

Tax Experimentation

Michael Abramowicz

Follow this and additional works at: <https://scholarship.law.ufl.edu/flr>



Part of the [Tax Law Commons](#)

Recommended Citation

Michael Abramowicz, *Tax Experimentation*, 71 Fla. L. Rev. 65 ().

Available at: <https://scholarship.law.ufl.edu/flr/vol71/iss1/2>

This Article is brought to you for free and open access by UF Law Scholarship Repository. It has been accepted for inclusion in Florida Law Review by an authorized editor of UF Law Scholarship Repository. For more information, please contact kaleita@law.ufl.edu.

TAX EXPERIMENTATION

*Michael Abramowicz**

Abstract

Random experiments could allow the government to test tax policies before they are enacted into general law. Such experiments can be revenue-neutral, with the tax authority ensuring *ex post* that average tax revenues received from taxpayers in the treatment and control groups are equal. Taxpayers might thus volunteer even for experiments that would broaden the tax base, for example by eliminating deductions. Continued participation by taxpayers in such experiments would indicate that the proposed reforms are efficient at least if externalities are disregarded. Non-revenue-neutral experiments raise greater concerns about horizontal inequity, but they may be helpful in addressing questions about effects of tax rates and in increasing participation.

INTRODUCTION66

I. REVENUE-NEUTRAL TAX EXPERIMENTS76

 A. *A Hypothetical Experiment*76

 1. The Potential Benefit to Taxpayers76

 2. Adjustments to Ensure Revenue Neutrality77

 3. Evaluation of Experimental Success80

 4. Enhancement of Experimental Interpretability81

 a. Varying the Treatment83

 b. Two-Level Randomization85

 c. Intent-to-Treat Randomization87

 B. *Other Revenue-Neutral Tax Experiments*87

 1. Deductions88

 a. Disallowing Other Deductions88

 b. Allowing New Deductions89

 2. Income90

 a. Imputing Income90

 b. Taxing Work Amenities91

 c. Changing Recognition Timing92

 3. Tax Procedure93

 4. Radical Tax Reform93

* Professor of Law, George Washington University. I am especially indebted to Andrew Blair-Stanek, a coauthor on a related article, for his many excellent ideas and suggestions.

II. CHALLENGES FOR TAX EXPERIMENTATION 95

 A. *Transitions* 95

 1. Increased Scope 95

 2. Opt-Out Experiments 97

 3. Involuntary Experimentation 98

 B. *Equity* 100

 1. Horizontal Equity 100

 a. The Case for Randomness 100

 b. *Ex Ante* Insurance to Avoid Inequity 103

 2. After-Tax Income Inequality 104

 C. *Manipulation* 106

 1. Eligibility and Matching 106

 2. Income and Deduction Shifting 108

 3. Status Changes 110

 D. *Externalities* 113

III. BEYOND REVENUE NEUTRALITY 116

 A. *Participation Inducements* 116

 B. *Experiments on Tax Rates* 117

 C. *Self-Executing Tax Experiments* 119

 D. *Larger Experimental Units* 120

CONCLUSION 122

INTRODUCTION

The Tax Cuts and Jobs Act of 2017¹ is an experiment on a grand scale. If, as critics claim, the statute encourages expensive tax avoidance strategies that will lead to even lower tax revenues than projected,² the

1. Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, 131 Stat. 2054. The bill was entitled the “Tax Cuts and Jobs Act” until right before it passed, when the Senate parliamentarian ruled that the title violated the Byrde rule. Eli Watkins, *Senate Rules Force Republicans to Go with Lengthy Name for Tax Plan*, CNN (Dec. 19, 2017), <https://www.cnn.com/2017/12/19/politics/tax-bill-name-delay/index.html> [<https://perma.cc/D8TZ-4XJZ>]. As a result, the final bill was entitled “An Act to provide for reconciliation pursuant to titles II and V of the concurrent resolution on the budget for fiscal year 2018.” *Id.* Most commentators continue to refer to the bill as the Tax Cuts and Jobs Act (or TCJA). *See, e.g.*, Stephen K. Cooper et al., *Eyes Turn Toward 2018 Tasks as Tax Reform Becomes Law*, 158 TAX NOTES 28, 28 (2018) (referring to “[t]he law, known informally as the Tax Cuts and Jobs Act (P.L. 115-97)”); Amy Hamilton, *Connecticut Finds a SALT Workaround That Would Actually Work*, 158 TAX NOTES 1328, 1328 (2018) (referring without qualification to the “Tax Cuts and Jobs Act”).

2. *See generally* David Kamin et al., *The Games They Will Play: Tax Games, Roadblocks, and Glitches Under the 2017 Tax Legislation*, 103 MINN. L. REV. (forthcoming 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3089423 [<https://perma.cc/MEG6-AHKK>]

inefficiencies and budget strains³ will be national in scope. The experiment has no control group. If the economy grows rapidly over the next few years, reform proponents will likely take credit, while critics will insist that the economic growth was exogenous.⁴ Both sides will make their case with anecdotal evidence rather than with the results of randomized experiments that economists generally prefer.⁵ Scholars have identified an empirical deficit in tax scholarship,⁶ but because federal tax law exhibits little exogenous variation,⁷ the deficit is difficult to correct.⁸

This Article describes an approach to tax reform that might have been, and that might still be, an experimental approach—in which proposed reforms are tested initially on groups of willing taxpayers. In recent years, legal literature has focused attention on the design⁹ and justifiability¹⁰ of randomized government policy in areas including consumer protection,¹¹

(identifying numerous mechanisms that taxpayers may use to avoid taxes as a result of the tax reform).

3. The official estimate of the Congressional Budget Office and the Joint Committee on Taxation is that the tax changes will cause an increase in deficits of \$1.8 trillion over ten years. See Letter from Keith Hall, Dir., Cong. Budget Office, to Ron Wyden, Sen. (Jan. 12, 2018), <https://www.cbo.gov/publication/53437> [<https://perma.cc/NW9Q-CQST>].

4. Economists had different estimates of the efficiency consequences of the last large tax reform. See Alan J. Auerbach & Joel Slemrod, *The Economic Effects of the Tax Reform Act of 1986*, 35 J. ECON. LIT. 589, 619–20 (1997).

5. See, e.g., Gary Burtless, *The Case for Randomized Field Trials in Economic and Policy Research*, 9 J. ECON. PERSP. 63, 64 (1995).

6. See, e.g., Michael J. Bommarito II et al., *An Empirical Survey of the Population of U.S. Tax Court Written Decisions*, 30 VA. TAX REV. 523, 526 (2011) (“Many authors have noted the absence of empirical research in tax law”); Nancy Staudt, *Empirical Taxation*, 13 WASH. U. J.L. & POL’Y 1, 1–2 (2003) (“[T]he literature gives legislators little guidance for selecting one proposal from another in the wide array of sound policy options.”).

7. The exceptional rigorous analyses of changes in federal law prove the rule. Ed Fox, for example, analyzes the effect of federal tax law on marriage by exploiting differences in *state* laws that led the federal law to apply differently in different states at the same time. See generally Edward Fox, *Do Taxes Affect Marriage? Lessons from History* (Law & Econ. Res. Paper Ser., Working Paper No. 17-15, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2988559 [<https://perma.cc/HW8V-7TBK>].

8. The deficit may also exist in part because tax law professors prefer theory to empirics. See Staudt, *supra* note 6, at 2 (“[F]ew attempt to test whether their hypotheses hold true in the real world.”). But even in a conference on tax empiricism, none of the papers directly assessed the effects of federal law. See *id.* at 5–8 (describing the contributions). With greater policy variation, especially random variation, scholarship would likely follow.

9. See, e.g., Michael Abramowicz et al., *Randomizing Law*, 159 U. PA. L. REV. 929, 974 (2011).

10. See generally Ronen Perry & Tal Z. Zarsky, “*May the Odds Be Ever in Your Favor*”: *Lotteries in Law*, 66 ALA. L. REV. 1035 (2015) (discussing justifications for randomized government policy).

11. See Talia B. Gillis, *Putting Disclosure to the Test: Toward Better Evidence-Based Policy*, 28 LOY. CONSUMER L. REV. 31, 38 (2015).

securities law,¹² patent law,¹³ and even regulation of food safety.¹⁴ Randomized experiments produce the best attainable evidence of legal policies' effect on behavior, but the only known tax-related randomized experiments involve tax procedure¹⁵ and welfare,¹⁶ not substantive tax law. Interest in the behavioral effects of tax law on taxpayers pervades the literature,¹⁷ but there has been virtually no consideration of whether tax law itself might benefit from systematic randomized experiments.¹⁸

Tax law is a promising field in which the government might run randomized experiments. The large number of taxpayers should make it possible to find voluntary treatment groups that are small relative to the population yet large enough to generate statistical power.¹⁹ Each tax change would apply to a random selection of qualified taxpayers agreeing to opt into the experiment, thus producing both a treatment group and a control group still subject to the status quo law. Experiments might test multiple tax law changes, disentangling any resulting interaction effects when individual taxpayers are in multiple treatment groups.²⁰ Meanwhile, tax experiments could be much cheaper than many other form of legal experimentation because the existing obligation to file tax

12. See Zachary J. Gubler, *Making Experimental Rules Work*, 67 ADMIN. L. REV. 551, 558 (2015).

13. See Lisa Larrimore Ouellette, *Patent Experimentalism*, 101 VA. L. REV. 65, 65 (2015).

14. See Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 28 (2017).

15. See, e.g., STATE OF CAL. FRANCHISE TAX BD., READYRETURN PILOT TAX YEAR 2004 STUDY RESULTS 2 (2006) (reporting results of an experiment on sending taxpayers pre-filled returns), <https://www.ftb.ca.gov/readyReturn/TY04RRFinalReport.pdf>. But the survey reported no results on perhaps the most interesting question, whether control group taxpayers paid more or less tax than treatment group taxpayers. *Id.* at 29–30 (comparing treatment group taxpayers reported state and federal income, but ignoring control group taxpayers); cf. Dennis J. Ventry, Jr., *Intuit's Nine Lies Kill State E-Filing Programs and Keep 'Free' File Alive*, 57 ST. TAX NOTES 555, 559 (2010) (arguing that taxpayers using ReadyReturn paid no more taxes than others, but without any comparisons between taxpayers and control group to support this argument).

16. See I DAVID KERSHAW & JERILYN FAIR, THE NEW JERSEY INCOME-MAINTENANCE EXPERIMENT: OPERATIONS, SURVEYS, AND ADMINISTRATION 3 (1976).

17. See, e.g., William J. Congdon et al., *Behavioral Economics and Tax Policy*, 62 NAT'L TAX J. 375 (2009).

18. A rare exception is the following single sentence: "Conceivably, field experiments could be designed where individuals are randomly assigned to different tax schedules in the spirit of the older negative income tax experiments in the United States." Emmanuel Saez et al., *The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review*, 50 J. ECON. LIT. 3, 43 (2012) (referring to experiments mentioned *infra* note 24 and accompanying text).

19. See, e.g., Jacob Cohen, *Statistical Power Analysis*, 1 CURRENT DIRECTIONS PSYCHOL. SCI. 98, 98 (1992) (noting that statistical power depends on sample size, amongst other variables).

20. See generally JAMES JACCARD & ROBERT TURRISI, INTERACTION EFFECTS IN MULTIPLE REGRESSION (2d ed. 2003) (describing techniques for identifying interactions between multiple variables of interests).

forms²¹ provides the government much of the evidence that it might need to assess the effects of an experiment.

A plausible explanation for the lack of attention to tax experimentation lies in tax law's core value of horizontal equity.²² This value makes especially salient the concern that experiments necessarily produce unequal treatment of similarly situated individuals. Randomly assigning some taxpayers to a tax law change violates horizontal equity because those taxpayers would then be treated differently from similarly situated taxpayers not so assigned. This concern diminishes if the treatment is a tax break. Randomized government experiments offering benefits are widely considered ethical,²³ and random experiments granting low-income individuals benefits have occurred.²⁴ Experiments often offer the treatment group something denied to the control group. In medical experiments, some patients who hope for a new treatment instead receive a placebo.²⁵ If uncertainty about relative efficacy suffices to justify medical experiments,²⁶ then legal experiments should be similarly defensible.

There are, however, problems with tax experiments limited to tax breaks. First, experimentation becomes a one-way ratchet, always expected to lower tax revenues. The Tax Cuts and Jobs Act increases deficits,²⁷ so future policymakers may need to raise more revenue, not select further tax breaks. Second, tax policy questions susceptible to experimentation often involve trade-offs, for example, whether taxpayers might be better off with higher rates but more deductions. Third,

21. See, e.g., I.R.C. § 6012 (West 2018) (specifying who must file U.S. federal tax returns).

22. Alan J. Auerbach & Kevin A. Hassett, *A New Measure of Horizontal Equity*, 92 AM. ECON. REV. 1116, 1116 (2002) (“There is virtual unanimity that horizontal equality, treating equals equally, is a worthy goal of any tax system.”).

23. See, e.g., Abramowicz et al., *supra* note 9, at 968 (“[W]hen scarce resources are distributed, randomization ensures that the distribution occurs without favor and in a way that limits rent-seeking . . .”). For a detailed treatment of ethical issues in social experiments, see generally THE BROOKINGS INST., *ETHICAL AND LEGAL ISSUES OF SOCIAL EXPERIMENTATION* (Alice M. Rivlin & P. Michael Timpane eds., 1975).

24. See, e.g., KERSHAW & FAIR, *supra* note 16, at 8.

25. The use of a placebo was initiated by Austin Flint in 1862. See AUSTIN FLINT, *A TREATISE ON THE PRINCIPLES AND PRACTICES OF MEDICINE* 1019–20 (4th ed. 1873).

26. The literature on research ethics suggests that experiments are permissible where practitioners are in “clinical equipoise” about the best course of treatment. See Franklin G. Miller & Howard Brody, *Clinical Equipoise and the Incoherence of Research Ethics*, 32 J. MED. & PHIL. 151, 152 (2007). A critique is that a doctor's individualized obligation is not to choose among treatments at random, but to choose a treatment thought to be best given available information for each individual patient. *Id.* at 156 (arguing that clinical equipoise is an incoherent theory and that ethical obligations in research are distinct from those in clinical settings).

27. See Letter from Keith Hall, Dir., Cong. Budget Office, to Ron Wyden, Sen., *supra* note 3.

taxpayers opting in to tax experiments will generally be well advised and wealthy,²⁸ so experimentation on tax breaks will often be inherently regressive.

These concerns, however, reflect a fallacy. If treatment group taxpayers benefit relative to the control group, then it might seem that all other taxpayers lose. After all, if some pay less revenue to the government, then others bear a higher proportion of taxation's overall burden. But taxpayers in the treatment group might benefit while paying the *same* level of tax on average as taxpayers in the control group. Many proposals for tax reform purport to correct inefficiencies in the tax code. Changes that remove economic distortions could be combined with rate changes in ways that would improve taxpayer welfare while maintaining government revenue.

This observation leads to the following insight: Tax experiments can be revenue neutral by design.²⁹ The treatment group in such an experiment, on average, pays taxes as high as those paid by the control group. Revenue neutrality substantially addresses the problems above. A revenue-neutral experiment is not a one-way ratchet. Instead, it allows questions of economic efficiency to be assessed independent of normative questions about overall tax levels and the distribution of the tax burden across income and other groups.³⁰ Revenue-neutral experiments necessarily allow the government to assess the impact of trade-offs. Even if only relatively wealthy taxpayers opted into these experiments, they would be guaranteed to pay as much in taxes as they would absent experimentation. Participating taxpayers anticipate benefiting from the trade-offs embodied by experimentation, but not at the direct expense of other taxpayers. Their experience, moreover, may help generate tax reform of broader benefit.

28. The government has sought to address this concern by providing grants for Low Income Taxpayer Clinics. See *Low Income Taxpayer Clinics*, IRS, <https://www.irs.gov/advocate/low-income-taxpayer-clinics> [<https://perma.cc/D62E-DZZB>] (last updated July 12, 2018).

29. Commentators have often suggested that tax reform should be revenue neutral. See, e.g., Jason S. Oh, *Will Tax Reform Be Stable?*, 165 U. PA. L. REV. 1159, 1200 (2017) (“In the current fiscal environment, it is widely accepted that any tax reform should not lose any revenue. . . . Revenue-neutrality is particularly relevant in base-broadening reform . . .”). But revenue neutrality has not been discussed in connection with experimentation.

30. The Joint Committee on Taxation estimates that the Tax Cuts and Jobs Act will result in a lower percentage of total tax revenues being paid by those with relatively high incomes and a higher percentage of total tax revenues being paid by those with relatively low incomes by 2027. JOINT COMM. ON TAXATION, JCX-68-17, DISTRIBUTIONAL EFFECTS OF THE CONFERENCE AGREEMENT FOR H.R.1, THE “TAX CUTS AND JOBS ACT” 5 (2017), <https://www.jct.gov/publications.html?func=startdown&id=5054> (projecting, for example, that taxpayers with income of \$1,000,000 and over will contribute 18.9% of total taxes, instead of 19.1% of total taxes under prior law).

What combination of tax changes might serve as the subject of a tax experiment? Consider, as a simple (at the risk of being trivial) illustration, the recent tax reform's limitations on the deductibility of entertainment expenses.³¹ Under prior law, taxpayers could deduct expenses such as tickets for a sporting event used to entertain a client,³² but taxpayers can no longer do so.³³ A longstanding argument for restricting deductibility of such expenses is that taxpayers derive utility from such expenditures.³⁴ Because the recipient of entertainment benefits generally need not include the benefits in income,³⁵ deductibility immunizes such expenditures from taxation.³⁶

Limitations on the entertainment deduction might seem a poor candidate for experimentation because no one would volunteer for increased tax liability. Whatever the act's overall merits, the Tax Cuts and Jobs Act demonstrates the political palatability of bundling taxpayer-unfriendly changes with taxpayer-friendly changes: A limitation on deductibility was joined to a rate reduction.³⁷ The entertainment-deduction limits recouped only a tiny fraction of the revenue loss from other parts of the tax bill.³⁸ But the lobbyists' embrace—or at least

31. I.R.C. § 274(a)(1)(A) (Supp. V 2018) (“No deduction otherwise allowable under this chapter shall be allowed for any item . . . [w]ith respect to an activity which is of a type generally considered to constitute entertainment, amusement, or recreation . . .”).

