Stat. 913, 961.

49. § 4, 91 Stat. at 961.

the USDA, which in turn was a continuance of the rationing and colonization policies from the earliest days of the nation.

The legislative history of the Food and Agriculture Act of 1977 provides some insight into why Congress singled out Indian Country for the unique dual commodity distribution and food stamp model. According to the House Committee Report on the bill, the reservation amendments were intended to be part of the overall reform objective of "facilitat[ing] the participation of the needy so that those who do need stamps do get them."[50] The Committee gave weight to hearing testimony stating that "the remote geographical location of many reservations" was the primary barrier to program participation.[51] Of sixty-two reservations surveyed for the Senate Select Committee on Nutrition and Human Needs, only twenty-two reported having a food store within fifty miles.[52] The Navajo Nation reported the estimated distance between food stores and the farthest homes on the reservation to be 300 to 400 miles.[53] While some reservations were closer to urban centers, the Senate Committee stated that it "believe[d] that the tribal organization [was] best equipped to make [the] decision" between food stamps, commodity distribution, or concurrent programs.[54]

Despite this stated belief in the ability of Tribes to decide which program is most appropriate for their citizens, the final language of the Act contained significant limits on Tribal sovereignty, the impacts of which are still felt today. First, individual reservation households were explicitly prevented from simultaneous participation in direct distribution and food stamps.[55] Second, there were explicit barriers to Tribal administration of both programs. Under the 1977 Act, administration of the commodity distribution program still rested with the state government, not with the Tribe, unless the Secretary of Agriculture made a "determin[ation] that the tribal

50. H.R. REP. NO. 95-464, at 2.
51. *Id.* at 132.
52. S. REP. NO. 95-180, at 126 (1977); H.R. REP. NO. 95-464, at 133.
53. H.R. REP. NO. 95-464, at 132.
54. S. REP. NO. 95-180, at 129.
55. Food and Agriculture Act of 1977, Pub. L. No. 95-113, sec. 1301, § 4, 91 Stat. 913, 961.

organization [was] capable of effectively and efficiently administering such distribution."[56] There were even more barriers to Tribal administration of the food stamp program. Under the Act, the Secretary of Agriculture had to both determine that the State failed to properly administer the program and that the Tribal organization was capable of doing so in the State's place, "in light of the distance of the reservation from State . . . certification and issuance centers" and other factors like Tribal fiscal organization.[57] Thus, the presumption was strongly against Tribal administration of the food stamp program, although states were required to engage in good faith consultation and "implement the program in a manner that [was] responsive to the needs of the Indians on the reservation."[58] These limits to dual participation and Tribal administration are still creating challenges more than four decades later.

The modern FDPIR program was officially established by federal regulation in 1979. A 1978 proposed rule suggested administering the Food Distribution Program as similarly to the food stamp program as possible, both to achieve uniformity and efficiency and because "the underlying purposes of [the programs] are closely related."[59] However, the necessity of special rules regarding the unique dual nature of enrollment quickly became apparent. Under the 1977 Food and Agriculture Act, Indian Tribal Organizations (ITOs) could choose to operate a food stamp program, a food distribution program, or both programs within a reservation boundary.[60] There was concern that if an ITO administered FDPIR only, reservation residents who were not part of the Tribe would become ineligible for both programs.[61] To avoid this, the agency's final rulemaking expanded baseline FDPIR eligibility to *all* households on a

56. § 4, 91 Stat. at 961.

57. § 11, 91 Stat. at 970.

58. § 11, 91 Stat. at 970.

59. Food Stamp and Food Distribution Programs; Indian Reservations, 43 Fed. Reg. 57798, 57799 (proposed Dec. 8, 1978) (to be codified at 7 C.F.R. pts. 271, 281, 283).

60. Food and Agriculture Act of 1977, Pub. L. No. 95-113, sec. 1301, § 4, 91 Stat. 913, 961.

61. Food Stamp and Food Distribution Programs on Indian Reservations, 44 Fed. Reg. 35904, 35911 (June 19, 1979) (to be codified at 7 C.F.R. ch. II).

reservation, not just those who were part of the Tribe.[62] Non-Tribal residents were also permitted to enroll in an off-reservation food stamp program if available, while on-reservation Tribal citizens were not.[63] Conversely, Tribal citizens who resided off-reservation but nearby could still participate in FDPIR.[64] ITOs could also split reservations into smaller areas with individual program designations for FDPIR, food stamps, or concurrent areas.[65] These basic parameters still govern the program today.

While the structure and eligibility rules remain similar today, the stated goals of FDPIR have evolved. The original commodity distribution program's purpose was twofold: to provide an outlet for domestically produced agricultural products and to provide nutritious food to Native American households.[66] With the decrease in domestic surpluses and an increased national concern for hunger in the 1970s, this second purpose became dominant.[67] The Food and Agriculture Act of 1977 states that distribution programs to reservations "shall improve the variety and quantity of commodities supplied to Indians in order to provide them an opportunity to obtain a more nutritious diet."[68] The Act sought to meet this nutrition goal by aligning the type and quantity of commodity foods with the thrifty food plan used to calculate food stamp benefits.[69] However, the USDA avoided specificity on food quantity and quality and insisted the program was supplementary only.[70] In response, the Department received almost thirty comments on its proposed rule asking for greater specificity and guarantees related to the contents of the food

62. *Id.*

63. *Id.* at 35912.

64. *Id.*

65. *Id.*

66. *See supra* text accompanying notes 31-35, 50.

67. Food and Agriculture Act of 1977, Pub. L. No. 95-113, sec. 1301, § 2, 91 Stat. 913, 958.

68. Sec. 1304, § 4(a).

69. H.R. REP. NO. 95-464, at 134 (1977).

70. The proposed regulation establishing FDPIR stated that the Department of Agriculture would offer "a variety and quantity of commodities for Indian households such that the commodity package represents an acceptable alternative" to food stamps only, further noting the joint conference committee report specifically stated that the commodity distribution program is not intended to provide a fully adequate nutritional diet alone. Food Stamp and Food Distribution Programs; Indian Reservations, 43 Fed. Reg. 57798, 57798 (proposed Dec. 8, 1978) (to be codified at 7 C.F.R. pts. 271, 281, 283).

package itself.[71] The USDA rebutted these requests, concluding "that a guarantee expressed in such literal measures would be impracticable and inappropriate."[72] Beyond budget limits and logistic concerns, the Department's primary reasoning was that it wanted the food packages to be responsive and "tailor[ed] . . . as closely as practicable to individual [T]ribal preferences."[73] The modern implementation of FDPIR shows that the reality of the program would sharply diverge from this goal of responsiveness.

