2019

Tax Experimentation

Michael B. Abramowicz
George Washington University Law School, abramowicz@law.gwu.edu

Follow this and additional works at: https://scholarship.law.gwu.edu/faculty_publications
Part of the Law Commons

Recommended Citation

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in GW Law Faculty Publications & Other Works by an authorized administrator of Scholarly Commons. For more information, please contact spagel@law.gwu.edu.
Tax Experimentation

Michael Abramowicz *

Forthcoming Florida Law Review

* Professor of Law, George Washington University.
Abstract

Random experiments could allow the government to test tax policies before enactment 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 may be helpful in addressing questions about effects of tax rates and in increasing participation.

Contents

I. REVENUE-NEUTRAL TAX EXPERIMENTS ..................................................12
   A. A Hypothetical Experiment .........................................................12
      1. The Potential Benefit to Taxpayers ....................................13
      2. Adjustments to Ensure Revenue Neutrality ...........................14
      3. Evaluation of Experimental Success ....................................16
      4. Enhancement of Experimental Interpretability ......................18
         a. Varying the Treatment ..................................................19
         b. Two-Level Randomization ............................................21
         c. Intent-to-Treat Randomization ....................................23
      B. Other Revenue-Neutral Tax Experiments .................................24
         1. Deductions ........................................................................24
            a. Disallowing Other Deductions ....................................24
            b. Allowing New Deductions ..............................................26
         2. Income ............................................................................26
            a. Imputing Income ..........................................................26
            b. Taxing Work Amenities ...............................................28
            c. Changing Recognition Timing .....................................28
         3. Tax Procedure ....................................................................29
         4. Radical Tax Reform ..........................................................29
   II. CHALLENGES FOR TAX EXPERIMENTATION .................................31
      A. Transitions ........................................................................31
         1. Increased Scope ..............................................................31
The Tax Cuts and Jobs Act of 2017\(^1\) 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 inefficiencies and budget strains\(^3\) will be national in scope. The experiment has no control group. If the

\(^1\) Pub. L. No. 115-97, 131 Stat. 2054 (2017). The bill was entitled the “Tax Cuts and Jobs Act” until right before it passed, when the Senate parliamentarian ruled that the title violated reconciliation rules. 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.” Most commentators continue to refer to the bill as the Tax Cuts and Jobs Act (or TCJA). \textit{See, e.g.}, Stephen K. Cooper, \textit{Eyes Turn Toward 2018 Tasks as Tax Reform Becomes Law}, 158 Tax Notes 28 (Jan. 1, 2018) (“The law, known informally as the Tax Cuts and Jobs Act (P.L. 115-97)...”); Amy Hamilton, \textit{Connecticut Finds a SALT Workaround That Would Actually Work}, Tax Analysts, Feb. 26, 2018 (referring without qualification to the “Tax Cuts and Jobs Act”).

\(^2\) David Kamin et al., \textit{The Games They Will Play: An Update on the Conference Committee Tax Bill} (Feb. 26, 2018), \textit{available at} https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3089423 (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 over 10 years of $1.8 trillion. \textit{See}
economy grows rapidly over the next few years, reform proponents will likely take credit, while critics will insist that the economic growth was exogenous. Both sides will make their case with anecdotal evidence rather than the results of randomized experiments that economists generally prefer. Scholars have identified an empirical deficit in tax scholarship, but because federal tax law exhibits little exogenous variation, the deficit is difficult to correct.

This Article describes an approach to tax reform that might have been and that still might be: an experimental approach, in which proposed reforms are tested initially on groups of willing taxpayers. In recent years, the legal literature has focused attention on the design and justifiability of randomized government policy, in areas including consumer protection, securities law, patent law, and even regulation of food safety. Randomized experiments produce the best attainable evidence of the effect of legal policies on behavior, but the only known


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 Edward Fox, Do Taxes Affect Marriage? Lessons from History (2017) (unpublished manuscript, on file with author).