32. Entertainment expenses were deductible if they were “directly related to” or “associated with” the taxpayer’s trade or business. I.R.C. § 274(a)(1)(A) (2012); *Walliser v. Comm’r*, 72 T.C. 433, 441–43 (1979) (finding otherwise deductible entertainment expenses failed both “directly related to” and “associated with” tests).

33. See I.R.C. § 274(a)(1)(A) (Supp. V 2018).

34. President Kennedy expressed this argument clearly: “Even though in some instances entertainment and related expenses have an association with the needs of business, they nevertheless confer substantial tax-free personal benefits to the recipients.” *President’s 1961 Tax Recommendations: Hearings on the Tax Recommendations of the President Contained in His Message Transmitted to the Congress Before the H. Comm. on Ways & Means*, 87th Cong. 12–13 (1961).

35. See I.R.C. § 132(a)(3), (d) (2012) (excluding fringe benefits from employee’s gross income when the expenditures would have been deductible by the employee if paid by the employee herself); *United States v. Gotcher*, 401 F.2d 118, 120, 121, 124 (5th Cir. 1968) (holding that taxpayer had no gross income from travel and apparent entertainment expenses provided by automobile manufacturer who wanted taxpayer to open a dealership).

36. The issues raised by the entertainment deduction arise as well with other potentially deductible expenses, such as expenses for meals, home offices, and education. For an early reform proposal accounting for these connections, see Daniel I. Halperin, *Business Deduction for Personal Living Expenses: A Uniform Approach to an Unsolved Problem*, 122 U. PA. L. REV. 859, 859 & n.1, 862–63 (1974).

37. Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, 131 Stat. 2054.

38. The Tax Cuts and Jobs Act’s limitation on meal and entertainment expenses was estimated by the nonpartisan Joint Committee on Taxation to raise \$23.5 billion over the ten-year budget window. JOINT COMM. ON TAXATION, JCX-67-17, ESTIMATED BUDGET EFFECTS OF THE

grudging acceptance—of this tax reform, as opposed to a reform that lowered rates less with no change to entertainment expense deductibility, suggests that the deduction was distortionary. It is hard to tell for sure.³⁹ But the example conceptually demonstrates that taxpayers should embrace revenue-neutral tax changes that remove decision-making distortions.

How might an experiment on entertainment deductibility have worked? Taxpayers could have volunteered to give up their entertainment deductions. The carrot to induce participation would be that these taxpayers receive lower rates, rates just low enough that government tax revenues would be identical. If the government selected the rate, it might miscalculate or intentionally set the rate low, disguising a tax cut as an experiment.⁴⁰ But there is a simple antidote: The law authorizing such experiments could provide that the total tax liability of treatment group taxpayers be multiplied by a factor sufficient to ensure that revenues from these taxpayers equal, in the aggregate, the revenues that they would have paid, as extrapolated from payments by control group taxpayers. For example, if the average treatment group taxpayer's liability, when measured after limiting interest deductions, would have been 1% more than the average control group taxpayer's liability, then each treatment group taxpayer might be given approximately a 1% discount on the nominal reported tax liability, calculated without granting the entertainment deduction.⁴¹

With so simple a formula for achieving revenue neutrality, the taxpayers most likely to opt in would be those that in the absence of the experiment would claim a relatively small entertainment deduction anyway. This would be unfortunate, as the greatest tax distortion presumably involves taxpayers who take large entertainment deductions. But a more sophisticated approach could calculate a custom multiplier for each participant. A statistical model developed after the experiment would feature *ex ante* attributes of the treatment and control group taxpayers, including previous entertainment expenses, as independent variables. The model would predict total tax liability reported by both control and treatment group taxpayers during the experimental period. If

CONFERENCE AGREEMENT FOR H.R.1, THE "TAX CUTS AND JOBS ACT" 4 (2017), <https://www.jct.gov/publications.html?func=startdown&id=5053>.

39. The combination of large corporation tax rate reductions with limitations on certain business deductions might be attributed to a political motive to highlight large reductions, if those reductions are greater in salience than the deduction limitations.

40. Cf. Karen C. Burke & Grayson M. P. McCouch, *Lipstick, Light Beer, and Back-Loaded Savings Accounts*, 25 VA. TAX REV. 1101, 1148 (2006) (criticizing purported tax reform proposals as "amount[ing] to little more than a disguised tax cut for high-income individuals").

41. More precisely, the tax payment multiplier would be $(100 / 101)$ to ensure revenue neutrality, so the discount is approximately 0.99%.

the model predicted that a taxpayer would have liability 3% higher if assigned to the control group, then the taxpayer would receive approximately a 3% tax rate reduction on the nominal reported tax liability if assigned to treatment. A treatment taxpayer's entertainment expenditures would not reduce tax liability, and the taxpayer therefore would have incentives to reduce entertainment expenditures on the margin.

This approach can also be used to experiment on the possibility of new deductions. For example, current tax law does not allow taxpayers to deduct commuting expenses.⁴² Commuting expenses may be seen either as enabling taxpayers to travel to work or as enabling taxpayers to live away from work, so the normative case for deductibility is close.⁴³ An experiment might test deductibility for all or only in limited circumstances, such as for commuting to an employer located far from affordable housing. Treatment group taxpayers would receive the deduction offered but would pay higher tax rates to maintain revenue neutrality relative to the control group. This could facilitate efficient behavioral changes, for example leading a taxpayer to work for an employer further away.⁴⁴ Or, the experiment might fail, suggesting the efficiency of the status quo.

Revenue-neutral tax experiments cannot answer all normative questions about tax law changes. First, these experiments assess efficiency only. A full normative analysis should embrace distributional concerns as well, which have received increased attention as inequality nationally has widened.⁴⁵ Nonetheless, the practice of revenue-neutral experimentation could promote attention to distributional issues. Some critics of the recent tax law changes argue that these changes amounted to a regressive tax cut disguised as tax reform.⁴⁶ Experimentation could generate relatively uncontroversial tax reform measures, reducing the need for omnibus tax reform packages focused on efficiency. A tax package featuring untested or rejected efficiency measures might face

42. I.R.C. § 262(a) (2012); *Comm'r v. Flowers*, 326 U.S. 465, 473–74 (1946).

43. See generally Tsilly Dagan, *Commuting*, 26 VA. TAX REV. 185, 201–34 (2006) (analyzing policy arguments); Ronald S. Ross, *Should Deductible Commuting Be Contingent on Principal-Place-of-Business Criteria?*, 83 J. TAX'N 88 (1995) (summarizing legal developments and arguments).

44. This argument is strongest if the taxpayer's residence is effectively fixed. See William A. Klein, *Income Taxation and Commuting Expenses: Tax Policy and the Need for Nonsimplistic Analysis of "Simple" Problems*, 54 CORNELL L. REV. 871, 880 (1969).

45. See, e.g., THOMAS PIKETTY, *CAPITAL IN THE TWENTY-FIRST CENTURY* 20 (2017) (documenting increased inequality within nations).

46. See, e.g., Thomas B. Edsall, *You Cannot Be Too Cynical About the Republican Tax Bill*, N.Y. TIMES (Dec. 21, 2017), <https://www.nytimes.com/2017/12/21/opinion/republican-tax-bill-trump-corker.html> [<https://perma.cc/X3AU-4X6D>].

criticism. Experimentation could thus foster a policy culture in which distributional issues are considered directly and acknowledged as policy choices rather than assumed to be inevitable epiphenomena of other tax policies.⁴⁷

Second, the experimental approach assumes that taxpayers internalize benefits of tax law provisions, but tax provisions may be motivated by anticipated third-party effects. For example, the mortgage interest tax deduction is purportedly motivated by a desire to provide incentives for home ownership, which supposedly generates positive externalities.⁴⁸ Many scholars are skeptical of this justification,⁴⁹ but the experimental approach that this Article describes does not provide a means for measuring such externalities or their absence. It might, however, be possible to design an experiment that can overcome this concern by experimenting at a level other than that of a taxpayer. For example, a homeowner's association might be allowed to opt in on behalf of all owners of homes in the association. If one assumes that this level internalizes externalities, then the willingness of such associations to participate would suggest that the benefits of the proposed treatment exceed the costs.

Third, revenue-neutral experiments cannot easily assess the macroeconomic consequences of changes in tax rates. Revenue neutrality would undo an isolated tax rate change. Proponents justify important features of the recent tax reform, such as the reduction in corporate tax rates⁵⁰ and the allowance of reduced rates for certain passthrough income,⁵¹ on the ground that they may improve the overall economic climate, thus benefiting workers as well as owners of capital.⁵² Non-revenue-neutral tax experiments may thus be needed.⁵³ For example, the

47. Robert Shiller has proposed that the government specify in advance the maximum level of inequality. See ROBERT J. SHILLER, *THE NEW FINANCIAL ORDER: RISK IN THE 21ST CENTURY* 149–64 (2003).

48. See, e.g., Denise DiPasquale & Edward L. Glaeser, *Incentives and Social Capital: Are Homeowners Better Citizens?*, 45 J. URB. ECON. 354, 383 (1999) (finding evidence that homeownership causes greater investment in social capital).

49. See, e.g., A. Mechele Dickerson, *The Myth of Home Ownership and Why Home Ownership Is Not Always a Good Thing*, 84 IND. L.J. 189, 207 (2009); Stephanie M. Stern, *Reassessing the Citizen Virtues of Homeownership*, 111 COLUM. L. REV. 890, 890–96 (2011).

50. I.R.C. 11(b) (Supp. V 2018) (cutting corporate tax rate from 35% to 21%).

51. *Id.* § 199A.

52. See Martin A. Sullivan, *Economic Analysis: Corporate Tax Incidence Made Simple*, 157 TAX NOTES 454, 456 (2017) (providing overview of the evidence that cutting corporate tax rates may, to some extent, benefit workers).

53. MARK P. KNIGHTLEY, CONG. RESEARCH SERV., R41596, *THE MORTGAGE INTEREST AND PROPERTY TAX DEDUCTIONS: ANALYSIS AND OPTIONS* 8–11 (2014) (reviewing the literature on whether the home mortgage interest deduction creates positive externalities to justify the tax benefit).

government might have selected random taxpayers to receive a 1% tax reduction in passthrough income and assessed the effects of such a reduction. Such experiments can provide information beyond the power of revenue-neutral tax experiments, at the cost of more serious horizontal equity objections. Without the revenue-neutrality constraint, the government can also grant inducements to tax experiment participants, such as a promise that treatment taxpayers will pay on average *less* than control taxpayers. Such inducements could reduce concerns about adverse selection and enable experiments with treatment groups covering many reforms.⁵⁴

With non-revenue-neutral experiments, success must be based on defined criteria, such as whether treatment businesses hire more employees. With opt-in revenue-neutral experiments, success can be measured based on the willingness of taxpayers to opt in, once early experimental periods produce information about the likely trade-offs. Such experiments can do little harm. At worst, only a few taxpayers, potentially unrepresentative, will participate, and the experiment will never scale up. If the experiment does scale, taxpayers would come to learn what tax rate discount or increase they might expect in exchange for the new tax treatment. Unless virtually all taxpayers wish to be part of an experiment, the government should be cautious in making an experiment permanent because opting-in taxpayers are self-selected. The government therefore might transition gradually, by making an opt-in experiment opt-out and then ultimately mandatory.⁵⁵ Alternatively, the government might allow all taxpayers to opt into the treatment group, without requiring it of any taxpayers.

Part I of this Article will describe revenue-neutral tax experiments in more detail, illustrating how they can be used to assess a wide range of tax policies. Part II will identify challenges for revenue-neutral experimentation, including scaling up experiments, addressing concerns about horizontal and vertical equity, countering the danger of taxpayer manipulation, and testing tax expenditures and other policies with goals beyond efficiency. Finally, Part III will consider non-revenue-neutral experiments. It will explain how such experiments can induce greater participation on a wider range of issues and can study the effect of changing marginal tax rates. It also describes the possibility of self-executing tax experiments, in which the law automatically will change in a direction indicated by the experimental results, and experiments in which individual taxpayers are not the unit of experimentation.

54. See *infra* Section III.A.

55. See *infra* Section II.A.

I. REVENUE-NEUTRAL TAX EXPERIMENTS

This Part discusses how the government might use revenue-neutral tax experiments to assess the efficiency of various features of the tax code. The approaches described here might be used by the U.S. federal government, by state or local governments, or by foreign governments. Thus, while this Article will use provisions from the Internal Revenue Code and recent tax reform statutes and proposals as examples, its analysis is not dependent on the structure of the U.S. income tax system. Section I.A elaborates the entertainment deduction experiment discussed in the introduction, and Section I.B describes other potential tax experiments.

A. *A Hypothetical Experiment*

The goal of revenue-neutral tax experimentation is to identify potential sets of changes to tax law that in combination would provide the government the same amount of revenue but that would reduce distortion of economic activity. This section elaborates how an experiment might have tested the ultimately adopted reform of removing the entertainment expenses deduction.

1. The Potential Benefit to Taxpayers

Why might a taxpayer be interested in the combination of tax rate reduction and loss of the deduction, assuming the taxpayer anticipates paying around the same amount as before? Consider a taxpayer with gross income of \$110,000, a tax rate of 25%,⁵⁶ and \$10,000 in certain entertainment expenses deductible under what we will assume is current law. Further, suppose that this taxpayer receives \$8,000 in subjective value from these expenses. That is, the taxpayer would be equally happy if the taxpayer could reallocate the \$10,000 entertainment expenses to \$2,000 in taxes and \$8,000 in cash. In the table below, this is listed as Scenario 1. Of course, not all taxpayers would value \$83,000 in take-home pay as much as \$75,000 plus \$10,000 in entertainment expenses, but some might. The immediate burden is not to show that a revenue-neutral change would be good for all taxpayers, just that it might be good for some.

56. These numbers are chosen to make the math easy, not to reflect the intricacies of the tax code. A more realistic example would take into account factors such as the standard deduction. See I.R.C. § 63(c) (2012 & Supp. V 2018) (allowing taxpayers to deduct a fixed amount of money in lieu of itemizing deductions).

Table 1. Hypothetical effects of an experiment

	Gross Income	Entertainment Expenses	Taxable Income	Tax Rate	Taxes	Take Home Income
Status quo	\$110K	\$10K	\$100K	25%	\$25K	\$75K
Scenario 1	\$110K	\$0	\$110K	24.5%	\$27K	\$83K
Scenario 2	\$108K	\$2K	\$106K	23.5%	\$25K	\$81K

This scenario illustrates that a tax law change can make the government better off (receiving \$2,000 more in revenue) while making a taxpayer no worse off. Though such tax changes are a plausible policy goal, a taxpayer would have no incentive to take the trouble to opt into such an experiment. The fact that this tax law change produces a surplus of \$2,000, however, suggests that it is possible to imagine a tax law change that could make this hypothetical taxpayer better off and the government no worse off. Consider, for example, Scenario 2. Here, assume that the taxpayer spends \$2,000 on entertainment and this reduction in spending leaves the taxpayer with only \$108,000 in gross income. With a 23.5% tax rate, the taxpayer pays \$25,000 in taxes, leaving the taxpayer with \$81,000 in take home income. Some taxpayers might prefer this to the status quo. If not, then the experiment would simply fail.

2. Adjustments to Ensure Revenue Neutrality

The sizes of the treatment and control groups could be equal, but that is not necessary. The advantage of using an equal number of taxpayers in each group is that this provides the greatest statistical power.⁵⁷ If either group is sufficiently small, then any differences between the two groups are more likely to be attributable to noise. On the other hand, placing most opting-in taxpayers in the treatment group maximizes the number who may receive their preferred tax treatment. For ease of exposition, however, this Article will assume that the groups are the same size.

The simplest technique for achieving revenue neutrality would be to calculate the quotient of the total taxes paid by the control group divided by the total taxes calculated by the treatment group, prior to applying a multiplier. Each taxpayer in the treatment group would then pay taxes

57. For a proof, see Hugh, Answer to *Sample Size proportion per control vs. experiment group*, STACK EXCHANGE (Mar. 5, 2017, 11:01 PM) <https://stats.stackexchange.com/questions/265622/sample-size-proportion-per-control-vs-experiment-group> [<https://perma.cc/F24K-96H4>].

equal to the amount the taxpayer reported, multiplied by this quotient. Suppose, for example, that the control group taxpayers paid a total of \$1 billion in taxes, and the treatment group taxpayers reported a total of \$1.04 billion in taxes. (A higher tax bill would be expected for this experiment, since the experiment is removing what we assume was a deduction available under current law.) Then, a treatment group taxpayer who had \$100,000 in gross income would have reported \$25,000 in taxes, regardless of the amount of entertainment expenses that the taxpayer incurred. Tax liability, however, would be only $\$25,000 * (1 / 1.04)$, or around \$24,000.

This approach to calculating the tax multiple enhances the likelihood that treatment group taxpayers will respond in the same way as they would if there were a broader change in tax law affecting all taxpayers. A taxpayer in the control group has experienced no legal change at all and thus should presumably behave in the same way as the taxpayer would have if the experiment had never occurred.⁵⁸ This is assuming, of course, that the experiment is large enough that each member of the treatment group will have only a negligible effect on the tax ratio. If one imagined the opposite—an experiment with two taxpayers, one in the treatment group and one in the control group—then the treatment taxpayer would expect his own tax liability to be equal to that of the control group taxpayer, and the treatment taxpayer would behave as if a lump sum tax would be imposed.⁵⁹ The law of large numbers is essential not only for statistical validity, but also because revenue neutrality means that the tax liability of each treatment group participant depends on all other participants' tax returns.⁶⁰

Calculating a single ratio for all taxpayers is not the best approach. Taxpayers with low levels of entertainment deductions would be especially likely to opt in. The ratio of taxes paid by taxpayers in the control and treatment groups would then be close to 1. Those with large entertainment deductions would thus opt out. The tax authority must calculate a different ratio for each treatment taxpayer, based on how placement in the control or treatment group affected taxpayers with similar characteristics. A simple version along these lines would be to define subgroups of taxpayers. A group, for example, might consist of all taxpayers from a particular industry with a particular level of income and

58. Below, this Article will consider the caveat that the treatment group's tax change might have some indirect effect on the control group. See *infra* Section I.A.4.