III. PRESENT: CHALLENGES AND CHANGES

FDPIR today looks very different from its first iteration in 1979, due in large part to advocacy from Indian Country following the first decade of the program. After the Agriculture Act of 1977 was passed, thirty-seven people testified at USDA hearings in October 1977 about the reservation-specific provisions of the Act.[74] Comments from the Tohono O'odham Nation[75] in response to the 1978 proposed rule ensured that state agencies were required to submit FDPIR operation plans to the relevant ITO and incorporate ITO comments before submitting a final plan to the Food and Nutrition Service (FNS).[76] Comments from the Navajo Nation increased the level of administrative funding from 50% to 75% with the option of increasing the amount up to 100% based on "compelling justification."[77] At the inception of the program in 1979, commenters were already asking for better labeling, a reduction in container size, and the addition of fresh and frozen foods.[78] Some of these requests were the same challenges that had been identified as early as 1939 in

---

71. Food Stamp and Food Distribution Programs on Indian Reservations, 44 Fed. Reg. 35904, 35922 (June 19, 1979) (to be codified at 7 C.F.R. ch. II).

72. *Id.*

73. *Id.*

74. Food Stamp and Food Distribution Programs; Indian Reservations, 43 Fed. Reg. at 57798.

75. The Tohono O'odham Nation was known as the Papago Tribe in 1977. *Tohono O'odham History*, TOHONO O'ODHAM NATION, [https://perma.cc/UQQ8-RUJA] (last visited Nov. 22, 2022).

76. Food Stamp and Food Distribution Programs on Indian Reservations, 44 Fed. Reg. at 35906, 35907.

77. *Id.* at 35921.

78. *Id.* at 35924.

the reports that led to the creation of the food stamp program.[79] Advocates continued to flag problems and push for solutions for FDPIR over the first decade of the program, leading to a series of dramatic changes in the 1990s and early 2000s.

In 1987, Congress passed the Commodity Distribution Reform Act requiring FNS and FDPIR administrators to collect information about commodity foods and participant feedback, which up until that point had been largely unaddressed due to the FNS's policy of reducing the administrative burden for the FDPIR program.[80] Partially in response to this call for information, the first national study of FDPIR was conducted in 1990.[81] While some of the study's conclusions on participant satisfaction with the food package conflict with other reports,[82] the study highlighted several issues with administration, particularly distribution, selection, and availability.[83] Some of the key challenges facing FDPIR today include the quality and cultural responsiveness of the food package, distribution and procurement at the local level, and tensions between federal paternalism and Tribal sovereignty.

A. Food Package Content: Quality and Selection

The actual content and quality of the food packages themselves have been one of the key challenges to successful implementation of FDPIR. In 1990, these problems came to a head at a hearing of the of the House Select Committee on Hunger at the Standing Rock Reservation in North Dakota, with

---

79. *See supra* notes 35-36 and accompanying text.

80. Commodity Distribution Reform Act and WIC Amendments of 1987, Pub. L. No. 100-237, 101 Stat. 1733 (1988).

81. CHARLES L. USHER ET AL., U.S. DEP'T OF AGRIC., EVALUATION OF THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) ES-1 (1990), [https://perma.cc/4QDG-FLRN].

82. One of the report's conclusions is that "[p]rogram participants express strong positive preferences for almost all commodity food items." *Id.* at ES-12. However, this conflicts with other reports of widespread community dissatisfaction during this time and uses consumer preference to gloss over the nutritional deficiencies of commodity foods. *See Standing Rock Sioux Reservation: A Case Study of Food Security Among Native Americans: Hearing Before the H. Select Comm. on Hunger*, 101st Cong. 22-23 (1990) (statement of Charles "Red" Gates, FDPIR Program Dir., Standing Rock Sioux Tribe); Mucioki et al., *supra* note 4, at 89; *see also* ECHO HAWK ET AL., *supra* note 8, at 45.

83. USHER ET AL., *supra* note 81, at ES-6, ES-7.

Congressmen Byron Dorgan (North Dakota), Chairman Tony P. Hall (Ohio), and Eni Faleomavaega (American Samoa).[84] In his role as FDPIR program director for the Standing Rock Sioux Tribe, Charles "Red" Gates gave memorable testimony about the quality of the FDPIR foods:

> Mr. GATES. At this time, I would like to open a couple of these cans, and show you what some of the people are forced to eat because that is all they have.
>
> Mr. HALL. Now, what is this? Is this canned meat?
>
> Mr. GATES. This is canned beef, supposedly. We get reports saying that it is supposed to be real good for us.
>
> Mr. FALEOMAVAEGA. Is this USDA-approved?
>
> Mr. GATES. Yes.
>
> Mr. DORGAN. This is sent through the commodities program from USDA; is that correct?
>
> Mr. GATES. Right. You can see the top of the can here, with the fat on it.
>
> Mr. HALL. Why don't you bring it up here? Why don't you take it up and show it to us? It does not look too good.
>
> Mr. GATES. No, it does not. If you dump it out in a plate, you will see blood vessels.
>
> Mr. DORGAN. Do you have a plate? Let us take a look at it. Do you have a stronger scooper?
>
> Mr. FALEOMAVAEGA. If the chairman will yield, I will wager the chairman that dog and cat food probably have more nutritional value than what this is offering.[85]

Congressman Dorgan went on to note the particular impact of the presented food given the health disparities of Indian Country, noting "it really is a disgrace to be providing that as representing nutritional commodities for people who are suffering from diabetes at the rate of 50 percent of the general population."[86] In response to questions about whether the Standing Rock Sioux Nation could ask for better quality items,

---

84. *Standing Rock Sioux Reservation: A Case Study of Food Security Among Native Americans: Hearing Before the H. Select Comm. on Hunger*, 101st Cong. 1 (1990).

85. *Id.* at 22.

86. *Id.* at 23.