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.


randomized experiments involve tax procedure\textsuperscript{15} and welfare,\textsuperscript{16} not substantive tax law. Interest in the effects of tax law on taxpayer behavior pervades the literature,\textsuperscript{17} has been virtually no consideration of whether tax law might benefit from systematic randomized experiments.\textsuperscript{18}

Tax law, however, 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.\textsuperscript{19} Each tax change would apply to a random selection of qualified taxpayers agreeing to opt in to 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.\textsuperscript{20} Meanwhile, tax experiments could be much cheaper than many other forms of legal experimentation, because the existing obligation to file tax forms\textsuperscript{21} 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.\textsuperscript{22} This value makes especially salient

\begin{itemize}
  \item \textsuperscript{15}See, e.g., STATE OF CALIFORNIA FRANCHISE TAX BOARD, READYRETURN PILOT: TAX YEAR 2004 STUDY RESULTS (2006) (reporting results of an experiment on sending taxpayers pre-filled returns), available at \url{https://www.ftb.ca.gov/readyReturn TY04RRFinalReport.pdf}. Id. at 2. 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, TAX ANALYSTS SPECIAL REPORT, Aug. 30, 2010, at 555, 559 (arguing that taxpayers using ReadyReturn paid no more taxes than others, but without any comparisons between taxpayers and control group to support this argument).
  \item \textsuperscript{16}See DAVID KERSHAW & JERilyn FAIR, 1 THE NEW JERSEY INCOME-MAINTENANCE EXPERIMENT: OPERATIONS, SURVEYS, AND ADMINISTRATION (1976).
  \item \textsuperscript{17}See, e.g., William J. Congdon et al., Behavioral Economics and Tax Policy, 62 NAT’L TAX J. 375 (2009).
  \item \textsuperscript{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. 1, 43 (2012) (referring to experiments mentioned infra note 24 and accompanying text).
  \item \textsuperscript{19}See, e.g., Jacob Cohen, Statistical Power Analysis, 1 CURRENT DIRECTIONS IN PSYCHOL. SCI 98, 98 (1992) (noting that statistical power depends on sample size).
  \item \textsuperscript{20}See generally JIM JACCARD & ROBERT TURRISSI, INTERACTION EFFECTS IN MULTIPLE REGRESSION (2003) (describing techniques for identifying interactions between multiple variables of interests).
  \item \textsuperscript{21}See, e.g., I.R.C. § 6012 (specifying who must file U.S. federal tax returns).
  \item \textsuperscript{22}Alan J. Auerbach & Kevin A. Hassett, A New Measure of Horizontal Equity, 92 AM. ECON. REV.
concerns 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, and random experiments granting low-income individuals have occurred. 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 placebo. If uncertainty about relative efficacy suffices to justify medical experiments, 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, so future policymakers may need to raise more revenue, not to 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, taxpayers opting in to tax experiments will generally be well-advised and wealthy, 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

1116, 1116 (2002) (“[T]here is virtual unanimity that horizontal equality—the extent to which equals are treated 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 ETHICAL AND LEGAL ISSUES OF SOCIAL EXPERIMENTATION (Alice M. Rivlin & P. Michael Timpane eds., 1975).

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

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. 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. See Franklin G. Miller & Howard Brody, Clinical Equipoise and the Incoherence of Research Ethics, 32 J. MED. & PHIL. 151, 156 (2007) (arguing that clinical equipoise is an incoherent theory and that ethical obligations in research are distinct from those in clinical settings).

27 See supra note 3.

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. The treatment group in such an experiment on average pays taxes as high as 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 about the distribution of the tax burden across income and other groups. Revenue-neutral experiments necessarily allow the government to assess the impact of trade-offs. Even if only relatively wealthy taxpayers opt 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 experiment, but not at the direct expense of other taxpayers. Their experience, moreover, may help generate tax reform of broader benefit.

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. Under prior law, taxpayers could deduct expenses such as tickets for a sporting event used to entertain a client, but taxpayers can no longer do so.

Commentators have often suggested that tax reform should be revenue neutral. See 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.

The Joint Committee on Taxation estimates that the Tax Cuts and Jobs Act will result by 2027 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. See Joint Committee on Taxation, Distributional Effects of the Conference Agreement for H.R. 1, The “Tax Cuts and Jobs Act” 5 (Dec. 18, 2017), available at https://www.jct.gov/publications.html?func=startdown&id=5054 (projecting, for example, that taxpayers with income of $1,000,000 or over will contribute 18.9% of total taxes, instead of 19.1% of total taxes under prior law).

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) as amended by Pub. L. 115-97 § 13304(a)(1)(A) (“No deduction otherwise allowable under this chapter shall be allowed for any item… With respect to an activity which is of a type generally considered to constitute entertainment, amusement, or recreation”).

Entertainment expenses were deductible if they were “directly related” to or “associated with” the taxpayer’s trade or business. I.R.C. § 274(a) (2012); Walliser v. Commissioner, 72 T.C. 433 (1974)
for restricting deductibility of such expenses is that taxpayers derive utility from such expenditures. Because the recipient of entertainment benefits generally need not include the benefits in income, deductibility immunizes such expenditures from taxation.