59. This is not entirely a bad result. Economists generally assume that lump sum taxes are the least distortionary. See Joseph E. Stiglitz, *Self-Selection and Pareto Efficient Taxation*, 17 J. PUB. ECON. 213, 217 (1982).

60. In a relatively small experiment, the government might calculate the ratio separately for each treatment group taxpayer—that is, making the ratio equal to the average tax reported of all taxpayers in the control group divided by the average tax reported of all taxpayers in the treatment group *other* than the taxpayer affected.

entertainment deductions from the year prior. This group would then be subdivided into treatment and control, and the same ratio would be used for all taxpayers in this subgroup.

A more sophisticated, yet still easily implementable, approach would use multivariate regression analysis. The government would estimate two regressions, one for the treatment group and one for the control group. Each would predict reported income after the experiment as a function of variables from past years' tax returns. The precise form of the regression does not matter much for this Article's purposes; a simple multiple linear regression model might work well,⁶¹ or the government might perform a nonlinear regression⁶² or even use machine learning techniques, such as a neural network regression⁶³ or a decision forest.⁶⁴ The government can thus calculate for each treatment taxpayer the ratio between the tax bill that would be expected if the taxpayer were in the control group divided by the expected unadjusted tax bill if in the treatment group. The ratio does not depend on the level of entertainment deductions or income claimed by the taxpayer in the treatment year. If the sum of liability applying this ratio does not produce precisely the revenue-neutral amount, then all treatment taxpayers' liability could be multiplied by a constant to ensure exact revenue neutrality.

The taxpayers would be informed of the ratio calculated at the end of the experiment and would receive a corresponding adjustment in liability. If the treatment taxpayers are entitled to a discount as a result, they would receive interest on the money for the period the government held it.⁶⁵ In a reverse experiment where taxpayers eventually pay more, if the taxpayers underestimated their final liability, they might pay interest, just as a taxpayer who is responsible for quarterly estimated tax payments may be required to pay interest when the quarterly payments are too low.⁶⁶ If final reconciliation of the experiment takes a while, for example because some taxpayers fail to file their tax returns on time,⁶⁷ the government could make an initial adjustment a few months after the

61. See generally PAUL D. ALLISON, *MULTIPLE REGRESSION: A PRIMER* (1st ed. 1998) (discussing multiple linear regression models).

62. See generally GEORGE A.F. SEBER & C.J. WILD, *NONLINEAR REGRESSION* (2003) (discussing nonlinear regression).

63. See generally, e.g., Donald F. Specht, *A General Regression Neural Network*, 2 IEEE TRANSACTIONS ON NEURAL NETWORKS 568 (1991) (discussing neural network regression).

64. See generally, e.g., Weida Tong et al., *Decision Forest: Combining the Predictions of Multiple Independent Decision Tree Models*, 43 J. CHEM. INF. COMP. SCI. 525 (2003) (proposing the decision forest model).

65. The IRS often pays interest on overpayments. See I.R.C. § 6611 (2012 & Supp. V 2018).

66. I.R.C. § 6654 (2012).

67. Taxpayers who file late must pay penalties. I.R.C. § 6651 (2012 & Supp. V 2018). Such penalties, excluding interest, could be included in the comparison between the treatment and control group, to account for the possibility that the tax regime may affect the timeliness of filing.

relevant taxable year and then a final adjustment some time later. If finality is more important than exact revenue neutrality, this final adjustment could be scheduled to come sooner rather than later.

These timing details aside, the statistical approach reduces the risk that because of adverse selection, those who opt into the experiment are those with characteristics that make them relatively immune to the tax law change at issue. Such taxpayers could still opt in, but if the relevant characteristics are captured by the model, then it would predict that these taxpayers' liability would not change much, and so the effect of the experiment on these taxpayers would be small. This does not solve the adverse selection problem completely, however. The independent variables are an incomplete list of factors that might affect taxpayers, and taxpayers can be expected to have private information about their future behavior.⁶⁸ A taxpayer who plans to reduce his entertainment deductions for reasons not apparent based on available data would be particularly likely to opt in.

3. Evaluation of Experimental Success

How can the government assess whether an experiment was successful? One question is whether the experiment led to behavioral changes among taxpayers. That might be discerned from tax returns filed, particularly if taxpayers in the control group are required to report the same data that they would have reported if they were in the treatment group, even if that data is no longer relevant in computing their tax liability. In this example, treatment group taxpayers might reduce their entertainment expenses. If entertainment expenses declined but gross income did not, then that would indicate that entertainment expenses in fact are largely not legitimate business expenses. More generally, the ratio of the decline in entertainment expenses to the decline in gross income provides a proxy for the success of the experiment. It is not obvious, however, what level marks the cutoff between success and failure, particularly because what matters is the *marginal* effect of the tax reduction on gross income.⁶⁹

This possibility highlights two points. First, even when tax rules are conventionally framed as binary choices, they often simply reflect polar points on the spectrum. This is more obvious in the context of entertainment deductions than in many areas of tax law because

68. See ROBIN BOADWAY, FROM OPTIMAL TAX THEORY TO TAX POLICY 50 (2012) (discussing the importance of asymmetric information to tax policy analysis).

69. See Manoj Viswanathan, *The Hidden Costs of Cliff Effects in the Internal Revenue Code*, 164 U. PA. L. REV. 931, 947 (2016) ("A taxpayer's marginal tax rate, in contrast to a taxpayer's average tax rate, is an effective indicator of how the Internal Revenue Code affects a taxpayer's decisions.").

businesses have long been allowed only partial deduction of certain classes of entertainment expenses.⁷⁰ Ideally, a process of experimentation might lead the government to home in on the efficient level of permissible deductibility. Second, it will not always be straightforward to interpret an experiment to determine whether it was successful or not, even as to the taxpayers who opted in.⁷¹ Therefore, the strongest indication that a tax change is efficient as to the taxpayers in the experiment is the mere fact that the taxpayers opted into the experiment. So long as an experiment is revenue neutral, taxpayers' willingness to opt into the experiment suggests that it is expected to leave the taxpayers better off, with no adverse consequences for the Treasury. Continued demand from taxpayers to participate in subsequent iterations of the experiment would strengthen this inference.

The inference, however, carries caveats. First, the experiment's success may not be generalizable to taxpayers who do not opt in. This highlights the question of how the government can scale up an experiment, a question to which we will soon return.⁷² Second, the efficiency of a tax experiment may depend not only on the effects of the tax on the taxpayers, but also the effect of the tax on third parties. In the context of the entertainment deduction, for example, this analysis has so far overlooked the clients who would have been wined and dined but no longer received such benefits. Any loss of utility that these clients suffer might count as a negative effect on social welfare. Or perhaps the effect is positive, if entertainment expenses represent kickbacks that distort decisions of economic agents,⁷³ especially if public officials receive benefits.⁷⁴ The lower the effects on third parties, the stronger the case that a revenue-neutral tax reform will be welfare neutral as to nonparticipants.

4. Enhancement of Experimental Interpretability

An additional caveat is that experimental subjects may behave differently than they would if they faced the same tax rules but outside an experimental context. This problem arises in medical experiments when participants may be able to deduce which group they are in.⁷⁵ In

70. I.R.C. § 274(n) (Supp. V 2018) (disallowing 50% of the deduction for most meal expenses).

71. *Cf. infra* Section II.A (considering generalizing experiments beyond initial participants).

72. *See infra* Section II.A.

73. *See* LEONARD J. BROOKS & PAUL DUNN, BUSINESS & PROFESSIONAL ETHICS FOR DIRECTORS, EXECUTIVES & ACCOUNTANTS 388 (7th ed. 2015).

74. A taxpayer may not deduct business expense payments made in violation of state or federal law. I.R.C. § 162(c)(2) (2012).

75. *See, e.g.,* Jefferson M. Fish, *The Trouble with Double-Blind Placebo Studies*, PSYCHOL. TODAY (Nov. 23, 2010), <https://www.psychologytoday.com/blog/looking-in-the-cultural->

social experiments, it is impossible to conceal group assignments. Taxpayers can respond to the economic incentives of a tax change only if they are aware of it, so there can be no placebo group.⁷⁶

“Hawthorne effects” occur when members of the treatment group behave differently because they know that they are in an experiment.⁷⁷ Subjects might, in a tax experiment, focus more on the relevant tax issue, giving it outsized importance. Or subjects might be regret averse;⁷⁸ that is, they wish to avoid feeling regret for the decision that they have already made to opt into the experiment.⁷⁹ In the entertainment deduction experiment, an exaggerated cutback on entertainment expenses reduces the risk that the taxpayer will find out that the taxpayer would have been better off with status quo law.

Meanwhile, “John Henry effects” occur when subjects in the control group behave differently than they would outside an experiment.⁸⁰ Annoyed at not being assigned to treatment, some might increase their entertainment expenses, so that they can profit by deducting even more than they would have. Or, they might cut back on such expenses, figuring they were planning to before being assigned to control. Determining which scenario is more likely is an exercise in speculative psychology. The premise of tax experimentation is that when taxpayers are faced with direct economic incentives, the treatment group’s behavioral responses are direct results of those incentives. But at least on the margins, psychological considerations related to the experimental setting may play a role.

A related but distinct concern is that group assignment might affect taxpayers’ reporting of their behavior. Control group taxpayers might decide to be more honest in reporting their entertainment deductions because they worry (even if falsely) that government investigators will be more likely to find fraud. Or, miffed at being assigned to control, such

mirror/201011/the-trouble-double-blind-placebo-studies [https://perma.cc/XET7-FVM7] (noting that patients often can determine whether a pill is a placebo or biologically active).

76. Double-blind medical experiments typically include three groups: a control group, a placebo group, and a treatment group. *See, e.g., id.*

77. *See* Stephen R. G. Jones, *Was There a Hawthorne Effect?*, 98 AM. J. SOC. 451, 452–53 (1992) (describing experiments in which such effects were claimed).

78. *See generally* Graham Loomes & Robert Sugden, *Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty*, 92 ECON. J. 805 (1982) (defining evidence for regret aversion).

79. For example, litigants may accept settlement offers because they wish to avoid the possibility of regret should they do worse at trial. *See* Chris Guthrie, *Better Settle Than Sorry: The Regret Aversion Theory of Litigation Behavior*, 1999 U. ILL. L. REV. 43, 45–46 (1999).

80. *See, e.g.,* Allen C. Barrett & Doris A. White, *How John Henry Effects Confound the Measurement of Self Esteem in Primary Prevention Programs for Drug Abuse in Middle Schools*, 36 J. ALCOHOL & DRUG EDUC. 87, 99 (1991) (describing an observed John Henry effect).

taxpayers might take shortcuts in reporting. Meanwhile, even if reporting is required for individuals in the treatment group and they face liability for misreporting,⁸¹ treated taxpayers might reason that their entertainment deductions no longer factor into their tax liability and thus not bother to collect all the underlying data. Or, they might reason that because the reported expense values will not reduce their liability, they might as well overreport to avoid any possible sanction for misreporting.

The tax authority might adopt various approaches to addressing these issues. The first and most plausible is simply to ignore them. The tax authority's principal job is to produce summary data to inform taxpayers, who can then make their own assessments of experimental results in deciding whether to opt in for future years. If the ultimate measure of a revenue-neutral tax experiment's success is demand to receive the tax treatment, then what matters is simply that the government report the data accurately. If, however, the goal is to enable both the tax authority and the taxpayer to make informed rather than speculative decisions, then some other approaches may be necessary if Hawthorne or John Henry effects are expected to be large. The following subsections will consider other approaches that will make experiments easier to interpret.

a. Varying the Treatment

An alternative strategy for reducing the magnitude of Hawthorne and John Henry effects is to define a range of treatment groups.⁸² For example, the experiment might feature eleven groups, with one receiving no deduction, one receiving 10% deductibility, and so forth. The group receiving the full deductibility allowed under current law would be the control group, and it might be larger than each other group.⁸³ Hawthorne and John Henry effects seem likely to be most pronounced at the extremes, when a taxpayer concludes that he is in the control or receiving the full treatment. Some components of Hawthorne or John Henry effects might gradually increase with the treatment level, but this approach would at least isolate the components of these effects that emerge at the extremes. An extrapolation of the trend between 10% and 90% to the

81. Ordinarily, tax law imposes no penalty on taxpayers who fail to file, but through withholdings have overpaid their tax, because penalties are based on the size of the deficiency. See *Patronik-Holder v. Comm'r*, 100 T.C. 374, 380 (1993).

82. This is often described as a dose-response design. See WILLIAM M. HOLMES, USING PROPENSITY SCORES IN QUASI-EXPERIMENTAL DESIGNS 17–18 (2014).

83. See, e.g., Simon Bate & Natasha A. Karp, *A Common Control Group—Optimising the Experiment Design to Maximise Sensitivity*, 9 PLOS ONE, Dec. 11, 2014, at 1, 9, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263717/pdf/pone.0114872.pdf> [<https://perma.cc/X4LU-CKV3>] (showing that with multiple treatments, sensitivity for comparison with the control is maximized with a larger control group).

extremes might be a more reliable gauge for policy than the extremes themselves.

Eleven is not a magic number. Taxpayers could be assigned to a smaller or larger number of groups. Revenue neutrality, however, must be retained. One strategy for achieving this with multiple treatment groups would be to calculate multipliers based on the performance of each treatment group individually, comparing to the single control group. But unless the number of taxpayers is very large, the multiplier levels might be noisy, influenced by randomness within the control group or a particular treatment group. Even if the control group and a treatment group as a whole are representative, a particular taxpayer might be matched⁸⁴ through regression to a relatively small number of taxpayers in the treatment group who had higher or lower than expected performance for reasons having nothing to do with the experiment.

An alternative approach would be to develop a single integrated regression model that allows a taxpayer's expected income, given prior years' data, to be estimated based on the degree of deductibility. The regression might include a term representing the degree of deductibility allowed, along with a square and perhaps a cube of that term. The regression would also include all other variables from previous years' tax data that allowed for multipliers to vary within the treatment group in the proposal.⁸⁵ Each non-control taxpayer's multiplier would then be calculated as the taxpayer's expected tax level if full deductibility were allowed, divided by the taxpayer's expected unadjusted tax level given the actual level of deductibility allowed.

Allowing for a wide range of treatment levels has an additional potential benefit beyond helping to highlight Hawthorne and John Henry effects: It may help identify tax changes where the optimum exists somewhere between the extremes. Perhaps the optimal degree of deductibility for entertainment expenses is somewhere between 0% and 100% because spending money on entertainment generally provides some consumption value but also contributes to taxpayers' income. On the other hand, if an experiment seems to suggest a result relatively near the zero-deductibility extreme, there may be a strong argument for tax law to move all the way to that extreme, namely that eliminating

84. The term "matching" is often used in statistics to refer to a particular experimental design that is an alternative to regression, in which baseline characteristics are used to divide subjects into treatment-control pairs. *See generally, e.g.,* Ruta Brazauskas & Brent R. Logan, *Observational Studies: Matching or Regression?*, 22 *BIOLOGY BLOOD MARROW TRANSPLANTATION* 557 (2016) (discussing this experimental concept). Whatever design is used, taxpayers will be affected by those with similar characteristics.

85. *See supra* notes 61–64 and accompanying text.

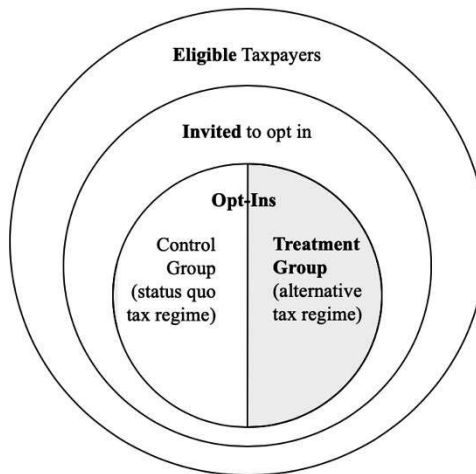
deductibility altogether is likely to reduce transaction costs.⁸⁶ If for political or transaction-costs reasons the law seems likely to settle at an extreme,⁸⁷ no matter the experimental results, then an experiment that simply compares a control group and a single treatment group may provide for a cleaner comparison.

b. Two-Level Randomization

With two-level randomization, the government first identifies taxpayers for whom the alternative regime might be appropriate. Then the government randomly selects a subset of these eligible taxpayers to be invited to opt in. Some taxpayers invited will decline. Of those who do opt in, some fraction must be randomly assigned to the control group, while others are subject to the alternative tax regime.

Figure 1 illustrates these different groups.

Figure 1. Two-Level Randomization



Two-level randomization affords two levels of comparison. As before, the government can compare nominal tax liability from those taxpayers subject to the alternative regime (i.e., those shaded gray in Figure 1) to tax revenues from the control group. But now, the government can also compare the tax returns of those taxpayers who were not invited to opt in, despite being eligible, to the tax returns of all taxpayers who were

86. For a discussion of transactions costs in the tax system, see Kneave Riggall, *Comprehensive Tax Base Theory, Transaction Costs, and Economic Efficiency: How to Tax Our Way to Efficiency*, 17 VA. TAX REV. 295, 307–08 (1997).

87. The current treatment of entertainment expenses is at the extreme of complete disallowance. I.R.C. § 274(a)(1) (Supp. V 2018). But meals with clients are 50% deductible. *Id.* § 274(n).

invited to opt in. Suppose, for example, that control group taxpayers seek to make up for the misfortune of being placed in the control group by working harder. This would be unfortunate for the members of the treatment group, who would receive less of a tax discount than they otherwise would, making the experiment seem less successful than it in fact was. The larger comparison might identify this behavior.

It may seem counterintuitive to make a comparison involving many taxpayers *not* participating in the experiment. One can, however, think of the eligible-but-invited taxpayers as serving as a control group of a sort for the treatment of being invited to participate in the experiment. Even assuming that being invited to participate has no direct effect on behavior,⁸⁸ Hawthorne and John Henry effects can be seen as indirect effects of being invited into the experiment. Two-level randomization allows for measurement of these indirect effects, which are likely to be more attenuated than the effect of being chosen for the treatment group but less susceptible to the Hawthorne and John Henry problems.