Gates testified that the Tribe's suggestions had "fall[en] on deaf ears" despite listing complaints on the required twice annual food acceptability report for over eight years.[87] Gates also touched on the importance of traditional foods, noting that although higher quality protein like albacore tuna is available, "the Indians on Standing Rock are not partial to fish."[88] Beyond the visceral example of the canned meat, Gates also highlighted the high sodium contents of canned vegetables, lack of labeling, and lack of nutritional education.[89] Gates continues to be interviewed to this day about this impactful testimony, as well as his contributions as a FDPIR program director for more than thirty years.[90]

In large part due to advocates like Gates, the quality and selection of the FDPIR food package has slowly improved, primarily in the last decade. In 1989, the National Association of Food Distribution Programs on Indian Reservations (NAFDPIR) was formed.[91] This group of Tribal representatives works with the USDA's FNS to promote advocacy, policy, and legislative changes relating to FDPIR.[92] In 2002, NAFDPIR passed a resolution requesting that FNS convene a group to regularly review the contents of the FDPIR food package.[93] The resulting FDPIR Food Package Review Work Group has eighteen voting members and multiple non-voting members, made up of NAFDPIR members, ITO representatives (including Charles "Red" Gates), federal and Tribal health professionals, and USDA

---

87. *Id.*

88. *Id.*

89. *Standing Rock Sioux Reservation: A Case Study of Food Security Among Native Americans: Hearing Before the H. Select Comm. on Hunger*, 101st Cong. 22-23 (1990).

90. Jennifer Churchill, *A 30-Year Journey to Bring Back Bison*, INDIAN GIVER (Nov. 29, 2019), [https://perma.cc/8YQW-MGX6].

91. Andi Murphy, *After a Fraught History, Some Tribes Finally Have the Power to Rethink 'Commodity Foods,'* CIV. EATS (Nov. 1, 2021), [https://perma.cc/TLA8-6CLC].

92. *Id.*

93. *FDPIR Food Package Review Work Group*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (Sept. 19, 2022), [https://perma.cc/FL6M-LWKY].

and FNS staff.[94] Since 2002, the group regularly meets to review and revise the food package.[95]

One of the biggest changes to the program in recent years was the advent of fresh produce. As of 2017, almost all ITOs participate in the Fresh Fruits and Vegetable Program operated through the Department of Defense.[96] Despite this, some FDPIR participants find the proportion of pasta and grains is still too large, leading to stockpiling.[97] The Food Package Review Workgroup has acknowledged this complaint and advocated for reduced grains and increased protein and vegetables in their November 2021 meeting.[98]

The second big change has been the introduction of culturally relevant foods. Traditional foods have been requested as part of the FDPIR food package since the program's inception in 1979.[99] It took almost forty years to make this request a reality. A 1996 pilot introduced ground bison to the food package, but it was not a permanent addition.[100] The creation of the Food Package Review Work Group in the early 2000s helped lead to congressional authorization for traditional and locally grown foods in the 2008 Farm Bill.[101] Despite this authorization, it took almost a decade of further advocacy for foods like blue cornmeal

94. *The Food Distribution Program on Indian Reservations: FDPIR Food Package Review Workgroup Members*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (Sept. 2022), [https://perma.cc/UYG4-AMH2]; *The Food Distribution Program on Indian Reservations: Food Package Review Work Group—Goals, Structure, and Review Process*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (Sept. 25, 2015), [https://perma.cc/4TYG-YDM7].

95. *Food Package Review Work Group—Goals, Structure, and Review Process*, *supra* note 94.

96. *USDA DOD Fresh Fruit and Vegetable Program: The Basics for FDPIR*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (June 26, 2017), [https://perma.cc/3UZA-JMJ3]; *see also* discussion *infra* Part IV.

97. Mucioki et al., *supra* note 4, at 92.

98. FOOD & NUTRITION SERV, U.S. DEP'T OF AGRIC., FDPIR FOOD PACKAGE REVIEW WORKGROUP STRATEGIC PLANNING MEETING MINUTES 2 (2021), [https://perma.cc/M6VS-SPG3].

99. Food Stamp and Food Distribution Programs on Indian Reservations, 44 Fed. Reg. 35904, 35912 (June 19, 1979) (to be codified at 7 C.F.R. ch. II) (noting FNS received three comments to the proposed FDPIR rule suggesting that "FNS make money available to Indian tribes so they can purchase native and cultural foods raised in their area for distribution under the program").

100. Murphy, *supra* note 91.

101. Mucioki et al., *supra* note 4, at 89-91; Food, Conservation, and Energy Act of 2008, Pub. L. No. 110-246, § 4211, 122 Stat. 1651, 1884.

and bison to become regular parts of the food package.[102] While this has been a significant victory, these foods are provided on a national scale to every ITO in the country. This inappropriately large scale creates cultural disconnect, as what is traditional for one Tribe is not necessarily traditional for another.

B. Distribution: National Systems and Contingency Plans

The national system of distribution is another challenge FDPIR faces today. Like other USDA feeding programs, FDPIR is treated as a nationwide, homogenous service despite the unique scope, history, and focus of the program.[103] To address the challenges of distributing commodities to remote and rural locations, many FDPIR programs, especially smaller programs, adopted "tailgate" distribution systems, where food packages were delivered via truck to individual communities, as opposed to having participants travel to a centralized warehouse.[104] However, this system came with its own set of issues. A 1990 study of FDPIR noted that tailgating led to reduced selection, with some "commodity items only [available] every two or three months."[105] Problems with distribution have continued to plague the program to this day.[106] Today, there are only two distribution centers, centrally located in Boise, Idaho, and Kansas City, Missouri, which service every reservation in the United States.[107] As it currently stands, by using such a national distribution and food sourcing system, the USDA ends up treating the Native American populations it serves as a monolith, despite the diverse reality of Tribal food traditions, geographies, and needs.[108]

The system of distribution also does not allow for adequate emergency preparedness. In the face of disaster, FDPIR does not

---

102. Mucioki et al., *supra* note 4, at 89-90.

103. ANDREW FISHER, BIG HUNGER: THE UNHOLY ALLIANCE BETWEEN CORPORATE AMERICA AND ANTI-HUNGER GROUPS 144 (Robert Gottlieb ed., 2017).

104. USHER ET AL., *supra* note 81, at ES-6.

105. *Id.* at ES-7.

106. *See* ECHO HAWK ET AL., *supra* note 8, at 45; Mucioki et al., *supra* note 4, at 95.

107. KARLI A. MOORE ET AL., UNIV. OF ARK. INDIGENOUS FOOD & AGRIC. INITIATIVE, OPTIMIZING DISTRIBUTION CENTER LOCATION AND DELIVERY SCHEMES FOR THE USDA'S FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) 3 (2018), [https://perma.cc/S4VX-STHQ].