Two-level randomization thus might help allow better interpretation of experiments. But if two-level randomization is used, the results also could be used to calculate a second multiplier that would then be used to achieve more accurate revenue neutrality. A simple approach would be to multiply the treatment group's total tax bills by a constant, applied on top of the original multiplier, to ensure that the average tax revenues received from those invited to the experiment are equal to the average tax revenues received from those eligible but not invited. This would assure potential participants in a tax experiment that their ultimate tax liability will not be affected by John Henry or Hawthorne effects arising from the division into treatment and control groups. Alternatively, a multiplier might apply to both the treatment and control group, or even to all taxpayers invited into the experiment. Such an application might be justified on the ground of horizontal equity, but it would effectively punish control group taxpayers who work harder as a result of John Henry effects or reward those who work less hard as a result.

The tax adjustments to achieve revenue neutrality could be even more sophisticated. A separate multiple regression could model the eligible-but-uninvited taxpayers. The government would then be able to calculate the expected tax liability of each member of the treatment group if that member had not been invited to participate. This allows for an individualized second multiplier to be calculated for each treatment group member. A third multiplier could then be applied to all treatment group participants to achieve revenue neutrality. Arguably, however, such

88. If the experiment is well-known even among those not invited, the nonreceipt of an invitation might affect behavior. This can produce a sort of John Henry effect of its own. See *supra* note 80 and accompanying text. But these effects seem likely to be smaller than the effects on those who have taken the affirmative step of trying to opt in.

adjustments may make the system too opaque. Thus, the tax authority reasonably might seek to achieve revenue neutrality only with the first layer of multipliers, even if using two-level randomization. At least, that might make sense in early experimentation on experimentation. If large discrepancies between eligible-but-uninvited taxpayers and invited taxpayers emerged, then such refinements might be necessary.

c. Intent-to-Treat Randomization

Yet another approach would be for the government to conduct just a single layer of randomization, but for that layer of randomization to be at the invitation stage. The tax authority would invite only some eligible taxpayers to participate but then allow all taxpayers who volunteer for the experiment to receive the alternative tax regime. This approach is appropriate if there is relatively little reason to worry that merely being offered or denied the alternative regime will change taxpayer behavior. The approach may be useful if it is seen as undesirable to randomly pick a taxpayer to participate in an experiment but then assign the taxpayer to the control group. A standard statistical methodology called “intent to treat” can be used to determine the statistical significance of the results, comparing results of those offered the treatment (eligible and invited taxpayers) with those not offered it (eligible but not invited), taking into account that many may decline the treatment.⁸⁹ Revenue neutrality could then be achieved with a simple multiplier applied to all invited and opting-in taxpayers, set at a level ensuring that revenues are equal among invited and not-invited taxpayers.

B. Other Revenue-Neutral Tax Experiments

This section considers a variety of other possible tax experiments, including tax experiments affecting what is deductible, experiments on the definition of income, experiments on tax procedure, and radical tax experiments exploring significant structural changes to the structure of taxation.

89. See Vikrant K. Bubbar & Hans J. Kreder, *The Intention-to-Treat Principle: A Primer for the Orthopaedic Surgeon*, 88 J. BONE & JOINT SURGERY 2097, 2099 (2006).

1. Deductions

Experiments could test both disallowing what otherwise would be deductions and allowing deductibility where it otherwise would not be possible.

a. Disallowing Other Deductions

The deduction for entertainment expenses has served as a useful example of what could have been an alternative to the Tax Cuts and Jobs Act because the Act eliminated the deduction while lowering rates.⁹⁰ It is useful in part because it is relatively trivial and uncontroversial, allowing the focus to be on the merits of revenue-neutral experimentation rather than the merits of the entertainment deduction itself. Yet revenue-neutral experimentation also could have been used to assess the impact of eliminating other deductions. For example, the tax reform reduced the availability of the business interest deduction.⁹¹ A justification for this reform is that the tax law previously advantaged equity relative to debt because equity is taxed twice.⁹² President Obama proposed eliminating the deduction in part because it increases leverage in the economy as a whole,⁹³ making it “more susceptible to severe downturns.”⁹⁴ Others have argued that it would be better to eliminate the double taxation of equity⁹⁵ or to allow a deduction based on a corporation’s combined debt and equity.⁹⁶ Meanwhile, there is a strong theoretical argument that business interest should be deductible.⁹⁷ A tax experiment might have assessed one or more possible reforms. High demand to participate in such an experiment would suggest a view that the current approach to business interest imposes significant distortions. Meanwhile, the government or

90. See Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, § 13304, 131 Stat. 2054, 2124–25 (amending I.R.C. § 274(a) (2012)).

91. See *id.* § 13301 (amending I.R.C. § 163(j)).

92. See, e.g., Curtis Dubay, *Taxation of Debt and Equity: Setting the Record Straight*, HERITAGE FOUND. (Sept. 30, 2015), <https://www.heritage.org/taxes/report/taxation-debt-and-equity-setting-the-record-straight> [<https://perma.cc/W57K-CUCX>].

93. WHITE HOUSE & DEP’T OF THE TREASURY, THE PRESIDENT’S FRAMEWORK FOR BUSINESS TAX REFORM: AN UPDATE 9 (2016), <https://www.treasury.gov/resource-center/tax-policy/Documents/The-Presidents-Framework-for-Business-Tax-Reform-An-Update-04-04-2016.pdf> [<https://perma.cc/3P52-DJRF>].

94. *Id.*

95. See Dubay, *supra* note 92 (arguing that eliminating double taxation should be preferred to eliminating the deduction).

96. Katherine Pratt, *The Debt-Equity Distinction in a Second-Best World*, 53 VAND. L. REV. 1055, 1058 (2000).

97. See, e.g., Charles O. Galvin, *The Deduction of Nonbusiness Interest: An Exercise in Planned Confusion*, 41 TAX LAW. 803, 803 (1988) (explaining the logic both for an income tax base and for a consumption tax base).

third parties could study how different tax treatment affects decisions about whether to finance with debt or equity.

Another possible subject for experimentation before enactment would have been one of the most controversial changes⁹⁸ in the statute: the imposition of caps on the deductibility of state and local taxes.⁹⁹ Many commentators have claimed that these caps represented a purely political calculation, as the states most adversely affected by this change were those with high taxes, and such states are generally “blue states,” which lean Democratic rather than Republican.¹⁰⁰ This highlights that tax policy has distributional consequences, not just efficiency consequences. Yet some argue that these deductions may cause states and localities to oversupply goods that might be more optimally supplied by the market,¹⁰¹ eliminating the tax allows taxpayers to be taxed on the consumption benefits that they receive.¹⁰² There exist counterarguments¹⁰³ and compromise proposals.¹⁰⁴ A revenue-neutral tax experiment on individual taxpayers could not allow full examination of these issues, because such an experiment could not establish how the existence of the deduction affects state and local policy. But it might highlight whether taxpayers believe that the deduction distorts their own behavior, for example by leading them to locate in areas with higher taxes.

b. Allowing New Deductions

Revenue-neutral tax experiments can also be used to test the efficiency of new deductions. A tax experiment, for example, could be used to test the possibility of reintroducing the entertainment expenses deduction. A taxpayer might believe that the new tax regime is inefficient, causing the taxpayer to spend too little on entertainment

98. A poll indicated that 75% of voters believed that such taxes should remain deductible. See Jonathan Easley, *Poll: Majority Oppose GOP Tax-Reform Bill*, HILL (Nov. 16, 2017, 12:38 PM), <http://thehill.com/policy/finance/360693-poll-majority-oppose-gop-tax-reform-bill> [<https://perma.cc/GPD5-5P4P>].

99. Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, § 11042, 131 Stat. 2054, 2085–86 (amending I.R.C. § 164 (2012)).

100. See, e.g., Alicia Parlapiano & K.K. Rebecca Lai, *Among the Tax Bill's Biggest Losers: High-Income, Blue State Taxpayers*, N.Y. TIMES, <https://www.nytimes.com/interactive/2017/12/05/us/politics/tax-bill-salt.html> [<https://perma.cc/Q3GZ-PC4E>] (last updated Dec. 5, 2017).

101. See, e.g., Louis Kaplow, *Fiscal Federalism and the Deductibility of State and Local Taxes Under the Federal Income Tax*, 82 VA. L. REV. 413, 417 (1996).

102. See Brian Galle, *Federal Fairness to State Taxpayers: Irrationality, Unfunded Mandates, and the “SALT” Deduction*, 106 MICH. L. REV. 805, 813 (2008).

103. Kaplow, *supra* note 101, at 486 (noting that the deduction may promote spending on undersupplied public goods).

104. See Gladriel Shobe, *Disaggregating the State and Local Tax Deduction*, 35 VA. TAX REV. 327, 327 (2016) (arguing for the deductibility of state but not local taxes).

relative to other ways of recruiting clients. Such a taxpayer should be willing to opt into an experiment in which the treatment group would receive the deduction. The taxpayer would then be subject to a multiplier that would *increase* nominal tax liability to ensure revenue neutrality. Thus, the participation carrot and stick are reversed, but the experiment can be run as before. Similarly, experiments could test reducing the limits on the business interest deduction and the state and local tax deduction; such experiments could be executed simultaneously with experiments on *increasing* the limits on such deductions.

Meanwhile, experiments could test the possibility of new deductions, such as a commuting deduction, perhaps limited to the cost of commuting from a workplace to the nearest location with a supply of affordable housing.¹⁰⁵ More ambitiously, an experiment might test a deduction for child care, perhaps limited to married taxpayers who both have full time jobs. The current lack of deductibility may lead parents (particularly mothers) not to re-enter the work force, even though they would do so in the absence of tax distortions.¹⁰⁶ A married couple might opt into such an experiment because they anticipate that many in the control group will not return to the work force. Couples in the treatment group might return to the work force in greater numbers, paying the same level of taxes as the control group on average but less than if they returned to the work force in the absence of the experiment.

2. Income

a. Imputing Income

An alternative policy that might cure the same alleged inefficiency would be to test imputing income to stay-at-home parents.¹⁰⁷ The theory is that a parent who works at home in effect is paying herself to take a job, but her income from doing so is not taxed.¹⁰⁸ Economists sometimes argue that imputed income should be taxable,¹⁰⁹ but taxing imputed income seems politically infeasible. Even so, parents might be willing to opt into an experiment in which time spent on childrearing, perhaps just during work hours but perhaps more broadly, is taxed. Tax experiments are not limited to questions of what count as deductions, and taxpayers

105. *See supra* notes 42–44 and accompanying text.

106. *See, e.g.*, Shannon Weeks McCormack, *Postpartum Taxation and the Squeezed out Mom*, 105 GEO. L.J. 1323, 1350–61 (2017) (proposing to allow deductibility).

107. *See, e.g.*, Nancy C. Staudt, *Taxing Housework*, 84 GEO. L.J. 1571, 1618–36 (arguing for imputing income from household labor).

108. *Id.* at 1576.

109. *See, e.g.*, Tsilly Dagan, *Itemizing Personhood*, 29 VA. TAX REV. 93, 115–16, 116 n.45 (2009) (discussing imputed income and citing sources).

might volunteer to accept a greater tax base (leading to greater taxable income) in exchange for lower tax rates (because of revenue neutrality). This experiment might interest similar taxpayers as an experiment on a child care deduction, but there could be some differences. Taxpayers who do not expect to be able to itemize deductions, for example, might still be interested in an imputed income experiment.¹¹⁰ Taxpayers with older children who can be left alone might prefer this experiment as well.¹¹¹

An argument against an imputed income experiment is that there is little reason to conduct an experiment on a tax policy that ultimately will be politically infeasible. On the other hand, perhaps experimentation might lead imputing income to become more palatable, particularly in combination with other policies lowering tax rates for parents. Even if this experiment is impractical, other imputed income experiments might be feasible. The second greatest category of tax expenditure as calculated by the Treasury arises from the lack of imputed income for rent from owners of housing who live in their homes.¹¹² A tax experiment might seek to move tax to the “baseline tax system”¹¹³ by imputing income for rent and then allowing “a deduction for expenses, such as interest, depreciation, property taxes, and other costs, associated with earning such rental income.”¹¹⁴ Such an experiment might help eliminate distortions leading to excessive owner-occupied housing.¹¹⁵ In principle, similar experiments could test other forms of imputed income.¹¹⁶

b. Taxing Work Amenities

Beyond the question of imputed income, experiments on definition of income can explore fundamental questions about whether the classic

110. Taxpayers who would be below the standard deduction even with child care expenses would not benefit from deductibility. See I.R.C. § 63(c) (2012 & Supp. 2018) (providing for a standard deduction).

111. This highlights that tax policies may have externalities. See *infra* Section II.D. Imputing income for child care might result in parents leaving their children alone more often, for better or more likely for worse.

112. DEP'T OF THE TREASURY, TAX EXPENDITURES 22 tbl.1 (2017).

113. *Id.* at 9–10 (arguing that imputed income is taxed in the baseline).

114. *Id.* at 10.

115. This thus implicates some of the same externality issues as the home mortgage interest deduction. See *infra* notes 227–30 and accompanying text.

116. There is, however, likely some limit, even if political feasibility is placed to the side. John Brooks asks, “If I am earning imputed rent from my home, what about from my car? My furniture? My computer? My dishwasher? My dishes?” John R. Brooks, *The Definitions of Income*, 71 TAX L. REV. 253, 255 (2018).

Haig¹¹⁷-Simons¹¹⁸ model is appropriate or can test small deviations from the Haig-Simons model. John Brooks critiques Henry Simon's argument that "psychic benefits" in income can be ignored because all workers are equally affected.¹¹⁹ Brooks points out that "[t]here is enough heterogeneity of psychic benefits across jobs and individuals to make universal assumptions unreasonable."¹²⁰ In theory, a tax authority might conduct a survey to rate the attractiveness of different jobs and impose a corresponding tax increase on relatively pleasant and cushy jobs—or, equivalently given revenue neutrality, a tax decrease for unpleasant jobs. Taxpayers might be willing to opt into such an experiment if they are willing to switch to a less pleasant job. A much more modest experiment aimed at the same theoretical point might seek to tax amenities at work, ranging from free meals¹²¹ to employee gyms.¹²²

c. Changing Recognition Timing

Tax experiments also might apply to issues of timing. Some tax scholars have advocated switching to a system in which investments could be marked to market, meaning that taxpayers would recognize gains and losses each year even if they do not sell the securities.¹²³ A tax experiment might help evaluate some of the criticisms of a mark-to-market system, such as that it imposes substantial transactions costs,¹²⁴ especially if the regime applies beyond publicly held securities to assets

117. Robert Murray Haig, *The Concept of Income—Economic and Legal Aspects*, in *THE FEDERAL INCOME TAX 1*, 7 (Robert Murray Haig ed., 1921) ("Income is the money value of the net accretion to one's economic power between two points of time.").

118. See HENRY C. SIMONS, *PERSONAL INCOME TAXATION: THE DEFINITION OF INCOME AS A PROBLEM OF FISCAL POLICY* 50 (1938) (defining income as "the result obtained by adding consumption during the period to 'wealth' at the end of the period and then subtracting 'wealth' at the beginning").

119. Brooks, *supra* note 116, at 264 ("[T]hese elements of unmeasurable psychic income may be presumed to vary in a somewhat continuous manner along the income scale.") (quoting SIMONS, *supra* note 118, at 53).

120. *Id.* at 265.

121. Meals are excluded from income if "furnished on the business premises of the employer" and "for the convenience of the employer." I.R.C. § 119(a) (2012).

122. "Gross income does not include the value of any on-premises athletic facility provided by an employer to its employees." Treas. Reg. § 1.132-1(e) (1989).

123. Currently mark-to-market is normally available solely for securities dealers like stock brokerages. I.R.C. § 475. *But see* I.R.C. § 1256 (West 2018) (providing mark-to-market on sophisticated financial instruments like future contracts and foreign currency contracts).

124. See Edward A. Zelinsky, *For Realization: Income Taxation, Sectoral Accretionism, and the Virtue of Attainable Virtues*, 19 *CARDOZO L. REV.* 861, 879–89 (1997) (providing detail on the issue of transactions costs).

such as art.¹²⁵ A significant complication with a tax experiment of this sort is that the tax experiment must occur over a relatively long time horizon. This Article will return to this issue, along to the related danger that participants in tax experiments may seek to shift income into or out of the experimental period, below.¹²⁶

3. Tax Procedure

Revenue-neutral tax experiments can also be used to test procedural changes. For example, an experiment might test a regime in which the government would agree to provide binding opinions on tax questions by phone or email. Because taxpayers might then take advantage of mistaken statements of tax law by tax authority employees,¹²⁷ the treatment group would likely pay lower taxes than the control group, leading to an ex post multiplier increasing tax rates. Ideally, the ex post adjustment should also compensate for the extra expenses borne by the tax authority in providing extra customer service to the treatment taxpayers.

Experiments could also test taxpayer-adverse procedural changes. For example, some taxpayers with relatively complex returns (say, taxpayers with foreign bank accounts) might opt into a regime in which those taxpayers agree to submit with their tax returns a report submitted by a privately selected auditor. Such taxpayers presumably would be less likely to engage in tax evasion. But in expectation, they would pay no higher taxes than before, since the multiplier would be less than one. Such an experiment could provide valuable information to the tax authority. If the control group taxpayers reported much less in nominal taxes before application of the multiplier, that would indicate a high degree of tax avoidance and a high degree of effectiveness of a private audit requirement. If, on the other hand, differences were small, incorporating the program into the baseline tax regime would generate little revenue.

4. Radical Tax Reform

Revenue-neutral tax experimentation can be used not only to assess relatively minor changes in deductions, but also significant differences in policy. For example, scholars have proposed abolishing the corporate tax—which historically took an average of 27% of each company's

125. See Noel B. Cunningham & Deborah H. Schenk, *Taxation Without Realization: A "Revolutionary" Approach to Ownership*, 47 TAX L. REV. 725, 801–02 (1992) (discussing applicable rules for art objects).

126. See *infra* Section II.C.2 (addressing income and deduction shifting).

127. Currently, when a taxpayer calling the IRS gets an answer from an IRS employee, the taxpayer has no legal entitlement to rely on that answer. Emily Cauble, *Detrimental Reliance on IRS Guidance*, 2015 WIS. L. REV. 421, 431–37 (2015).

profits¹²⁸—and instead granting the government a roughly equivalent claim on the corporation’s equity. Thus, the government might receive 27% of a corporation’s stock.¹²⁹ This reform promises significant efficiency gains, as tax-minimization goals would no longer distort decisions, and because corporations would no longer need legions of well-paid tax advisors.¹³⁰

It might seem that this is not amenable to a revenue-neutral tax experiment. If the government receives equity, there is no way to ensure that the equity that the corporation contributes will produce equal revenue over time. But a change in the experimental design could enable the reform proposal to be tested. The government could auction rights to some percentage of the tax revenue that it will receive from both the treatment group and the control group. The government might then adjust the ownership percentage that the government takes until both revenue streams sell to the market for the same price. For example, the government might initially offer 27%, but then increase this if it was unable to find a sufficient number of purchasers of the revenue stream at that price. As the government changes the ownership percentage, some corporate taxpayers might change their mind about whether to participate in the experiment. But ultimately, the government should be able to identify an ownership percentage that is revenue neutral in expectation.

That does not mean that every radical tax reform is amenable to tax experimentation. It seems unlikely, for example, that a value added tax could be implemented on an experimental basis.¹³¹ A value-added tax requires each producer to pay tax on the value it adds; so, for example, a producer purchasing a product for \$100 and selling it for \$150 would pay tax only on the \$50 difference.¹³² If enacted economy-wide, then a combination of taxpayers in the value chain pay tax on all of the \$150. If

128. JANE G. GRAVELLE, CONG. RESEARCH SERV., R41743, INTERNATIONAL CORPORATE TAX RATE COMPARISONS AND POLICY IMPLICATIONS 3 (2014).

129. See Dean Baker, Opinion, *A Progressive Way to End Corporate Taxes*, N.Y. TIMES (Jan. 12, 2016), <https://www.nytimes.com/2016/01/13/opinion/a-progressive-way-to-replace-corporate-taxes.html> (providing additional details of this proposal, including that the shares should be nontransferable); cf. Mihir A. Desai et al., *Theft and Taxes*, 84 J. FIN. ECON. 591, 592 (2007) (“The state, thanks to its tax claim on cash flows, is *de facto* the largest minority shareholder in almost all corporations.”).

130. Baker, *supra* note 129.

131. Consumption taxes that are administered in much the same way as income taxes, rather than like sales taxes, could be a subject of experimentation. See generally David F. Bradford, *The X-Tax in the World Economy* (Nat’l Bureau of Econ. Research, Working Paper No. 10676, 2004), <https://www.nber.org/papers/w10676.pdf> [<https://perma.cc/2LWF-MN2V>] (describing such an approach to taxation).

132. For a discussion of different ways of calculating the tax, see generally Itai Grinberg, *Where Credit Is Due: Advantages of the Credit-Invoice Method for a Partial Replacement VAT*, 63 TAX L. REV. 309 (2010).

only some producers were required to pay tax, the system could become much more complex. Value-added taxes already create challenges of international harmonization,¹³³ but this would magnify those challenges on a domestic level.

II. CHALLENGES FOR TAX EXPERIMENTATION

This Part addresses challenges for tax experimentation, specifically questions of how to transition and scale up successful experiments, whether tax experiments will adversely affect tax equity, whether taxpayers might be able to manipulate tax experimentation to their advantage, and whether tax experiments can work with tax expenditures without adversely affecting other tax policy goals.

A. Transitions

This Section discusses how the government can determine whether to scale up an experiment or transition to a later stage of experimentation, such as an opt-out or involuntary experiment, prior to enactment as general law.

1. Increased Scope

The goal of tax experimentation is to provide information that can change baseline tax policy. Although appropriate experimental design can improve policymakers' ability to judge whether an experiment is a success,¹³⁴ the most easily accessible benchmark of success is simply participation. Given the constraint of revenue neutrality, taxpayer demand to be in the experiment suggests that the tax change is Pareto-improving, at least so long as the tax change does not induce behaviors that have effects on third parties.¹³⁵ But that leaves unclear whether the experiment would benefit others. Perhaps other taxpayers who might benefit from the experiment have not signed up simply because they did not know about it. But they may have chosen not to enroll because they expect that they would do less well under the experimental conditions.

Thus, a first step in transitioning from experiment to a tax law change is to invite more taxpayers to receive the treatment. If in initial rounds only some taxpayers were invited to participate in the experiment, then more might be allowed to participate, still assigning the same proportion of enrollees to treatment and control as before. With two-level randomization,¹³⁶ for example, a higher proportion of eligible taxpayers

133. See, e.g., *id.* at 321–22 (discussing some international issues).

134. See *supra* Section I.A.4 (addressing enhancement of experimental interpretability and appropriate experimental design).

135. The possibility of externalities is addressed *infra* Section II.D.

136. See *supra* Section I.A.4.b (addressing two-level randomization).

might be invited to participate. Alternatively or as a supplement, a higher percentage of opt-ins might be selected for the treatment group. Indeed, if an experiment continues to be successful, it is possible that all taxpayers who wish to opt-in might ultimately be assigned to the treatment group.

At this point, the experiment ceases to be an experiment. Rather, some may see it as tax reform in and of itself, though of a different form from what one might expect. The tax reform amounts to giving taxpayers the *option* to elect a particular tax regime in the subsequent year, in combination with a tax rate change.¹³⁷ Tax law already allows taxpayers to elect various options.¹³⁸ If the change is still to aim at revenue neutrality, then, as in the experiment itself, different taxpayers should receive different rate reductions based on data from past tax returns. Without a control group, it will be impossible to do this precisely, though the government might approximate this by using data from when a control group still existed. Making the treatment generally available will not likely reduce government revenues, as those who did not opt into the experiment in the past will tend to be those who benefit less from the experimental treatment.

There may, however, be a strong reason not to allow all taxpayers into the treatment group, even if all taxpayers (or at least all taxpayers who know about the program and might plausibly be affected by it) would choose to opt in. Once an alternative tax regime has expanded to be available as an option to all taxpayers meeting specified criteria, then it becomes impossible to run an experiment confirming that the alternative regime continues to produce at least as much tax revenue as the generally applicable tax law. The same problem applies with medical trials: once a drug is generally available, there must be “clinical equipoise” if the drug is to be tested against placebo.¹³⁹ Thus, the case for maintaining a control group is similar to the case for revenue-neutral tax experimentation generally. Even if only a few taxpayers wind up in the control group, this allows for continued study of the tax rule, which may be especially important if the efficiency of the rule changes over time. This is especially true with revenue-neutral experimentation, since the control group serves a critical role in allowing the tax rate to vary across treatment group taxpayers.

137. I.R.C. § 451(c)(2)(B) (West 2018).

138. See, e.g., *id.* § 451(c)(2) (allowing taxpayers to elect a particular approach to advance payments).

139. See Franklin G. Miller & Howard Brody, *What Makes Placebo-Controlled Trials Unethical?*, 2 AM. J. BIOETHICS 3, 3 (2002); see also Alex John London et al., *Rethinking Research Ethics: The Case of Postmarketing Trials*, 336 SCIENCE 544 (2012) (discussing ethical issues in postapproval studies).

2. Opt-Out Experiments

A voluntary tax experiment can either be opt-in, where a taxpayer who takes no action remains subject to generally applicable tax law, or opt-out, where a taxpayer will be subject to the alternative tax regime without affirmatively taking a step to elect generally applicable law. So far, we have assumed that all tax experiments would be opt-in. This approach fits better into the landscape of typical randomized governmental experiments.¹⁴⁰ Such experimentation is generally rare, but when the government has engaged in it, usually there is some benefit (e.g., a welfare program,¹⁴¹ a child's eligibility for a school voucher program¹⁴²) that the government makes available to volunteers (i.e., those who opt in). Randomization thus has the dual benefit of allocating a scarce resource and providing the government with better information. Moreover, such experiments help ensure that experimentation is Pareto-optimal, since properly informed taxpayers will opt into an alternative tax regime only if they expect it to increase their utility.

In an opt-in experiment, the government should provide a meaningful disclosure to invited taxpayers. At least, such a disclosure should explain that assignment to the treatment group will change the taxpayer's baseline tax liability, but that later adjustments will ensure that treatment taxpayers on average pay on average the same amount of taxes as those in the control group. The disclosure should also indicate how long the taxpayer will be subject to the treatment regime. We have assumed that experiments would be for one year only, but taxpayers could be placed into a treatment group for longer, particularly if the relevant tax provisions affect behavior with long-lasting tax consequences, such as purchasing depreciable assets.¹⁴³

Such disclosures could also be used in opt-out experiments, and indeed may be more important given that participation is the default. A transition from an opt-in to an opt-out experiment allows for an intermediate step, less drastic than permanent experimental adoption. Behavioral economics teaches that there may be large differences in

140. See generally Alice M. Rivlin & P. Michael Timpane, *Introduction and Summary*, in *ETHICAL AND LEGAL ISSUES OF SOCIAL EXPERIMENTATION 1* (Alice M. Rivlin & P. Michael Timpane eds., 1975) (discussing a wide range of experiments).

141. See, e.g., KERSHAW & FAIR, *supra* note 16, at 4.

142. See generally, e.g., Paul E. Peterson et al., *School Vouchers: Results from Randomized Experiments*, in *THE ECONOMICS OF SCHOOL CHOICE 107* (Caroline M. Hoxby ed., 2003) (discussing school voucher programs).

143. See I.R.C. § 167 (2012) (allowing depreciation deductions for assets held for business or investment purposes); see also I.R.C. § 168 (2012 & Supp. V 2018) (providing method for calculating depreciation deductions for tangible property, typically over multiple years, even decades).

responses depending on whether an experiment is opt-in or opt-out.¹⁴⁴ Thus, a change in an experiment from opt-in to opt-out is a plausible strategy for expanding the scope of an experiment, including far larger numbers of taxpayers and reducing the risk that any beneficial experimental outcomes are due to selection effects. The principal drawback is that opt-out creates the risk that many taxpayers who would not wish to participate will be enrolled because of inertia or because they ignored any information that they received about the experiment. Opt-out thus does less to ensure that the alternative tax regime increases expected taxpayer utility.

Nonetheless, an opt-out experiment remains less coercive than an actual change in the law. If the tax change being experimented with is a plausible candidate for ultimate adoption into the tax code, then an opt-out experiment is a modest step. Moreover, revenue neutrality may make opt-out experimentation less problematic than it would be with some other experiments. Revenue neutrality does not mean that a taxpayer should be indifferent as to which group the taxpayer is assigned; after all, a primary justification for revenue-neutral experimentation is that even revenue-neutral tax changes can benefit taxpayers. But revenue neutrality is a significant constraint on tax experiments. Members of the treatment group will be harmed only if the tax experiment in fact turns out to provide less efficient incentives to those in the group. The tax authority presumably will choose experiments that it believes have a substantial chance of producing some benefit, so taxpayers who fail to consider the merits of the experiment will likely not be harmed by being in the treatment group.

3. Involuntary Experimentation

Revenue neutrality also makes an involuntary tax experiment more plausible than involuntary social experiments ordinarily would be. At the least, involuntary experimentation is a useful step after an opt-out experiment before a tax law change is adopted for all taxpayers. If the experiment is to continue at this stage, presumably the tax authority has concluded at least tentatively that the tax change is beneficial. Thus, it seems likely that those placed in the treatment group benefit relative to those in the control group. The virtue of an involuntary experiment is that it eliminates concerns about selection effects, virtually ensuring that any

144. See William J. Congdon et al., *Behavioral Economics and Tax Policy*, 62 NAT'L TAX J. 375, 375–76 (2009) (noting the distinction in result between requiring opt-out to offering opt-in). See generally Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSP. 193 (1991) (discussing the status quo bias); William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988) (same).

difference discerned between the treatment and control groups would apply if the tax law change were made universal.

Some taxpayers who might not opt into an opt-in experiment or who might opt out of an opt-out experiment nonetheless might be happy to be included in an involuntary experiment. This would not be true with most social experiments but is a result of revenue neutrality. A taxpayer might believe that a revenue-neutral elimination of the entertainment deduction would be efficient as applied to its activities, but the taxpayer still might worry that the other taxpayers willing to participate in a voluntary experiment are those who expect their entertainment expenses to fall for largely exogenous reasons. This adverse selection is no longer a concern with involuntary experimentation.

The argument for involuntary experimentation is thus akin to an argument for mandatory purchases of insurance. Coverage requirements exist in part for this reason in the automobile insurance market,¹⁴⁵ and individuals are required to obtain health insurance under the Affordable Care Act for this reason.¹⁴⁶ In the health insurance context, the worry is that relatively healthy good risks will decide not to buy health insurance, making the market less attractive for slightly-less-good risks and potentially leading to a “death spiral” in which no one buys insurance,¹⁴⁷ even though many would like insurance at actuarially fair rates. In this context, a death spiral would be an experiment in which no one participates, even though many would participate absent adverse selection. In principle, this prospect can provide a normative case for involuntary experimentation even as the first step of an experiment, though that might be politically untenable.

In the insurance context, however, a coverage requirement is equitable; all drivers must purchase coverage, not half of all drivers chosen at random. Whether experiments are opt-in, opt-out, or mandatory, the prospect of horizontal inequity is likely to be the most significant obstacle to tax experimentation. This Article thus turns to that concern.

145. See Mila Araujo, *Understanding Minimum Car Insurance Requirements*, BALANCE, <https://www.thebalance.com/understanding-minimum-car-insurance-requirements-2645473> [<https://perma.cc/5T8S-TJ3T>] (last updated Nov. 1, 2018) (noting that forty-seven states require automobile insurance).

146. The Tax Cuts and Jobs Act eliminated the tax for noncompliance by setting the penalty to \$0. See Tax Cuts and Jobs Act of 2017, Pub. L. No. 115–97 § 11081, 131 Stat. 2054, 2092 (amending I.R.C. § 5000A (2012)).

147. For an argument that the risk is exaggerated, see Peter Siegelman, Essay, *Adverse Selection in Insurance Markets: An Exaggerated Threat*, 113 YALE L.J. 1223, 1254–58 (2004).

B. Equity

This section addresses equity objections to tax experimentation, considering both horizontal equity and after-tax income inequality.

1. Horizontal Equity

a. The Case for Randomness

Horizontal equity is an often-cited tax policy, the gist of which is that people with the same income should pay the same amount of tax.¹⁴⁸ The concern that tax experimentation might violate horizontal equity reflects the more general argument that randomization of government policy is inherently unequal.¹⁴⁹ If the government randomizes taxpayers who have opted into an alternative tax regime into a control group, subject to the generally applicable law, then the control-group taxpayers are arguably being treated inequitably. Similarly, with two-level randomization, taxpayers who are eligible for an alternative tax regime, but who are not invited to opt in, are arguably being treated inequitably compared to those invited.¹⁵⁰

The general consensus in the literature has been that experimentation is acceptable if there is a sufficient justification for the difference in treatment.¹⁵¹ Involuntary experimentation is most common in contexts in which individuals are thought to have lost their rights as a result of committing crimes.¹⁵² In one notable criminal justice experiment, domestic violence perpetrators in the Bronx were randomly assigned to one of four different treatment programs.¹⁵³ That experiment concluded

148. See BORIS I. BITTKER & LAWRENCE LOKKEN, FEDERAL TAXATION OF INCOME, ESTATES, AND GIFTS ¶ 3.1.4 (Thomson Reuters 2d/3d ed. 1993 & Supp. II 2018); see also Alan J. Auerbach & Kevin A. Hassett, *A New Measure of Horizontal Equity*, 92 AM. ECON. REV. 1116, 1116 (2002) (“There is virtual unanimity that horizontal equality, treating equals equally, is a worthy goal of any tax system.”).

149. See Abramowicz et al., *supra* note 9, at 967. For the counterargument, see Adam M. Samaha, *Randomization in Adjudication*, 51 WM. & MARY L. REV. 1, 18–21 (2009) (arguing that randomization can be consistent with equality).

150. See *supra* Section I.A.4.b.

151. See, e.g., Abramowicz et al., *supra* note 9, at 967–68; Maurice Rosenberg, *The Impact of Procedure-Impact Studies in the Administration of Justice*, 51 LAW & CONTEMP. PROBS. 13, 16 (1988).

152. See, e.g., Denise C. Gottfredson & M. Lyn Exum, Research Note, *The Baltimore City Drug Treatment Court: One-Year Results from a Randomized Study*, 39 J. RES. CRIME & DELINQ. 337, 343 (2002) (assigning offenders either to a drug court or to standard criminal process).

153. MELISSA LABRIOLA ET AL., CTR. FOR CT. INNOVATION, TESTING THE EFFECTIVENESS OF BATTERER PROGRAMS AND JUDICIAL MONITORING: RESULTS FROM A RANDOMIZED TRIAL AT THE BRONX MISDEMEANOR DOMESTIC VIOLENCE COURT v (2005), <http://www.courtinnovation.org/sites/default/files/battererprogramseffectiveness.pdf> [<https://perma.cc/985N-Q6WA>].

that treatment programs commonly employed throughout the nation for batterers may not be effective.¹⁵⁴ An experiment giving individuals who had *not* been convicted of domestic violence incentives to enroll in various programs might well seem inequitable.

The most common justification for randomization in government experiments is scarcity.¹⁵⁵ For example, experiments on supplements to the Earned Income Tax Credit (EITC)¹⁵⁶ have given the benefits to only some eager to participate.¹⁵⁷ Because the government expects to lose money on each member of the treatment group, revenue constraints limit the number of people included. Revenue neutrality might seem to weaken the scarcity justification for a tax experiment, as neither group costs the government more than the other. For revenue neutrality to work, however, there must be a control group, so that multipliers can be calculated. Membership in the treatment group is thus inherently scarce.

A scarcity argument is much more difficult to make in the context of involuntary experimentation. With a mandatory experiment, taxpayers forced into the treatment group who may be treated inequitably. But this is just a question of baselines. One could frame the treatment group as embodying new legal policy, subject to confirmation of the experiment's success. Then, membership in the control group is scarce, and so treatment group members are not being treated inequitably. The groups are being treated differently, but only because membership in each group must be scarce for an experiment to proceed.

These arguments for scarcity may seem artificial because the desired form of taxation, whether it is the treatment or the control, could at least in principle be given to all taxpayers, just not in the form of an experiment. But that is true with all social experimentation. Indeed, even with medical experimentation, scarcity is artificial. Presumably, all patients prefer the treatment, but we insist that they take some risk of

154. *Id.* at ix (“Regrettably, our study suggests that some of the most prevalent court responses to domestic violence crime may be ineffective . . .”).