108. FISHER, *supra* note 103, at 144.

have a contingency plan in place to ensure that those highly dependent on the program continue to receive service. The danger of this gap was made clear following the "nightmarish scenario" of the 2013 government shutdown, which led to extensive food spoilage and program failure.[109] As a result, Tribal leaders have been asking the USDA for a FDPIR contingency plan to continue program administration in the face of such unforeseen challenges.[110] NAFDPIR President Mary Greene Trottier again called for contingency planning in front of the House Committee on Natural Resources following the 2019 government shutdown.[111] With eerie foresight, Trottier noted, "shutdowns are not the only catastrophes that imperil our program and impact our nation's food system: so do natural disasters and commodity food shortages."[112] Just one year later, COVID-19 did just that, exposing the frailties of the U.S. food supply chain and their disproportionate impact on Native Americans and programs like FDPIR.[113] During a crisis, the current system of national distribution, as opposed to a regional system, limits both program efficiency and potential economic support for local, Native American producers.[114] Paired with the lack of contingency planning, the challenges facing the present distribution system are clear.

C. Paternalism: Administration and Consultation

Another significant challenge to FDPIR is addressing the historic pattern of paternalism associated with federal Indian Country policies. This dates back to the original 1977 Food and

109. *Oversight Hearing on "Shutdown Impacts in Indian Country" Before the H. Comm. on Nat. Res*., 116th Cong. (2019) (written testimony of Mary Greene Trottier, President, Nat'l Ass'n of Food Distrib. Programs on Indian Rsrvs.).

110. *Id.*

111. *Id.*

112. *Id.*

113. *Buying Native American: Federal Support for Native Business Capacity Building and Success: Oversight Hearing Before the S. Comm. on Indian Affs.*, 117th Cong. (2022) (written testimony of Janie Simms Hipp, Gen. Couns., U.S. Dep't of Agric.); JOEL ANDERSON & OLIVIA CHAN, CULTIVATING RESILIENCE IN INDIAN COUNTRY: AN ASSESSMENT OF COVID-19'S IMPACT ON TRIBAL FOOD SYSTEMS 1-4 (2021), [https://perma.cc/YRF8-WM4N].

114. ANDERSON & CHAN, *supra* note 113, at 32.

Agriculture Act. While the House Committee noted that the barriers to Tribal administration of commodity distribution were intentionally lower than Tribal administration of the food stamp program,[115] Tribal organizations still had to be found "effective and efficient managers" at the discretion of the Secretary of Agriculture before they could administer the program themselves.[116] After the initial FDPIR rule was proposed, several states protested their new responsibilities of administering FDPIR.[117] As a result, the USDA shifted the burden of administration to ITOs, changing the rule to be that only upon "a finding of *incapability* on the part of the ITO, would the state government be required to administer the program on behalf of the Indian tribe."[118] By 1990, the program was administered almost entirely by ITOs.[119]

While FDPIR is largely ITO-administered, other nutrition programs that are closely connected to FDPIR, such as SNAP, face more barriers to Tribal control. Section 4004 of the Agricultural Act of 2014 required a feasibility study of ITO administration of all federal nutrition assistance programs.[120] The resulting study found that "[n]early all Tribes that participated . . . expressed interest in administering [f]ederal nutrition assistance programs," as long as there was sufficient federal funding and technical assistance.[121] However, administration of a program is not the same as control of the program. For FDPIR, ITOs are limited to roles like food ordering and delivery, while most program decisions rest with the federal government.[122]

---

115. H.R. REP. NO. 95-464, at 133-35 (1977) ("The Committee was not as willing to permit [T]ribal organizations to administer the food stamp program . . . . It should be emphasized that the Secretary's designation of a [T]ribal organization as administrator of the food stamp program . . . should not be construed or interpreted in any way as an official recognition of [T]ribal sovereignty.")

116. *Id.* at 134.

117. Food Stamp and Food Distribution Programs on Indian Reservations, 44 Fed. Reg. 35904, 35913 (June 19, 1979) (to be codified at 7 C.F.R. ch. II).

118. *Id.* (emphasis added).

119. USHER ET AL., *supra* note 81, at I-2.

120. Agricultural Act of 2014, Pub. L. No. 113-79, § 4004, 128 Stat. 649, 785.

121. STEVEN GARASKY ET AL., U.S. DEP'T OF AGRIC., FEASIBILITY OF TRIBAL ADMINISTRATION OF FEDERAL NUTRITION ASSISTANCE PROGRAMS 39, 42 (2016), [https://perma.cc/YDH4-5YG9].

122. *Id.* at 166-68.

In practice, federal paternalism has led to abrupt, often unpopular, changes that catch many participants by surprise. Towards the end of the Bush administration, a workgroup of representatives from the FNS, Centers for Disease Control, and Indian Health Services unilaterally removed butter from the FDPIR food package over health concerns.[123] No Tribe was consulted, and no replacement was initially provided.[124] Participants were outraged. In a 2009 study of FDPIR, each of the seven reservations, nations, and rancherias participating in the study independently listed the removal of butter as a top complaint.[125] Gloria Goodwin of Minnesota's White Earth Indian Reservation brought the issue to the attention of the U.S. House Committee on Agriculture in a 2010 hearing to review FDPIR.[126] Despite the uproar, it took the USDA's senior advisor for Tribal relations, Janie Simms Hipp, more than a year and half to reintroduce butter into the food package.[127] This incident showcases how paternalistic federal policy in Indian Country can be disastrous when Tribal nations are not given a voice.

One solution that Indian Country advocates have pushed hard for is better use of consultation. Consultation is a recognition of the government-to-government relationship between Tribes and the federal government.[128] President Clinton's 2000 Executive Order 13175 required all agencies, including the USDA, to engage in "regular and meaningful consultation" with Tribal nations for all federal policies with Tribal implications.[129] Under President Obama,[130] as reaffirmed

---

123. *Detailed Information on the Food Distribution Program on Indian Reservations Assessment*, EXPECTMORE.GOV (Sept. 6, 2008), [https://perma.cc/8W44-MVS8]; FISHER, *supra* note 103, at 143.

124. FISHER, *supra* note 103, at 143; FINEGOLD ET AL., *supra* note 6, at 80, 126.

125. FINEGOLD ET AL., *supra* note 6, at 80, 96, 105, 113, 126, 146, 155.

126. *Hearing to Review the Food Distribution Program on Indian Reservations Before the Subcomm. on Dep't Operations, Oversight, Nutrition, and Forestry of the H. Comm. on Agric.*, 111th Cong. 42-43 (2010) (statement of Gloria Goodwin, Sec'y, Nat'l & Reg'l Bd. of Dirs., NAFDPIR).