155. *See, e.g.*, DAVID GREENBERG ET AL., SOCIAL EXPERIMENTATION AND PUBLIC POLICYMAKING 225 (2003) (“Random assignment usually became more acceptable [when officials] recognized that they did not have sufficient funding to serve their entire caseload and, hence, that some mechanism was needed to determine who would be denied services.”).

156. I.R.C. § 32 (2012 & Supp. V 2018). *See generally* BITTKER & LOKKEN, *supra* note 148, ¶ 37.1 (discussing the EITC in detail).

157. An experiment called “Paycheck Plus” on supplementing the EITC for single workers without children, for whom the EITC is very limited, has been ongoing in New York City and Atlanta, Georgia. *See* Rachel Pardoe & Dan Bloom, *Paycheck Plus: A New Antipoverty Strategy for Single Adults*, MDRC POL’Y BRIEF (MDRC, New York, N.Y.), May 2014, at 1, <http://www.mdrc.org/publication/paycheck-plus/file-full> [<https://perma.cc/LU73-XNHP>]. This program involves both treatment groups (receiving the extra EITC) and control groups (who do not). *Id.* at 4.

being placed in the control group because we are not sure that the treatment is better than placebo.¹⁵⁸ The baseline in which the patients do not have access to the treatment is purely a result of law, both in the medical context and in the tax context.

The ultimate justification for experimentation in both contexts is informational.¹⁵⁹ As with medical experiments, the justifiability of a social experiment depends on an evaluation of whether the treatment has a significant chance of being beneficial and whether the experiment may succeed in producing useful information.¹⁶⁰ One might argue that tax experimentation is more troubling than other legal experiments because horizontal equity is an important tax-law goal.¹⁶¹ On this argument, it might be acceptable to have random experiments concerning patent policy,¹⁶² where the primary goal is to maximize economic efficiency,¹⁶³ but less acceptable to allow inequalities to creep into tax policy. Some tax scholars, however, have criticized horizontal equity as a meaningless concept.¹⁶⁴ Indeed, economist Richard Musgrave argued that the only purpose of horizontal equity was as a “safeguard against capricious discrimination—a safeguard which might be provided equally well by a *requirement that taxes be distributed at random.*”¹⁶⁵

Tax law has long used randomization.¹⁶⁶ For example, the IRS uses randomization in selecting which taxpayers’ returns to audit.¹⁶⁷

158. See Miller & Brody, *supra* note 26, at 157.

159. Abramowicz et al., *supra* note 9, at 965 (“[R]andom policy experimentation . . . will produce better information than nonrandomized experiments.” (emphasis omitted)).

160. Macartan Humphreys, *Reflections on the Ethics of Social Experimentation*, 6 J. GLOBALIZATION & DEV. 87, 87–88 (2015).

161. See *supra* note 148 and accompanying text.

162. See, e.g., Ouellette, *supra* note 13, at 96–97.

163. Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 357–58, 377 (2010) (describing the “reward” of exclusive patent rights as a “dominant justificatory theor[y] of patent law” that “largely motivates current patent doctrine”).

164. See, e.g., Paul R. McDaniel & James R. Repetti, *Commentary, Horizontal and Vertical Equity: The Musgrave/Kaplow Exchange*, 1 FLA. TAX REV. 607, 621 (1993) (arguing that horizontal equity lacks “independent normative content, and that content must be supplied by reference to economic assumptions and a theory of justice”); accord James Repetti & Diane Ring, *Horizontal Equity Revisited*, 13 FLA. TAX REV. 135, 135 (2012).

165. RICHARD A. MUSGRAVE, *THE THEORY OF PUBLIC FINANCE* 160 (1959) (emphasis added).

166. Sarah B. Lawsky, *Fairly Random: On Compensating Audited Taxpayers*, 41 CONN. L. REV. 161, 163 (2008) (“[T]he tax system currently employs explicit randomness . . .”).

167. The IRS has three basic methods for selecting taxpayers to audit, two of which involve randomization. See *id.* at 164–68 (discussing these three categories). First, some taxpayers’ returns have features that virtually always merit audit, such as clearly missing income. See IRM 4.1.5.1.3.4 (Oct. 20, 2017). Such audits are not random. Second, the IRS has a highly confidential statistical methodology called the Discriminant Index Function (DIF) that scores the likelihood of an audit that increases tax revenue; the higher the DIF score, the greater the probability of being audited. IRM 4.1.2.7 (Oct. 19, 2017). Third, pursuant to the IRS’s National Research Program

Experimentation involving randomization would likely survive constitutional scrutiny. Tax classifications are subjected to rational basis review, meaning they will be upheld if “rationally related” to a “legitimate” government interest.¹⁶⁸ Increasing taxpayer utility while maintaining at least the same tax revenue is a legitimate government interest,¹⁶⁹ and randomization allows the government to pursue these interests. Indeed, the Supreme Court has expressly stated that the government has particularly broad constitutional latitude with tax law.¹⁷⁰

b. *Ex Ante* Insurance to Avoid Inequity

If the horizontal inequity of tax experimentation is still thought to present a powerful objection, the government might seek to reduce the inequity. Allowing treatment group taxpayers to choose to be in the control group or vice-versa would defeat the purpose of the experiment. But the government could seek to make it so that taxpayers will *ex ante* be indifferent to which group the taxpayer is assigned.

The government might do this by offering taxpayers randomization insurance. Suppose, for example, that equal numbers of participants (whether opt-ins, non-opt-outs, or conscripts) are to be assigned to treatment and control. The government could allow any taxpayer to pay \$1,000 to purchase a unit of insurance that would pay \$2,000 if the taxpayer is assigned to the group that the taxpayer prefers less. For example, a taxpayer who would like the treatment might purchase for \$10,000 insurance that would pay \$20,000 if the taxpayer ended up assigned to the control group. A risk-averse taxpayer should purchase enough insurance on the initially less-favored option to make the taxpayer indifferent between the options.¹⁷¹

(NRP), some taxpayers are selected entirely at random. IRM 4.22.1.3(4) (Sept. 6, 2017) (“As randomly selected returns, NRP taxpayers can represent thousands of similar taxpayers in the population.”).

168. *Regan v. Taxation with Representation of Wash.*, 461 U.S. 540, 547 (1983); *Apache Bend Apartments, Ltd. v. United States*, 964 F.2d 1556, 1561 (5th Cir. 1992), *rev’d on standing grounds en banc*, 987 F.2d 1174 (5th Cir. 1993); *see also* *Armour v. City of Indianapolis*, 566 U.S. 673, 680 (2012) (“As long as the City’s [tax] distinction has a rational basis, that distinction does not violate the Equal Protection Clause.”); *Williamson v. Lee Optical of Okla., Inc.*, 348 U.S. 483, 491 (1955) (applying the rational basis test).

169. *Colangelo v. United States*, 575 F.2d 994, 998 (1st Cir. 1978) (Coffin, C.J., dissenting) (maintaining tax collections is a legitimate government interest).

170. *Armour*, 566 U.S. at 680 (“[W]e have repeatedly pointed out that “[l]egislatures have especially broad latitude in creating classifications and distinctions in tax statutes.” (second alteration in original) (quoting *Regan*, 461 U.S. at 547)).

171. Lawrence Blume & Daniel L. Rubinfeld, *Compensation for Takings: An Economic Analysis*, 72 CAL. L. REV. 569, 601 (1984) (“If an individual is averse to risk and actuarially fair insurance can be purchased (from a risk-neutral third party insurer), then it is not difficult to show that the individual will completely insure against the risk.”).

Because assignment to a group is a product of pure chance, there is no danger of adverse selection with such insurance. An insurance payout is a lump sum, rather than an entitlement to be taxed according to the other tax regime, so there is also little danger that the insurance program will change the insureds' behavior. Meanwhile, the program could be inexpensive to administer, with insurance payments and payouts calculated on tax returns. For the government, which is effectively risk neutral, to offer actuarially fair insurance has no expected budget impact. Private parties could also offer such insurance, but the product is so simple and so connected to the government program of experimentation that it is likely easier for the government to offer the product directly. Actuarially fair government insurance should be more popular than insurance in which much of the insurance premium covers insurance company functions like rating and underwriting.

2. After-Tax Income Inequality

The above analysis suggests that the intuition that tax experimentation entails serious horizontal inequities is weak. Nonetheless, one might worry about *vertical* inequities. Tax experimentation's benefits may primarily flow to higher income taxpayers, worsening after-tax income inequality, if tax experimentation is opt-in.¹⁷² This may happen for several reasons. First, higher income taxpayers have better tax advisors or are more tax savvy (or both). They will thus be more likely to opt into alternative tax regimes that they expect to benefit them. Second, higher income taxpayers may be subject to more complicated tax rules,¹⁷³ thus creating more opportunities for alternative tax regimes that benefit them. Third, wealthier taxpayers tend to be less risk-averse.¹⁷⁴ Even if the

172. This concern about worsening inequality is related to but not the same as "vertical equity." Vertical equity is the idea that there should be an "appropriate" pattern of differentiation between those of different levels of economic income. McDaniel & Repetti, *supra* note 164, at 607; *see also* Louis Kaplow, *Horizontal Equity: Measures in Search of a Principle*, 42 NAT'L TAX J. 139, 140–41 (1989) (explaining the idea of vertical equity). "The vertical equity principle," however, "does not prescribe whether tax rates should be proportional, progressive, or regressive; nor, if progression or regression is the chosen mode, does it indicate how steep the slope should be." BITTKER & LOKKEN, *supra* note 148; *accord* McDaniel & Repetti, *supra* note 164, at 610 ("The word 'appropriate' is not self-defining . . ."); *id.* ("VE [vertical equity] could apply to a tax system that is progressive, proportional or regressive.").

173. The simplest tax return, the 1040EZ, can only be used by taxpayers with taxable income below \$100,000. *See Topic Number 352 - Which Form - 1040, 1040A, or 1040EZ?*, IRS, <https://www.irs.gov/taxtopics/tc352> [<https://perma.cc/3AMM-HY6N>] (last updated Feb. 12, 2018). The most complicated return, the 1040, includes numerous schedules for more complex tax situations. *Id.*

174. *See* KENNETH J. ARROW, *The Theory of Risk Aversion*, in *ESSAYS IN THE THEORY OF RISK-BEARING* 90, 92–93 (Julius Margolis ed., 1971); A. MITCHELL POLINSKY, *AN INTRODUCTION TO LAW AND ECONOMICS* 55 n.31 (2d ed. 1989).

alternative regime appears *ex ante* to offer a better expected outcome for the taxpayer, it may turn out to be worse if circumstances change. This danger is more likely to scare off lower-income taxpayers.

Arguably, one should see even a policy that dominantly benefits wealthy taxpayers as enhancing social welfare if other taxpayers are not harmed. Placing aside externalities,¹⁷⁵ revenue-neutral tax experimentation, unlike virtually all other conceivable tax reform,¹⁷⁶ is Pareto optimal,¹⁷⁷ and many often think Pareto improvements to be welfare-improving,¹⁷⁸ despite experimental evidence that people sometimes are willing to accept Pareto-dominated outcomes.¹⁷⁹ The argument against tax law benefits for the wealthy that do not reduce resources to others requires a modification of the Pareto criterion so that each person's welfare depends on *relative* wealth.¹⁸⁰ Revenue-neutral tax experiments from which wealthy taxpayers benefit may still be Pareto-improving, however. In the example of the entertainment deduction, it seems unlikely that the utility of non-wealthy taxpayers falls because wealthy taxpayers reallocate spending from the entertainment category. Indeed, such a change plausibly might reduce the perception of inequality, as opting-in taxpayers have lower after tax income. Thus, if one measures inequality solely through a measure like the Gini coefficient,¹⁸¹ inequality is reduced.¹⁸² In this case, experimentation might produce greater inequality of happiness, but not greater financial inequality.

175. See *infra* Section II.D.

176. See, e.g., Arthur Cockfield, *Income Taxes and Individual Liberty: A Lockean Perspective on Radical Consumption Tax Reform*, 46 S.D. L. REV. 8, 42–43 (2001) (noting that “any type of radical tax change” is likely to violate the Pareto criterion, at least in the short term).

177. A pareto improvement is one that makes some better off and no one worse off. See JOHN BLACK ET AL., OXFORD DICTIONARY OF ECONOMICS 301 (4th ed. 2012).

178. For a defense of using the Pareto principle in policy analysis, see Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961, 1015 (2001).

179. See, e.g., Dorothea K. Herreiner & Clemens Puppe, *Inequality Aversion and Efficiency with Ordinal and Cardinal Social Preferences—an Experimental Study*, 76 J. ECON. BEHAV. & ORG. 238, 252 (2010).

180. See Sven Ove Hansson, *Welfare, Justice, and Pareto Efficiency*, 7 ETHICAL THEORY & MORAL PRAC. 361, 377–79 (2004) (suggesting such an adjustment); see also Khandakar Qudrat-I Elahi, *Economic Inequality and Paretian Welfare Economics: Some Insinuating Questions*, 35 F. SOC. ECON. 19, 28–32 (2005) (critiquing the Pareto criterion given concerns about inequality).

181. See generally Corrado Gini, *Concentration and Dependency Ratios*, 87 RIVISTA DI POLITICA ECONOMICA 769 (1997) (providing English translation of Gini's 1909 article in Italian).

182. The Gini coefficient is generally reported as a function of household income. See, e.g., JESSICA L. SEMEGA ET AL., U.S. CENSUS BUREAU, INCOME AND POVERTY IN THE UNITED STATES: 2016, at 8 (2017), <https://www.census.gov/content/dam/Census/library/publications/2017/demo/P60-259.pdf> [<https://perma.cc/S4LW-QACL>]. Household income would not include deductible business expenses.

Suppose, however, that tax experimentation might increase inequality under some definition to sufficiently outweigh any efficiency benefits. At least in principle, any benefit accruing to high-income taxpayers could be reallocated in part to low-income taxpayers. That is, if high-income taxpayers expect to receive a utility benefit equal to \$1,000 from some sort of tax experimentation, then the law creating such experimentation could impose a new tax on higher income taxpayers and redistribute the receipts to lower income taxpayers. This argument is frequently made on behalf of programs that increase economic efficiency but may have negative redistributive consequences,¹⁸³ and it leads to the retort that the mere possibility of redistribution does not justify a policy if the law creating the efficient policy does not affect redistribution.¹⁸⁴ In principle, however, a hypothetical statute that authorized tax experimentation in conjunction with other progressive tax changes could answer the distribution objection.

C. Manipulation

Taxpayers may attempt to exploit the availability of tax experiments¹⁸⁵ in several ways. First, taxpayers who know about an experiment may change their behavior to become eligible for invitation into the alternative regime. Relatedly, taxpayers may change their behavior or their reporting so that they are matched to taxpayers in the control group likely to pay low taxes. Second, taxpayers subject to an alternative tax regime may minimize their taxes by shifting gross income and deductions between the tax years subject to the alternative regime and tax years not subject to the alternative regime. Third, taxpayers may change their tax status. This Article will consider each of these in turn.

1. Eligibility and Matching

Some taxpayers always attempt to game the tax laws to their advantage, and the same will doubtless be true of eligibility for tax experiments. Tax experiments might be limited to particular taxpayers, such as members of a particular industry. For example, the government

183. See, e.g., Louis Kaplow & Steven Shavell, *Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income*, 29 J. LEGAL STUD. 821, 822 (2000).

184. See Lee Anne Fennell & Richard H. McAdams, *The Distributive Deficit in Law and Economics*, 100 MINN. L. REV. 1051, 1082–83 (2016) (arguing that just as transactions costs can prevent achievement of efficiency goals, so too can political action costs prevent achievement of distributive goals).

185. See David A. Weisbach, *Formalism in the Tax Law*, 66 U. CHI. L. REV. 860, 869 (1999) (“Uncommon transactions that are taxed inappropriately become common as taxpayers discover how to take advantage of them.”).

might consider limiting eligibility for an experiment on eliminating deductibility of travel expenses to an industry in which such expenses generally seem unnecessary, such as health professionals who currently can deduct trips to conferences at fancy vacation destinations.¹⁸⁶ This is, however, an incomplete solution. A taxpayer who does not expect to need to take significant travel deductions anyway might seek to classify as a health professional, even if the taxpayer's business is only marginally related to health.

A taxpayer might have similar manipulation incentives even in an experiment for which the taxpayer is clearly eligible or for which there are no eligibility limitations. Recall that after a treatment group taxpayer calculates tax liability under the new rules, this amount is multiplied by a number chosen to ensure revenue neutrality. That multiplier is based on the quotient of the tax liability one would expect the taxpayer to have based on the taxpayer's past data if assigned to the control group, divided by the nominal tax liability expected if assigned to the treatment group.¹⁸⁷ If the government uses a multivariate regression, the taxpayer hopes to have characteristics that lead to a prediction that the taxpayer will have much higher nominal liability in the treatment group. A taxpayer planning ahead may be able to generate data in a year before the experiment that will lead to such predictions.

The taxpayer's manipulation incentives will be the opposite in a tax experiment assessing the efficiency of a new deduction. Then, the taxpayer would like to be matched to other participants whose behavior will change the least. For example, if the experiment considers the possibility of a new deduction for certain commuting expenses, then the taxpayer would like to be matched with taxpayers who are likely to receive only a small benefit from the availability of the commuting deduction. A taxpayer who can claim a large amount for an experimental new deduction will be able to keep most of this benefit, if the regression predicts that the taxpayer's nominal taxable income would be similar whether the taxpayer is in the control or treatment group. In an extreme case, a taxpayer might move very close to work in the year before the experiment and then move much further away in the experimental year.

The tax authority might take several steps to limit the success of such manipulation. First, the government might define the relevant tax provision narrowly to reduce such gaming. The travel experiment might affect only deductions for travel to health-related conferences, and the commuting deduction might be eligible only for taxpayers who live a certain distance from work. Second, the government can keep

186. Cf. DR.'S REV., <http://www.doctorsreview.com> [<https://perma.cc/X62R-ACQB>] (allowing search for medical conferences based on desired destination).