127. FISHER, *supra* note 103, at 143.

128. Exec. Order No. 13175, 65 Fed. Reg. 67249 (Nov. 6, 2000).

129. *Id.*

130. Memorandum on Tribal Consultation, 2009 DAILY COMP. PRES. DOC. 1 (Nov. 5, 2009).

in January 2021 by the Biden administration,[131] agencies are further required to create and regularly update a plan to implement Executive Order 13175. The USDA's Tribal Consultation directive requires every agency within the USDA to "provide an opportunity for Tribes to participate in policy development to the greatest extent practicable and permitted by law."[132] Since at least 2015, Tribal citizens have called for regular consultations with the FNS solely about FDPIR.[133] After a lack of federal response, the National Congress of American Indians issued a resolution in the winter of 2016, again calling for regular FDPIR consultation.[134]

As a result of these requests, the FDPIR Tribal Leaders Consultation Working Group (TLCWG) was created to establish regular consultations with the USDA about issues related solely to FDPIR.[135] Composed of Tribal leaders, USDA deputy secretaries, the FNS Undersecretary, and the Director of Tribal Relations, the TLCWG is the only standing consultation body between the USDA and Tribal leadership.[136] The TLCWG has consulted on every aspect of FDPIR, including funding, nutrition education, distribution and delivery problems, emergency contingency planning, and demonstration projects.[137] In just a

---

131. Memorandum on Tribal Consultation and Strengthening Nation-to-Nation Relationships, 2021 DAILY COMP. PRES. DOC. 1 (Jan. 26, 2021).

132. U.S. DEP'T OF AGRIC., DR 1350-002, TRIBAL CONSULTATION, COORDINATION, AND COLLABORATION (2013), [https://perma.cc/U4QS-5WDK].

133. U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS TRIBAL LEADER CONSULTATION WORKING GROUP (Dec. 12, 2017) [hereinafter TLCWG Dec. 2017 REPORT-OUT], [https://perma.cc/9XS9-3E2B]; NAT'L CONG. OF AM. INDIANS, RES. #SPO-16-058, CALLING ON THE USDA TO INITIATE TRIBAL CONSULTATION ON ISSUES RELATED TO THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (2016), [https://perma.cc/GB52-AUVK].

134. CALLING ON THE USDA TO INITIATE TRIBAL CONSULTATION ON ISSUES RELATED TO THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS, *supra* note 133.

135. *See* TLCWG Dec. 2017 REPORT-OUT, *supra* note 133, at 1.

136. *Breaking New Ground in Agribusiness Opportunities in Indian Country: Hearing Before the S. Comm. on Indian Affs.*, 115th Cong. 17 (2018) (prepared statement of Janie Simms Hipp, Dir., Indigenous Food & Agric. Initiative, Univ. of Ark. Sch. of L.).

137. *See* U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: USDA FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) TRIBAL LEADERS CONSULTATION WORK GROUP (Feb. 13, 2020) [hereinafter TLCWG Feb. 2020 REPORT-OUT], [https://perma.cc/VZP2-96Q7]; U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR): DEMONSTRATION PROJECT FOR TRIBAL ORGANIZATIONS TRIBAL CONSULTATION

few short years, the TLCWG has made significant progress in making FDPIR more responsive to participant needs. In addition to providing significant oversight for a critical FDPIR self-determination demonstration project,[138] the TLCWG continued consultation through the COVID-19 pandemic, addressing issues surrounding CARES Act benefits, parity with SNAP, personal protective equipment procurement, and justification for budget changes.[139]

Janie Simms Hipp highlighted the significance of the TLCWG in a 2018 hearing before the U.S. Senate Committee on Indian Affairs, explaining, "We know the importance of consultation because we have seen it. It has solved longstanding problems in the FDPIR but [has] also shown USDA the power of actually getting Tribal governments in the room to problem-solve in a deeper and more meaningful way."[140] Regular consultation highlights the importance of a government-to-government relationship in the FDPIR program. It is also a critical first step towards a future of robust Tribal sovereignty.

MEETING (July 15, 2020) [hereinafter TLCWG July 2020 REPORT-OUT], [https://perma.cc/S2M8-2S9Q]; U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR): DEMONSTRATION PROJECT FOR TRIBAL ORGANIZATIONS TRIBAL CONSULTATION MEETING (Oct. 28, 2020) [hereinafter TLCWG Oct. 2020 REPORT-OUT], [https://perma.cc/5D7A-LR86]; U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: TRIBAL LEADERS CONSULTATION WORKING GROUP FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) (July 6, 2021) [hereinafter TLCWG July 2021 REPORT-OUT], [https://perma.cc/GCT8-8GPW]; U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: TRIBAL LEADERS CONSULTATION WORKING GROUP FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) (Aug. 11, 2021) [hereinafter TLCWG Aug. 2021 REPORT-OUT], [https://perma.cc/MKU4-LNCD]; U.S. DEP'T OF AGRIC., CONSULTATION REPORT-OUT: TRIBAL LEADERS CONSULTATION WORKING GROUP FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR) (Dec. 7, 2021) [hereinafter TLCWG Dec. 2021 REPORT-OUT], [https://perma.cc/U9PY-S7BB].

138. *See* discussion *infra* Section IV.A.

139. *See* TLCWG July 2020 REPORT-OUT, *supra* note 137; TLCWG Oct. 2020 REPORT-OUT, *supra* note 137; TLCWG July 2021 REPORT-OUT, *supra* note 137.

140. *Breaking New Ground in Agribusiness Opportunities in Indian Country: Hearing Before the S. Comm. on Indian Affs.*, 115th Cong. 17 (2018) (prepared statement of Janie Simms Hipp, Dir., Indigenous Food & Agric. Initiative, Univ. of Ark. Sch. of L.).

IV. FUTURE: FDPIR AND TRIBAL SOVEREGINTY

Tribal sovereignty predates the founding of the United States. In its most basic form, it represents the right of Tribal nations to be governed by their own laws, recognizing their sovereignty over Tribal land and citizens.[141] Tribal sovereignty also implicates a complex and often violent history of legal precedent.[142] Given this legal background, food sovereignty in the context of Indian Country is closely related to Tribal sovereignty. The First Nations Development Institute defines food sovereignty as "the legal ability of a tribal community to define their own food system and laws and regulations that may affect their food system."[143] For FDPIR, a more robust expression of both Tribal sovereignty and food sovereignty must include the expansion of Tribal self-governance to all federal feeding programs, elimination of the prohibition on dual enrollment of FDPIR and SNAP, and improvements in the procurement system to allow for more Tribally produced and procured foods on a smaller geographic scale.

A. Expanding 638 Authority

In 1975, just three years before the creation of FDPIR, Congress passed the Indian Self-Determination and Education Assistance Act, Public Law 93-638, which created what is commonly called "638 authority" for Tribes.[144] In its simplest terms, 638 authority allows Tribal nations to enter into contracts or compacts with certain federal agencies for responsibility over authorized programs, functions, services and activities.[145] Under a 638 contract or compact, a Tribal nation gains full control of the relevant program, while the federal government occupies a role

141. *See Tribal Sovereignty*, PAUMA BAND OF LUISEÑO INDIANS, [https://perma.cc/VNJ3-EBH6] (last visited Nov. 23, 2022); ECHO HAWK ET AL., *supra* note 8, at 20.

142. *Tribal Sovereignty*, *supra* note 141.

143. FIRST NATIONS DEV. INST., FOOD SOVEREIGNTY (2015), [https://perma.cc/5Y2Y-27Z4].

144. Indian Self-Determination and Education Assistance Act, Pub. L. No. 93-638, 88 Stat. 2203 (1975).

145. *Id.*

more akin to technical assistance.[146] Currently, 638 contracting and compacting is only permitted with the Department of the Interior and the Department of Health and Human Services.[147] While some scholars argue that 638 authority should not be conflated with absolute Tribal sovereignty, as it still relies on government contracts within the American legal system,[148] this legal mechanism provides an important practical tool for more robust self-determination. It also has a long track record of success, with almost every Tribal entity using some form of 638 authority today.[149] This authority is also an important recognition of the diversity of Tribal nations in the United States, as each contract or compact can be specifically tailored to fit a particular Tribe's needs.

The 2018 Farm Bill expanded 638 authority to the USDA for the first time through a pilot project specifically aimed at FDPIR.[150] In late 2021, the USDA awarded initial demonstration projects to eight eligible nations, allowing them to select food for the FDPIR packages and purchase directly from commercial vendors, including local Native vendors, for the first time.[151] The TLCWG was instrumental in guiding the project criteria and application process. For example, the USDA initially proposed capping the awards to only five projects, but after pushback from the TLCWG, this cap was removed.[152] The TLCWG also ensured that traditional foods would be eligible for purchase and reduced potential barriers for Native American vendors.[153]

The Intertribal Agriculture Council emphasized the relationship between the USDA 638 demonstration and

---

146. RICHELLE GROGG, CONG. HUNGER CTR., A PRIMER ON 638 AUTHORITY: EXTENDING TRIBAL SELF-DETERMINATION TO FOOD AND AGRICULTURE 5-6 (2019), [https://perma.cc/9MN8-DEUU].

147. *Id.* at 2.

148. *See* Danielle A. Delaney, *The Master's Tools: Tribal Sovereignty and Tribal Self-Governance Contracting/Compacting*, 5 AM. INDIAN L.J. 308, 309-10 (2017).

149. *See* GROGG, *supra* note 146, at 7.

150. Agriculture Improvement Act of 2018, Pub. L. No. 115-334, § 4003, 132 Stat. 4490, 4624-27.

151. *USDA Invests $3.5 Million to Provide Food Purchasing Options to Tribal Communities*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (Nov. 1, 2021), [https://perma.cc/Z8GF-VZZW].

152. TLCWG Oct. 2020 REPORT-OUT, *supra* note 137.

153. *Id.*

self-determination, stating the project was "an important acknowledgement of Tribal sovereignty that opens the door to food purchasing decisions that allow for more traditional, Tribally grown, local and regionally produced foods."[154] Overall, the demonstration project has been a success on this front. Every participating Tribe made a purchase from a Native American producer.[155] For example, as a result of the project, the Oneida Nation and Menominee Indian Tribe were able to offer traditional foods sourced from the Oneida Tribe, the Fond du Lac Reservation, and the Red Cliff Band of Lake Superior Chippewa.[156]

Despite this success, participating ITOs are already finding that producer growth is being limited by the project being viewed as a temporary "pilot."[157] Responding to this limitation, in a December 2021 consultation, the TLCWG recommended making the demonstration project permanent and providing full funding, as well as moving "from self-determination to self-governance with [the] ability to decide unique needs going beyond supplanting current food items."[158] The 2023 Farm Bill provides a perfect opportunity to confer 638 authority to the USDA in full.

While Tribes have administered some federal programs for decades, Indian Country advocates and scholars have argued that Tribal control of decision-making at all levels, such as the control provided by 638 authority, provides a more complete expression of Tribal sovereignty.[159] This is especially clear for FDPIR, which has been long administered by ITOs but still has a decades-long history of challenges stemming from ultimate federal control of the program.[160] With the 2023 Farm Bill just around the corner, expanding sovereignty-centered policies like the 638 demonstration project will be essential to the continued growth

---

154. *Exploring the Possibilities: FDPIR 638 Self-Determination Demonstration Project*, INTERTRIBAL AGRIC. COUNCIL (Jan. 29, 2021), [https://perma.cc/UCA6-YZUS].

155. *Buying Native American: Federal Support for Native Business Capacity Building and Success: Oversight Hearing Before the S. Comm. on Indian Affs.*, 117th Cong. (2022) (written testimony of Janie Simms Hipp, Gen. Couns., U.S. Dep't of Agric.).

156. *Id.*

157. TLCWG Dec. 2021 REPORT-OUT, *supra* note 137.

158. *Id.*

159. *See, e.g.*, Mucioki et al., *supra* note 4, at 89, 96; GROGG, *supra* note 146, at 2.