187. See *supra* Section I.A.2.

manipulable eligibility criteria and the variables used to match taxpayers confidential. The IRS already does something similar, keeping confidential the criteria that affect the probability that the IRS will audit a taxpayer.¹⁸⁸ On the other hand, there may be value in publishing such data to allow the public to better inform itself about whether an experiment was successful. Third, the government can use older data that predate the announcement of the experiment (not just data that predate the experimental year) to filter out opportunistic taxpayers. Fourth, the government might collect and then heavily weight data not easily manipulated by taxpayers, such as the college or graduate degrees received by the taxpayer, or the industry of past employers.

Ultimately, some manipulation is still likely to occur on the margins. The problem, however, should not be overstated. Tax law already presents many opportunities for fraud, and criminal and civil liability may deter many taxpayers from fraudulently filling out their tax returns in the hope of receiving a deduction.¹⁸⁹ If, nonetheless, many taxpayers seek to define themselves in a way that they anticipate will lead to better treatment, the strategy will be self-defeating. For example, if many non-health professionals classified themselves as health professionals, then the non-health professionals would be matched with many other non-health professionals. The problem is then reduced to the danger that real health professionals would not want to participate in the experiment for fear of being matched with non-health professionals. But the experiment failing for this reason would simply indicate that the underlying hypothetical tax provision (elimination of a deduction but only for health professionals) is unworkable because the eligibility criteria are easily manipulated. This itself is a valuable lesson for a tax experiment.

2. Income and Deduction Shifting

Taxpayers who are subject to an alternative tax treatment may try to minimize their taxes by shifting income and deductions from years in which they are subject to the treatment to years when the generally applicable tax rules apply to them. Or, taxpayers may try to shift gross income and deductions in the opposite direction, to years covered by the alternative tax regime. For example, if the alternative tax regime were a lower marginal rate of taxation in exchange for losing a deduction, a taxpayer may shift as much income as possible into years covered by the alternative regime. The taxpayer might do this by negotiating with their employer to forgo a bonus in a year subject to generally applicable tax

188. IRM 4.1.2.7 (Oct. 19, 2017) (“DIF mathematical formulas are confidential and for official use only. The DIF score assigned to a return should not be disclosed.”).

189. If the federal government has insufficient resources to prosecute tax fraud, *qui tam* suits could enable greater enforcement. See Franziska Hertel, Note, *Qui Tam for Tax?: Lessons from the States*, 113 COLUM. L. REV. 1897, 1922 (2013).

law with the understanding that the taxpayer would receive a larger bonus the next year, when the alternative regime would apply. Meanwhile, in the opposite type of experiment, where the taxpayer expects to pay a higher tax rate, the taxpayer could seek to shift income to a later year.

Experiments can address these problems. The alternative tax regime could apply for multiple years to minimize the potential for short-term mischief. Taxpayers might be informed of their group assignment on January 1 to prevent any manipulation in advance of the experiment. For each year of the experiment, a separate model of the behavior of control group and treatment group taxpayers would be developed so the multiplier a taxpayer receives would vary from year to year. The government would assess the experiment on a year-to-year basis, and it might focus analysis especially on taxpayers entering the first year of the experiment, since income shifting is likely to be most plausible near the end of an experiment. In principle, the government could even design an experiment to be permanent. A taxpayer giving up a deduction would be giving that deduction up permanently, and each year would receive some discount in exchange. If the tax law changed so that no taxpayers received this deduction anymore, then the multiplier would end up being closer to one.

A less drastic approach is for the separate tax treatment to apply only during the experimental period (perhaps even just a year), but for a multiplier to be calculated for each taxpayer for each subsequent year. This addresses the concern that even if an experiment is revenue-neutral during the period of the experiment, it might not be revenue-neutral afterward. So long as the multiplier approach is used in all subsequent years,¹⁹⁰ the experiment is guaranteed to be revenue neutral over taxpayers' entire lives. Thus, if treatment taxpayers were shifting income into the years of the experiment, they would receive large discounts during this time, but they might report less nominal tax liability after the experiment ends and therefore have to pay higher tax rates then.

The length of the experiment aside, tax law includes tools to discourage gaming. Auditors should rigorously apply the existing tax-law doctrines that prevent shifting gross income and deductions between years, such as the constructive receipt doctrine¹⁹¹ and the economic

190. Eventually, too few taxpayers might remain living to enable meaningful comparison of the treatment and control groups. At this point, it would likely be appropriate to use data from previous years.

191. This doctrine prevents cash method taxpayers (which are the vast majority of individual taxpayers) from postponing the reporting of gross income by failing to exercise the power to collect it. BITTKER & LOKKEN, *supra* note 148, ¶ 5.9.

performance doctrine.¹⁹² Meanwhile, the government might prioritize experimenting with alternative tax regimes that limit the scope of income-shifting. Capital gains, for example, might be particularly easy to shift,¹⁹³ and so the government might prioritize tax experiments that apply to only ordinary income (that is, income that is not capital gains).¹⁹⁴

3. Status Changes

Individuals can get married or divorced; corporations can merge with others or divide into multiple corporations (for example, by spinning off a business into a new corporation). If a taxpayer is subject to an alternative tax regime, how would such status changes apply? The rules governing such status changes must be designed with care to prevent opportunistic, inefficient behavior. For example, suppose that company *T* is a treatment group taxpayer subject to an alternative tax regime, while company *A* is not. Suppose *A* is in a position so that the alternative regime would reduce its taxes because it would benefit from the experiment more than most members of the treatment group. Suppose further that aside from tax considerations, it makes no economic sense for *A* to acquire *T*. If *A* could become subject to the alternative regime by merging with *T*, the result would be an economically inefficient merger, plus a tax windfall to *A*.

Further, status changes can complicate the goal of achieving revenue neutrality. The multiplier that would be applied to *T*'s nominal tax liability depends on *T*'s pre-experiment tax returns. But if *T* is suddenly much larger after acquiring *A* and is paying far greater taxes, that would have the effect of lowering the taxes of all treatment group taxpayers. The average multiplier will be equal to the total tax liability of control group taxpayers divided by the total nominal tax liability of treatment group taxpayers, so increasing the denominator reduces the multiplier. This would particularly be a problem if treatment group taxpayers are systematically more likely to acquire other companies than control group taxpayers. The reverse tendency would also create problems of evaluation.

192. This doctrine prevents accrual method taxpayers from taking a deduction before they have provided economic performance, such as delivering goods or services to the buyer. I.R.C. § 461(h) (West 2018). *See generally* 1 STEPHEN GERTZMAN, FEDERAL TAX ACCOUNTING ¶ 4.04[3] (2d ed. 1993 & Supp. Sept. 2018) (discussing economic performance rules).

193. Joseph J. Cordes & Harvey Galper, *Tax Shelter Activity: Lessons from Twenty Years of Evidence*, 38 NAT'L TAX J. 305, 322 (1985) (describing that when capital gains rates are low, investments often shift to activities generating capital gains).

194. Many other countries' tax systems have very different treatment for income from capital and income from labor. *See, e.g.*, Klaus Sieker, *Business Operations in Germany*, B.N.A. TAX MGMT. PORTFOLIO NO. 7140 ¶ IV.B.1 (2017) (discussing Germany's special treatment of capital income like dividends and interest); *cf.* I.R.C. § 64 (2012) (defining "ordinary income").

But tax law already deals with similar problems of status changes.¹⁹⁵ Individuals' divorces and corporations' divisions are easy. If a married couple is subject to an alternative tax regime and they get divorced, then both ex-spouses should remain subject to the alternative regime. If a corporation is subject to an alternative regime and divides, such as by spinning off a business into a new corporation,¹⁹⁶ then both the original corporation and the spun-off company will remain subject to the alternative regime.¹⁹⁷ The model used to determine multipliers would be generated based on the combined nominal tax of any spouses or pair of companies that split up. Then, the multiplier would apply equally to each spouse paying separately or to each entity.

Marriages and especially corporate mergers are more complicated. To prevent marriages from affecting tax liability, the most straightforward approach is to require all participants in an experiment to file individually, whether as a single individual or as married filing separately.¹⁹⁸ This would be true even for control group taxpayers, to facilitate the comparison between the control group and the treatment group. The drawback of this approach is that eliminating the option to file a joint return will be disadvantageous to some taxpayers. Although some taxpayers suffer a "marriage penalty,"¹⁹⁹ others enjoy a "marriage bonus,"²⁰⁰ and under current U.S. tax law, people generally see the married-filing-separately status as disadvantaged relative to the others,²⁰¹

195. See, e.g., I.R.C. § 382 (2012 & Supp. V 2018) (dealing with corporate tax attributes like Net Operating Losses in the context of corporate acquisitions).

196. E.g., *id.* § 355 (providing tax rules for corporate separations like spin-offs, split-offs, and split-ups).

197. Cf. 1 BORIS BITTKER & JAMES EUSTICE, *FEDERAL INCOME TAXATION OF CORPORATIONS AND SHAREHOLDERS* ¶ 11.12 (7th ed. 2000 & Supp. III 2018) (discussing how, in a corporate separation under I.R.C. § 355, certain tax attributes are continued).

198. Some circumstances already exist where one spouse's unusual tax situation bars a married couple from filing joint returns, including one spouse using the calendar year as a taxable year and the other spouse using a different taxable year, I.R.C. § 6013(a)(2) (2012), and one spouse being a nonresident alien, *id.* § 6013(a)(1). See generally BITTKER & LOKKEN, *supra* note 148, ¶ 111.5.2 (discussing married-filing-jointly returns versus married-filing-separately returns in depth). Moreover, filing separately sometimes allows taxpayers to get certain tax benefits that would not be available (or not as available) if filing jointly. See *id.* (discussing how filing separately can maximize the deduction under I.R.C. § 213(a)).

199. See, e.g., Margaret Ryznar, *A Practical Solution to the Marriage Penalty*, 44 PEPP. L. REV. 647, 653–58 (2017) (providing statistics and history).

200. For a detailed statistical analysis of who receives a penalty and who receives a bonus under current tax law, see Amir El-Sibaie, *Marriage Penalties and Bonuses Under the Tax Cuts and Jobs Act*, TAX FOUND. (Feb. 14, 2018), <https://taxfoundation.org/tax-cuts-and-jobs-act-marriage-penalty-marriage-bonus/> [<https://perma.cc/QRZ7-JV54>].

201. See, e.g., Ryznar, *supra* note 199, at 655 ("Filing separately is, on average, not advantageous because fewer credits and deductions are available, and the tax brackets are narrower than those of single filers.").

with some exceptions.²⁰² Thus, taxpayers might elect to participate in tax experiments only if they are willing to forego this option. The best long-term solution might be for all taxpayers to file separately.²⁰³ While this would be a significant change in U.S. law, most developed countries have always had individual filing or have made this switch in recent decades,²⁰⁴ and there are strong arguments that the United States should switch as well.²⁰⁵ Of course, the United States might use a tax experiment to determine whether to eliminate this option. Meanwhile, there is no impediment to tax experimentation in countries with individual filing.

In corporate mergers,²⁰⁶ where the target company is subject to an alternative regime but the acquiring company is not, there are two possible options. First, the government may require the target to maintain separate corporate form as the acquirer's subsidiary²⁰⁷ and continue to file its own separate tax return using the alternative tax regime, without the option of consolidating its returns with the acquirer's.²⁰⁸ So long as the acquirer and the target remain legally separate entities, the accounting is straightforward. A complication here is that the combined entities might engage in transactions seeking to take advantage of the differential tax treatment of the two entities. This problem is familiar to tax law,²⁰⁹

202. See, e.g., Laura Saunders, *When 'Married, Filing Separately' Lowers Your Tax Bill*, WALL STREET J. (Feb. 23, 2018, 5:30 AM), <https://www.wsj.com/articles/when-married-filing-separately-lowers-your-tax-bill-1519381801> [<https://perma.cc/PAC8-GPPJ>].

203. But see Yair Listokin, *Taxation and Marriage: A Reappraisal*, 67 TAX L. REV. 185, 196–201 (2014) (suggesting more modest changes to tax law).

204. Fox, *supra* note 7, at 9–10.

205. See, e.g., *id.* at 48–49 (arguing that the current regime distorts couple's marriage decisions).

206. E.g., I.R.C. § 368(a)(1)–(2) (2012) (defining the various types of reorganizations available for the tax-free merger of two corporations).

207. There are a number of methods of tax-free acquisition already available that would allow the target and acquirer to remain separate corporate forms, including “B” reorganizations under I.R.C. § 368(a)(1)(B), forward triangular mergers under I.R.C. § 368(a)(2)(D), and reverse triangular mergers under I.R.C. § 368(a)(2)(E).

208. See generally BITTKER & LOKKEN, *supra* note 148, ¶97.2 (describing consolidated returns). The government has long limited the eligibility of subsidiaries to consolidate their returns with their parent's corporate group, see I.R.C. § 1504(b), and subsidiaries (e.g., Target in the example above) subject to alternative tax regimes would simply be a new limitation. Non-consolidation of returns would not lead to taxation of the subsidiary's income at both the subsidiary and parent corporation level, thanks to the dividends received deduction at I.R.C. § 243(a)(3).

209. Transfer pricing also may become an issue as the IRS seeks to limit the scope of pass-through income rules. See Kamin et al., *supra* note 2, at 22–23 (“[T]he IRS has had only very limited success in preventing these kinds of transfer-pricing and other valuation games among related parties in other contexts.”).

with transfer pricing regulation seeking to ensure that transactions reflect what parties would agree to in arms-length negotiations.²¹⁰

Second, the government may allow the target to fully merge and consolidate with the acquirer, but the combined corporation would have to calculate its taxes once using the alternative regime and a second time using the default regime. The actual tax liability would be a weighted average of the two, in proportion to the relative value of the target and acquirer at the time of the merger.²¹¹ The government would then use the proportion of nominal tax liability attributable to the treatment group entity in modeling the effects of the experiment more broadly. The principal challenge with this approach is valuation. If the target and acquirer are public corporations, then the government can use their stocks' market capitalization to determine valuation.²¹² In other circumstances, the relative valuations might be a matter of dispute between the tax authority and the taxpayer, as the taxpayer might seek a larger or smaller capitalization based on which tax treatment seem more attractive. The costs of such gaming are likely to be low, however, given that the revenue-neutrality constraint is unlikely to make either option much more attractive than the other. Nonetheless, the transactions costs associated with arguing about valuation might discourage corporate participation in experiments on relatively minor tax code provisions.

D. Externalities

Many tax provisions, such as business expense deductions,²¹³ exist to properly calculate taxpayer income.²¹⁴ But other tax provisions are “tax expenditures” that further some additional goal,²¹⁵ such as creating

210. See, e.g., I.R.C. § 482 (Supp. V 2018).

211. I.R.C. § 382(b)(1) uses an analogous regime to limit the use of tax attributes (like net operating losses and built-in losses) of an acquired corporation in proportion to the value of the acquired corporation. See generally 2 BITTKER & LOKKEN, *supra* note 148, ¶ 14.42[3] (explaining functioning of § 382's limitations).

212. Treas. Reg. § 1.382-2(a)(3)(i) (as amended in 1999) (providing that in calculating “value,” control premiums or similar considerations are ignored).

213. E.g., I.R.C. § 162 (2012 & Supp. V 2018).

214. 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) (distinguishing “the structural provisions necessary to the application of a normal income tax, such as the definition of net income” from “special preferences found in every income tax”).

215. 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, 136–38 (2010). But see Borris I. Bittker, *Accounting for Federal “Tax Subsidies” in the National Budget*, 22 NAT'L TAX J. 244, 247 (1969) (criticizing the concept of tax expenditures); Douglas A. Kahn & Jeffrey S. Lehman, *Commentary, Tax Expenditure Budgets: A Critical View*, 54 TAX NOTES 1661, 1661 (1992) (criticizing the same).

positive externalities or reducing negative externalities.²¹⁶ A potential objection to tax tailoring is that allowing taxpayers to opt out of such externality-addressing tax expenditures will adversely affect third parties. For example, the R&D tax credit²¹⁷ is often defended as encouraging scientific research that will benefit society as a whole²¹⁸ because companies cannot appropriate the full value of their research efforts.²¹⁹ Similarly, tax expenditures subsidizing higher education²²⁰ are justified by its positive externalities.²²¹ Other examples include the various credits,²²² deductions,²²³ and exclusions²²⁴ subsidizing energy efficiency and clean energy, all of which are justified as reducing the negative externalities from fossil fuels.²²⁵

One could make an argument for preventing alternative tax regimes from removing such tax expenditures. Similarly, one could justify the absence of some deductions on externality grounds. For example, environmental considerations might justify the absence of a deduction or exclusion for the cost of gas used commuting by car.²²⁶ Allowing

216. See *infra* notes 218–25 and accompanying text; see also Martin Feldstein, *A Contribution to the Theory of Tax Expenditures: The Case of Charitable Giving*, in THE ECONOMICS OF TAXATION 99, 99 (Henry J. Aaron & Michael J. Boskin eds., 1980) (“[T]ax incentives should be used to encourage charitable gifts that will promote the achievement of desired social goals.”).

217. I.R.C. § 41 (2012 & Supp. V 2018); see also *id.* § 174 (allowing immediate deduction for R&D expenditures, in contravention of the general principle of capitalization for expenditures creating multi-year benefits).

218. See, e.g., Martin A. Sullivan, *Economic Analysis: Can a Patent Box Promote Advanced Manufacturing?*, 147 TAX NOTES 1347, 1348 (2015) (“[D]ecades of research by leading economists indicates that externalities from R&D not only exist but are very large.”).

219. See CONG. RESEARCH SERV., 112TH CONG., TAX EXPENDITURES: COMPENDIUM OF BACKGROUND MATERIAL ON INDIVIDUAL PROVISIONS 104 (Comm. Print 2012) (“[B]usinesses in general are unlikely to invest in R&D in amounts consistent with its social returns.”).

220. E.g., I.R.C. § 25A(i) (2012 & Supp. V 2018); *id.* § 117 (scholarships); *id.* § 221 (deduction on student-loan interest); *id.* § 222 (deduction for qualified tuition); *id.* § 529 (education savings accounts); I.R.C. § 127 (2012) (exclusion for employer-provided educational assistance).

221. See CONG. RESEARCH SERV., *supra* note 219, at 633.

222. I.R.C. § 45 (2012 & Supp. V 2018) (allowing credit for renewable energy generation, as from wind).

223. *Id.* § 179D (allowing deduction for all or part of the cost of energy-efficient commercial building property).