160. *See* discussion *infra* Section III.C.

and success of FDPIR. Tribes have been calling for a full amendment to the Indian Self-Determination and Education Assistance Act to allow 638 contracting for all federal nutrition programs.[161] Pairing this legal authority with federal funding for administrative costs will be critical.[162] Opportunities should also be explored for smaller Tribes, for whom full 638 contracts may be administratively infeasible.[163]

B. Allowing Dual Enrollment and Tribal Administration of SNAP

SNAP and FDPIR are intricately connected. Because the modern version of FDPIR was created as an alternative to SNAP, it is intended to mirror the program in terms of benefits provided.[164] However, there is not always parity between the two programs. During the COVID-19 pandemic, emergency benefits for SNAP increased, but this increase was not automatically reflected for FDPIR participants, causing some FDPIR participants to switch over to SNAP out of necessity.[165] The TLCWG is still investigating this parity gap.[166]

Allowing enrollment in both programs could also improve food access for Native Americans who need support the most. The 1977 Act establishing FDPIR mandated that "there shall be no distribution of federally donated foods to households under the authority of any law" in food stamp jurisdictions, except for temporary emergency situations and the commodity supplemental food program, which is distinct from FDPIR.[167] The Act goes on to carve out a separate exception specifically for

161. GARASKY ET AL., *supra* note 121, at 74.

162. *Id.* at 39; *see also* NAT'L CONG. OF AM. INDIANS, TRIBAL FOOD SOVEREIGNTY: INDIAN COUNTRY'S POLICY RECOMMENDATIONS FOR THE FEDERAL GOVERNMENT (2021), [https://perma.cc/W6YM-YYZ7].

163. GARASKY ET AL., *supra* note 121, at 39-40.

164. *See supra* note 59 and accompanying text.

165. TLCWG Dec. 2021 REPORT-OUT, *supra* note 137, at 4.

166. *Id.*

167. Food and Agriculture Act of 1977, Pub. L. No. 95-113, sec. 1301, § 4, 91 Stat. 913, 961.

Tribal organizations,[168] the foundation for FDPIR, but notes that the Secretary of Agriculture "shall not approve any plan for such distribution which permits any household on any Indian reservation to participate simultaneously in the food stamp program and the distribution of federally donated foods."[169] Today, although eligible households can switch between the two programs, they cannot participate in both programs within the same month.

The reason for this prohibition is unclear, but the challenges it has created today are obvious, including the parity gap and difficulties navigating administration of two separate systems.[170] Even though FDPIR is intended to be supplemental only, more than 38% of participants rely on the program for all of their food needs.[171] For a person facing food insecurity, deciding whether to enroll in SNAP or FDPIR presents a difficult, confusing, and perhaps unnecessary challenge. Advocates have called for eliminating the barrier to dual enrollment of SNAP and FDPIR, especially given the food insecurity challenges revealed by the COVID-19 pandemic.[172] Ending this prohibition can put the decision in the hands of individual Native Americans, allowing for more effective administration of federal nutrition programs, reduced food insecurity among the most vulnerable, and increased self-determination and Tribal sovereignty.

Expanding Tribal administration of SNAP and other federal feeding programs can also help improve FDPIR by allowing more streamlined administration specific to each Tribe's needs.

---

168. § 4, 91 Stat. at 961 ("Distribution of commodities, with or without the food stamp program, shall also be made whenever a request for concurrent or separate food program operations, respectively, is made by a [T]ribal organization.").

169. *Id.*

170. TLCWG Dec. 2021 REPORT-OUT, *supra* note 137, at 4.

171. NANCY M. PINDUS ET AL., U.S. DEP'T OF AGRIC., STUDY OF THE FOOD DISTRIBUTION PROGRAM ON INDIAN RESERVATIONS (FDPIR), at xviii (2016), [https://perma.cc/46FA-Q42J].

172. NATIVE FARM BILL COAL., COVID-19 CRISIS RESPONSE: PROPOSED LEGISLATIVE TEXT (2020), [https://perma.cc/3YTQ-M4AK] (calling for temporary waiver of the dual SNAP/FDPIR prohibition due to COVID-19 food insecurity); Mucioki et al., *supra* note 4, at 96 (suggesting that policy makers "[r]econsider policy restrictions that inhibit vulnerable populations from accessing FDPIR when enrolled in other welfare assistance programs"); NAT'L CONG. OF AM. INDIANS, *supra* note 162 (including waiver of the dual SNAP/FDPIR prohibition as a top-line policy priority).

Although Indian Country is incredibly diverse and each nation has its own goals, many ITOs surveyed in the FNS's 2014 Tribal administration feasibility study cited "the ability to exercise their sovereignty . . . [and] offer culturally appropriate programming and services" as key reasons for wanting to administer federal nutrition assistance programs.[173] Most Tribes already have experience with administration of federal programs, although the study emphasized that one of the key remaining barriers to Tribal administration is a lack of federal funding for administrative costs.[174] Recent bipartisan legislation has been introduced to allow Tribal administration of SNAP.[175] This solution, when paired with ending the dual prohibition on enrollment, would reduce FDPIR participation barriers while also allowing for a more robust expression of Tribal sovereignty.

C. Changing the Procurement System

In order to enable more culturally relevant food, especially food that is Tribally produced and culturally appropriate at the granular Tribal level, significant changes need to be made to the current food procurement system for FDPIR. Currently, the USDA's Agriculture Marketing Service (AMS) is responsible for sourcing all USDA food items included in the FDPIR package.[176] Producers must undergo a lengthy USDA vendor certification process before they can be selected as AMS vendors.[177] In addition, producers must be able to provide a particular product on a national scale—for all FDPIR participants—in order to be accepted as an AMS vendor.[178] The high quantity can be a

173. GARASKY ET AL., *supra* note 121, at vii.

174. *Id.* at ix.

175. Press Release, Mike Rounds, U.S. Sen., Rounds, Smith Introduce Bipartisan Legislation to Promote Tribal Self-Governance for Federal Food Assistance Programs (Sept. 22, 2021), [https://perma.cc/ECH4-3Y2D].

176. *Food Distribution Program on Indian Reservations: Vendor*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (July 8, 2019), [https://perma.cc/TC3R-HA38].

177. INDIGENOUS FOOD & AGRIC. INITIATIVE, FDPIR 638 FOOD SOURCING APPLICATIONS (Jan. 22, 2021), [https://perma.cc/574D-PXUZ].

178. *Id.*; *see also* AGRIC. MKTG. SERV., U.S. DEP'T OF AGRIC., SUPPLEMENT 603 TO THE AMS MASTER SOLICITATION FOR THE PURCHASE OF FROZEN BISON PRODUCTS FOR DISTRIBUTION TO FEDERAL FOOD AND NUTRITION ASSISTANCE PROGRAMS 8 (2021),

significant barrier for smaller Tribal producers, especially those who produce culturally relevant foods, such as bison, which traditionally have much lower yields.[179]

To make matters worse, this process is further complicated when it comes to fresh fruit and vegetables, a highly sought and highly fought for addition to the food package. Fresh produce is sourced by the Department of Defense through their Fresh Fruits and Vegetables Program (USDA DOD Fresh), which is administered by the Defense Logistics Agency (DLA).[180] The DLA system is entirely separate and distinct from the AMS system. Therefore, a Tribal producer who wants to provide fresh produce for FDPIR must go through multiple channels, agencies, and certification processes before they can be accepted as a vendor.