224. I.R.C. § 136 (2012) (allowing exclusion for energy conservation subsidies provided by utilities).

225. See CONG. RESEARCH SERV., *supra* note 219, at 113–53.

226. See Roberta F. Mann, *On the Road Again: How Tax Policy Drives Transportation Choice*, 24 VA. TAX REV. 587, 635–40 (2005) (discussing how the existing tax benefit for employer-provided parking creates externalities, including more car-generated pollution and congestion); *Comm’r v. Flowers*, 326 U.S. 465, 472–73 (1946) (denying deduction for commuting

taxpayers to opt into such a deduction in exchange for higher tax rates might reduce social welfare. Thus, the tax authority should hesitate to offer a tax experiment that would eliminate for treatment taxpayers a tax provision justified (or justifiable) based on its effects on others.

There may be some situations, however, in which concerns about externalities could be addressed by changing the *unit* offered the alternative tax regime. Consider, for example, the home-mortgage interest deduction.²²⁷ The deduction's defenders argue that it encourages homeownership, producing positive externalities including better maintained neighborhoods and greater social capital in neighborhoods.²²⁸ Allowing opting into an alternative tax regime without the home-mortgage interest deduction (in exchange for lower rates) might then have negative effects on neighborhoods.

A response to this problem would be to allow entire neighborhoods (or entire collections of neighborhoods) to opt into such alternative regimes. Specifically, the decisions to opt in could be made by a homeowners' association (HOA) or a municipality. This election would "run with the land," so that home mortgage interest could never be deducted for a home within that HOA or municipality, but any taxpayer principally residing in a home would benefit from the lower rates. The opt-in is thus "in rem," attaching to the property (that is, the homes), regardless of whether it is subsequently transferred. Allowing the opt-in to be done by the HOA or the municipality ensures that externalities are internalized at the level making the decision to opt in. This would, of course, require enabling legislation, above and beyond that needed for tax experimentation in general.

Random assignment of some HOAs or municipalities that opt into a control group would allow rigorous determination of whether the home mortgage interest deduction produces the predicted neighborhood benefits. In theory, the deduction should lead to more investment in housing, so assuming the experiment includes at least some not-yet-developed land, one should expect that such land would be more likely to be developed than corresponding land in the control group. Property values similarly ought to be higher in areas with the deduction than without, as those in the treatment group allocate more of their money to non-house spending.²²⁹ Of course, if the primary goal of an experiment

expenses based on predecessor to I.R.C. § 262, which denies a deduction for personal, family, and living expenses).

227. I.R.C. § 163(h)(2)(D).

228. See KNIGHTLY, *supra* note 53 (reviewing literature on this matter). See generally William M. Rohe et al., *The Social Benefits and Costs of Homeownership*, in *LOW-INCOME HOMEOWNERSHIP* 381 (Nicolas P. Retsinas & Eric S. Belsky eds., 2002) (reviewing literature on the social impacts of homeownership).

229. For a general argument that laws affect housing prices, see Anup Malani, *Valuing Laws as Local Amenities*, 121 *HARV. L. REV.* 1273, 1274 (2008).

were informational, then a home mortgage interest deduction experiment with individual subjects might be more effective. If the deduction truly has externalities, then one should expect houses near those randomized to the treatment of giving up the deduction to decline in value relative to houses near those randomized to control.

The government also might be able to experiment with tax provisions thought to provide positive externalities by defining a new treatment group that is designed to promote externalities but in a different way. For example, the government might consider replacing the R&D tax credit with a system in which the government directly grants subsidies in the form of tax credits to companies based on the quality of their research proposals or the importance of research undertaken.²³⁰ Such an experiment could still be revenue neutral. Taxpayers might be willing to opt in, if they think they would benefit from greater tax benefits for more important research, with reduced tax benefits for less important research. The government might then seek to evaluate both whether the change is attractive to taxpayers and whether the change indeed produces more valuable research. As with assessing externalities from the home-mortgage interest deduction, this analysis would require data beyond that available from tax returns.

III. BEYOND REVENUE NEUTRALITY

This Article argues for revenue-neutral experimentation in part because such experimentation entails little risk to taxpayers. It is possible, however, to imagine more adventuresome experiments that would not necessarily be revenue-neutral. Section III.A discusses the possibility that participation might be subsidized, with participants on average paying less tax than others. Section III.B suggests the possibility of non-revenue-neutral experiments on tax rates to help assess the impact of taxation on taxpayer behavior. The results of these or other experiments might automatically lead to permanent changes in law, as discussed in Section III.C's analysis of self-executing experiments. Finally, Section III.D assesses the possibility that tax experiment subjects might be states, counties, or localities instead of individuals.

A. *Participation Inducements*

Perhaps the most significant challenge to tax experimentation is that few taxpayers may agree to participate. Opting in to an experiment requires some research into the tax law issue. Taxpayers also need to assess the possibility of adverse selection. Once again, if the taxpayers

230. See generally Daniel J. Hemel & Lisa Larrimore Ouellette, *Beyond the Patents-Prizes Debate*, 92 TEX. L. REV. 303 (2013) (discussing the varieties of incentives governments can provide for innovation, beyond just R&D tax credits).

most likely to participate are those with private information that they would gain greatly from a tax law change, then taxpayers who expect to benefit only slightly with respect to their nominal tax liability should be expected to lose money once revenue-neutral adjustments apply. Above,²³¹ this Article discussed one solution to adverse selection that makes taxpayers hesitant to participate: designating an experiment opt-out or even mandatory. But it also recognized strong arguments against such an approach to these strategies.

An alternative possibility would be for the government to offer inducements to taxpayers willing to participate. All taxpayers who opt into the experiment, or perhaps only those assigned to treatment, might receive some tax discount (perhaps a small percentage discount on their tax returns). Even relatively modest inducements might be sufficient to overcome adverse selection problems. With such participation inducements, tax experimentation would not be truly revenue neutral; treatment group taxpayers would be promised that they would pay less on average. The government would be sacrificing revenue for the information to be obtained from a tax experiment. But taxpayers in the treatment group would still be matched to control group taxpayers. This has the useful effect of ensuring that taxpayers will be willing to participate only if they believe that the tax changes will not overly distort their own incentives.

The government might offer inducements for particular experiments, but it also might offer inducements to taxpayers willing to participate in a wide range of tax experiments. The government might simultaneously execute a wide range of experiments on various tax issues. This can help limit adverse selection. Because taxpayers will anticipate that the chances of being randomized into the treatment group of any particular experiment will be relatively low, participation is not likely to be limited to taxpayers who expect to benefit disproportionately from that tax experiment.

B. *Experiments on Tax Rates*

The government can also use participation inducements to facilitate experiments on tax rates. Consider, for example, a tax experiment in which some taxpayers will be randomly selected to pay an extra 2% in taxes, while others receive a 3% discount on their taxes. The purpose of such an experiment would be to assess the effect of the level of taxation on other variables, such as taxpayer work effort and job creation. These are among the most contentious issues in tax policy,²³² but randomized

231. See *supra* Section II.A.

232. See generally Saez et al., *supra* note 18 (providing a review of the literature).

studies have been unavailable²³³ and interpretive issues make it virtually impossible to generalize from existing evidence about long-term effects of marginal tax rate changes.²³⁴ An important question for policymakers is whether higher tax rates reduce taxpayers' willingness to work—that is, whether the income or substitution effects of taxes dominate.²³⁵

Such experiments might produce greater objections than revenue-neutral experimentation, because the horizontal inequities resulting are greater. This is indeed a principal reason that this Article has focused on revenue-neutral experimentation. Nonetheless, the possibility of such experiments should not be dismissed, because they would produce a form of data otherwise almost impossible to obtain. Moreover, this Article has shown above that concerns about horizontal inequity may be overstated. Taxpayers might purchase insurance against being randomized to the group that receives a tax increase,²³⁶ with such insurance paying an *ex ante* lump sum to such taxpayers. That would limit the experiments' usefulness in providing insight into the effects of overall wealth on individual behavior, but such experiments could still help identify marginal incentive effects of government policy.

With revenue-neutral tax experiments, one can often consider an experiment successful simply on the basis that informed taxpayers would prefer the treatment group to the control group. With non-revenue-neutral experiments, the mere existence of participating taxpayers cannot be a strong argument in favor of the policy. Still, the government might measure success based on the effects of the experiment. For example, if an experiment demonstrated that taxpayers who suffered an increase in the tax rate did not reduce work effort as measured by pretax income, or even increased work effort, that would furnish an argument in favor of increasing the tax rate. On the other hand, if an increase in the tax rate in fact lowered tax revenues, that would indicate that the government was

233. *Id.* at 23 (“Although we refer in this section to income tax rate schedule changes as a treatment, they certainly do not represent a classical treatment in which a random selection of taxpayers is presented with a changed tax rate schedule, while a control group of taxpayers is not so subject.”).

234. *Id.* at 43 (“Estimates of the elasticity of taxable income in the long run (i.e., exceeding a few years) are plagued by extremely difficult issues of identification, so difficult that we believe that there are no convincing estimates of the long-run elasticity of reported taxable income to changes in the marginal tax rate.”).

235. The substitution effect suggests that workers will work harder with lower taxes, substituting labor for leisure, while the income effect suggests that lower taxes may cause workers to engage in more leisure because they have less need to work. *See, e.g.,* LIBBY RITTENBERG & TIMOTHY TREGARTHEN, *PRINCIPLES OF MICROECONOMICS* 309–13 (2009). If only marginal tax rates change, then there should be no income effect, but virtually any tax reform will affect both inframarginal and marginal tax rates.

236. *See supra* Section II.B.1.b.

already on the wrong side of the Laffer Curve²³⁷ and that a tax increase would not be advisable. Between these extremes, analysts might reach different normative conclusions.

C. Self-Executing Tax Experiments

How should the tax authority decide whether to expand an experiment when it shows some effects that would be generally viewed as beneficial and some that would be generally viewed as harmful? And at what point should the legislature make a tax provision permanent? These are fundamentally political trade-offs that experimentation can produce information about, but public officials ultimately must decide how to act based on such information. But a legislature authorizing tax experimentation could specify *ex ante* what the consequences of such an experiment might be. For example, a legislature authorizing an experiment on a tax increase might provide that the tax increase will be applied to all taxpayers so long as the experiment establishes that a 1% increase in tax leads to no more than a 0.5% decrease in work effort.

Such self-executing experiments²³⁸ can facilitate legislative transparency and honesty. During the debate on the recent tax reform, brief consideration was given to the possibility of a trigger that would automatically increase tax rates if revenues fell short of projections, but the Senate parliamentarian rejected the plan.²³⁹ The virtue of such a trigger is that it punishes cheap talk. A legislator endorsing a tax cut can make more credible the legislator's claimed revenue projections by agreeing to a trigger. But there are problems with such a trigger as well.²⁴⁰ Revenue may fall short of projections for exogenous reasons, such as a recession. Moreover, increasing taxes would generally be a poor fiscal policy response to a recession. Taxes generally fall during recessions, and this "automatic stabilizer" may reduce the extent of the economic

237. See Arthur B. Laffer, *The Laffer Curve: Past, Present, and Future*, HERITAGE FOUND. (June 1, 2004), <https://www.heritage.org/taxes/report/the-laffer-curve-past-present-and-future> [<https://perma.cc/FT4H-3THC>] (discussing the curve, which postulates that tax revenues will be zero at both 0% and 100% tax rates).

238. See Abramowicz et al., *supra* note 9, at 985–87 (discussing the possibility of self-execution).

239. See *Senate Republicans Scramble to Find Revenue for Tax Bill with Vote Expected Friday*, N.Y. TIMES (Nov. 30, 2017), <https://www.nytimes.com/2017/11/30/us/politics/tax-overhaul-senate-debate.html> [<https://perma.cc/TXM4-GWHS>].

240. Some of these problems are political. See, e.g., Stephen K. Cooper, *Senate Leaders Agree to Add Fiscal Trigger to Tax Reform Bill*, TAX NOTES TODAY (Nov. 29, 2017), <https://www.taxnotes.com/tax-notes-today/tax-reform/senate-leaders-agree-add-fiscal-trigger-tax-reform-bill/2017/11/29/1xcfm> [<https://perma.cc/E8PN-UCXU>] (reporting Sen. Kennedy's reaction to the prospect of automatic tax increases).

downturn.²⁴¹ An experiment, however, can provide a better measure of a policy's effects, because the treatment group is compared to the control group. Thus, if legislators favor a tax cut because they believe that it will largely be self-financing, it might make sense for the tax cut initially to be experimental, with the result of the experiment determining whether they should make the tax cut permanent.

A reasonable objection is that it will be difficult to craft—in advance—a measure that fully captures a tax experiment's effects. Legislators might hope that a tax cut not only will lead to more work effort on the part of the recipients of the tax cut, but also greater job creation by these individuals, who might be more likely to start or expand businesses. Yet it will not always be straightforward to attribute job gains and losses to particular taxpayers. The self-execution outcome might thus depend on easily measured variables that may be proxies for broader economic effects.

One form of a self-executing tax experiment might allow for tax law to move in opposite directions depending on the result, for example with a failed tax cut experiment leading to an automatic tax increase. Suppose that legislators disagree about the effects of increasing taxes on the wealthy, with more conservative legislators worrying that this will lead the wealthy to work less hard and more liberal legislators taking the opposite position. This is an empirical issue, and it is an empirical issue to which an experiment is responsive. But many legislators seem unlikely to change their positions even if an experiment provides powerful evidence, but self-execution might force a policy result. Such a self-executing experiment might be easier to pass than an experiment that can only move the law in one direction. If legislators on each side truly believe that the experiment will vindicate their respective empirical positions, then both sides should be eager to support such a self-executing experiment. Meanwhile, a refusal to agree to such a self-executing experiment may expose legislators who claim that a tax law change would lead to a particular result but do not in fact believe this to be true.

D. Larger Experimental Units

Non-revenue-neutral experiments, whether self-executing or not, will produce information about how individual taxpayers respond to changes in tax law, such as increases or decreases in tax rates. But often, the relevant question will be how these changes affect others. This Article noted above that people might be interested in whether businesses affected by a tax law change increase or decrease their hiring levels.

241. See generally Alan J. Auerbach & Daniel Feenberg, Symposium, *The Significance of Federal Taxes as Automatic Stabilizers*, 14 J. ECON. PERSP., Summer 2000, at 21, 37 (discussing automatic stabilizers in fiscal policy).

Moreover, even when they are interested in taxpayer responses, that information might not be clear from the taxpayers' tax returns. For example, an experiment might impose a tax penalty on taxpayers who fail to purchase health insurance. The Affordable Care Act included such an "individual mandate,"²⁴² which the Supreme Court upheld as a tax,²⁴³ but this individual mandate was removed in the most recent tax reform.²⁴⁴ Tax returns might tell whether a penalty leads more taxpayers to purchase insurance, but it cannot provide information about how it affects these taxpayers' health.

A possible solution would be for tax experiments to occur at levels greater than the individual taxpayer. This Article has already discussed one example of this, the possibility of a revenue-neutral tax experiment for the home mortgage interest deduction, where homeowners' associations could decide to participate or not.²⁴⁵ Tax experiments also might be executed at the level of a state, county, or locality. These approaches might better indicate the effect of tax rates on hiring than an experiment applying to individuals. It may be easy to obtain data on business activity or economic growth in a region or for an industry, even when it might be difficult to attribute that hiring to any particular individual. Similarly, unless the government plans to start collecting health information on tax returns, it may be more capable of evaluating health statistics already collected at various levels of government.²⁴⁶

Such tax experiments reduce but do not eliminate the danger that a tax experiment will simply shift economic activity.²⁴⁷ If a tax experiment results in some individuals receiving a tax increase and others receiving a tax decrease, then taxpayers may seek to take advantage of these discrepancies. For example, a married couple filing separately²⁴⁸ might seek to ensure that income appears on the tax form of the member of the couple who enjoys the lower tax rate.²⁴⁹ Such games are more difficult to play when a tax is applied to an entire city, county or state. There is,

242. See I.R.C. § 5000A (2012).

243. *Nat'l Fed'n of Indep. Bus. v. Sebelius*, 567 U.S. 519, 566 (2012).

244. Tax Cuts and Jobs Act of 2017, Pub. L. No. 115-97, § 11081, 131 Stat. 2054, 2092 (codified as amended at I.R.C. § 5000A (Supp. V 2018)) ("ELIMINATION OF SHARED RESPONSIBILITY PAYMENT FOR INDIVIDUALS FAILING TO MAINTAIN MINIMUM ESSENTIAL COVERAGE.").

245. See *supra* Section II.D.

246. The National Center for Health Statistics, for example, collects much relevant data. See *National Center for Health Statistics*, CENTERS FOR DISEASE CONTROL & PREVENTION, <https://www.cdc.gov/nchs/index.htm> [<https://perma.cc/F4EG-VBS6>] (last visited Nov. 7, 2018) (providing information on the Center).

247. See *supra* Section II.C.2.

248. I.R.C. § 1(d) (2012) (allowing married individuals to file a return that is not joint with his or her spouse).

249. See, e.g., *Poe v. Seaborn*, 282 U.S. 101, 108–09 (1930) (example of such an arrangement).

however, still a danger that economic activity may move across city or state borders. At least on the margins, businesses will have incentives to move activity from a location in a high-tax treatment group to a location in a low-tax treatment group. Thus, tax experiments can produce distortions. Moreover, tax experiment may give one business an advantage over a competitor, making tax changes much more consequential than if they were applied uniformly. This is problematic also because it may lead to misinterpretation. But if experimental units are sufficiently large, such effects may be relatively small, and experiments of this sort may provide the best available evidence of the effects of tax policy.

CONCLUSION

This Article has considered a range of possible applications of tax experimentation, from relatively small issues such as the entertainment deduction to foundational questions about the effect of marginal tax rates on labor supply and even the effect of tax policies on taxpayer health. Perhaps one reason that the possibility of tax experiments has been neglected is that it seems politically implausible that the government would randomly assign taxpayers to different tax rates, let alone assign different municipalities to different tax rules. This Article's ambition has been to show that more modest tax experimentation, featuring revenue-neutral designs and voluntary participation, might sometimes be possible, and that interpretive challenges can be addressed. Initial forays involving discrete code provisions could lead to voluntary experiments in which taxpayers receive some tax rate reward in exchange for their willingness to being assigned to any of many tax treatments. Such a program could increase the odds that the next great tax reform is based on a solid foundation of evidence.