Tribal leaders have proposed multiple solutions to this procurement challenge. First, advocates have sought regional sourcing with lower production thresholds to make it easier for smaller producers, including Tribal producers, to be selected as vendors, especially when harvests are low.[181] This could also allow for culturally relevant foods to be truly relevant to a specific Tribe instead of considering Indian Country as a monolith. With changes in the centralized warehousing and distribution system, this change could also lead to improved food quality and availability.[182] However, the USDA has been reluctant to implement this change, perhaps due in part to fear of litigation based on the geographic differentiation of the food package.

Second, advocates have called for the sourcing of fresh fruits and vegetables to all be housed under one roof with the AMS.[183] This would significantly ease the certification process and lower

---

[https://perma.cc/BK9F-DFYZ] (requiring the purchase unit size for bison to be 40,000 pounds net weight, or 1,000 shipping containers).

179. *Bison Production*, PENNSTATE EXTENSION (Oct. 10, 2005), [https://perma.cc/59VW-8KAN] (noting that bison are mostly appropriate for small-scale operations and that the average herd size in Pennsylvania is sixteen bison).

180. *USDA DOD Fresh Fruit and Vegetable Program*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (July 12, 2022), [https://perma.cc/VMT7-9VD3].

181. JANIE SIMMS HIPP & COLBY D. DUREN, SEEDS OF NATIVE HEALTH, REGAINING OUR FUTURE: AN ASSESSMENT OF RISKS AND OPPORTUNITIES FOR NATIVE COMMUNITIES IN THE 2018 FARM BILL 55-56, 58 (2017), [https://perma.cc/FD6J-VDX9].

182. MOORE ET AL., *supra* note 107, at 13-14.

183. HIPP & DUREN, *supra* note 181, at 55-56.

the administrative burden of potential producers.[184] Requests of this nature made during formal consultation have had little to no success, in part due to the fact that the DOD has not been part of the consultation process, limiting what the USDA can do on its own.[185] Tribal requests to change the procurement system have been denied.[186]

However, during the COVID-19 pandemic, the USDA created a new emergency program, the Farmers to Families Food Box.[187] This program was intended to alleviate both hunger and the economic stresses on farmers.[188] Similar to FDPIR, the Farmers to Families Food Box provided a direct distribution food package that included both shelf-stable products and fresh produce.[189] However, under the new program, AMS procurement included direct purchase of fresh produce, and the food was procured using a regional vendor system.[190] These are enhancements that FDPIR advocates, including the TLCWG, have previously been denied. Now that the AMS has demonstrated that these changes are possible, it is critical that they be applied to FDPIR. These changes would increase Tribal sovereignty by increasing the participation of Native American producers in FDPIR, allowing Tribes the opportunity to reconnect to traditional foodways and feed themselves.

The challenges to procurement could also be easily alleviated by the expansion of 638 authority to the USDA. Under 638 authority, Tribes would become responsible for procurement; the AMS would be no longer involved. Thus, this tool for expanding Tribal sovereignty could serve a dual purpose by eliminating the procurement challenge. USDA 638 authority

---

184. *Id.*

185. *See* TLCWG Feb. 2020 REPORT-OUT, *supra* note 137 (noting that the USDA provided updates on behalf of the DOD, and that the DOD and the FNS meet regularly, but giving no indication that the DOD would join any future consultations); TLCWG Dec. 2021 REPORT-OUT, *supra* note 137, at 6-7 (noting that the AMS stated that taking over produce procurement from the DOD would "not be conducive" under the current system).

186. *See* TLCWG Dec. 2021 REPORT-OUT, *supra* note 137, at 6-7.

187. *USDA Farmers to Families Food Box*, U.S. DEP'T OF AGRIC. AGRIC. MKTG. SERV. (May 28, 2021), [https://perma.cc/RZ6L-N8TZ].

188. *USDA Announces Coronavirus Food Assistance Program*, U.S. DEP'T OF AGRIC. (Apr. 17, 2020), [https://perma.cc/26WY-X2WH].

189. *USDA Farmers to Families Food Box*, *supra* note 187.

190. *Id.*

would also make it easier for Native American and local producers to become vendors, as they would not have to jump through the dual hoops of both the AMS and DOD Fresh.[191] While this is an excellent solution for Tribes that can take on 638 authority, the broader issues with the AMS and procurement still need to be addressed, as 638 authority is not always a viable option for smaller Tribes. The future of FDPIR procurement requires a multifaceted solution that includes expanding 638 authority, adopting regional distribution, and transferring all fresh produce sourcing to either the AMS or to Tribes through 638.

V. CONCLUSION

For the first time in a long time, FDPIR enrollment is showing potential for growth.[192] This is in part due to the recent wins Indian Country advocates have long fought for. Although rooted in the colonizing practice of rations and a long history of federal neglect, FDPIR has become an integral part of many Tribal food systems. The program's unique scope as the only federal nutrition program provided specifically for Native Americans makes it a critical tool for addressing the current challenges to food security and health in Indian Country.

While recent advocacy from Tribal leaders and allies has led to significant improvements in the last two decades, the persistence of centuries-old challenges is telling. Many of the challenges that have persisted the most, such as the lack of healthy or culturally specific foods, distribution challenges, and administrative uncertainty, can be traced to a lack of Tribal sovereignty. In contrast, the most effective solutions have come from participants and Tribal nations themselves. As a result, the advancement of Tribal sovereignty will be critical for the advancement of FDPIR.

---

191. *See* INDIGENOUS FOOD AND AGRIC. INITIATIVE, *supra* note 177.

192. *Food Distribution Program Tables*, U.S. DEP'T OF AGRIC. FOOD & NUTRITION SERV. (Nov. 10, 2022), [https://perma.cc/E7MY-WQ73] (selecting most recent data titled "Participation or Meals Served Participation (FDPIR and CSFP), Meals Served (NSIP)") (showing peak participation of roughly 140,000 participants in 1989, down to 75,600 in 2013, and back up to 87,200 in 2018).

FDPIR provides a clear mechanism for Tribal nations to secure culturally essential and nutritious foods for their people, but the history of the program in federal control has fallen well short of this goal. Recent successes like the USDA 638 demonstration project are a step in the right direction, and policy makers should expand on this success in the 2023 Farm Bill and beyond. Only by addressing FDPIR's fraught history can we ensure the program's future.