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. The entertainment deduction limits recouped only a tiny fraction of the revenue loss from other parts of the tax bill. But the lobbyists’ embrace or at least 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. But the example conceptually demonstrates that taxpayers should embrace revenue-neutral tax changes that remove decisionmaking distortions.

How might an experiment on entertainment deductibility have worked? Taxpayers could have volunteered to give up their entertainment deductions. The

---

33 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.” Hearings on Tax Recommendations of the President Contained in His Message Transmitted to Congress Before the House Comm. on Ways & Means, 87th Cong., 1st Sess. 12-13 (Apr. 20, 1961).

34 See United States v. Gotcher, 401 F.2d 118 (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); see also I.R.C. § 132(a)(3) & (d) (excluding fringe benefits from employee’s gross income when the expenditures would have been deductible by the employee if paid by the employee him- or herself).

35 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 (1974).


37 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. [difficulty of inferring intent of multimember body]
carrot to induce participation would be that these taxpayers would receive lower rates, just low enough that government tax revenues would be identical. If the government selected this rate, it might miscalculate or intentionally set the rate low, disguising a tax cut as an experiment.\textsuperscript{38} But there is a simple antidote: The law authorizing such experiments could provide that the total tax liability of treatment group taxpayers will be multiplied by a factor sufficient to ensure that revenues from these taxpayers will 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.\textsuperscript{39}

With so simple a formula for achieving revenue neutrality, the taxpayers most likely to opt in will 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 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 equally be used to experiment on the possibility of new deductions. For example, currently tax law does not allow taxpayers to deduct commuting expenses.\textsuperscript{40} 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.\textsuperscript{41} An experiment might test deductibility

---


\textsuperscript{39} More precisely, the tax payment multiplier would be \( \frac{100}{101} \) to ensure revenue neutrality, so the discount is approximately 0.99%.

\textsuperscript{40} I.R.C. § 262; Commissioner v. Flowers, 326 U.S. 465 (1946).

\textsuperscript{41} See generally Tsilly Dagan, \textit{Commuting}, 26 \textit{VA. TAX REV.} 185, 201-34 (2006) (analyzing policy
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.\textsuperscript{42} 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.\textsuperscript{43} 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.\textsuperscript{44} 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 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.\textsuperscript{45}

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.\textsuperscript{46} Many scholars are skeptical of this justification,\textsuperscript{47}
but the experimental approach that we describe here 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 we assume 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 rates48 and the allowance of reduced rates for certain passthrough income,49 on the ground that they will improve the overall economic climate, thus benefiting workers as well as owners of capital.50 Non-revenue-neutral tax experiments may thus be needed.51 For example, the 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 participants in tax experiments, 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.52

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

---

48 I.R.C. 11(b), as amended by Pub. L. 115-97 § 13001(a) (cutting corporate tax rate from 35% to 21%).
50 See Martin A. Sullivan, Corporate Tax Incidence Made Simple, 157 Tax Notes 454 (Oct. 23, 2017) (providing overview of the evidence that cutting corporate tax rates will, to some extent, benefit workers).
51 MARK P. KEIGHTLY, CONG. RESEARCH SERV., R41596, THE MORTGAGE INTEREST AND PROPERTY TAX DEDUCTIONS: ANALYSIS AND OPTIONS 8–11 (Mar. 18, 2014) (reviewing the literature on whether the home mortgage interest deduction creates positive externalities to justify the tax benefit), reprinted in 2014 TAX NOTES TODAY 55-17.
52 See infra Part III.A.
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. Alternatively, the government might allow all taxpayers to opt into the treatment group, without requiring it of any taxpayers.

Part I 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.

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 we will use provisions from the Internal Revenue Code and recent tax reform statutes and proposals as examples, our analysis is not dependent on the structure of the U.S. income tax system. Part I.A elaborates the entertainment deduction experiment discussed in the introduction, and Part 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.

53 See infra Part II.A.
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%, 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. Our 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.

**Table 1. Hypothetical effects of an experiment**

<table>
<thead>
<tr>
<th></th>
<th>Gross income</th>
<th>Entertainment Expenses</th>
<th>Taxable income</th>
<th>Tax rate</th>
<th>Taxes</th>
<th>Take home income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status quo</td>
<td>$110K</td>
<td>$10K</td>
<td>$100K</td>
<td>25%</td>
<td>$25K</td>
<td>$75K</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>$110K</td>
<td>0</td>
<td>$110K</td>
<td>24.5%</td>
<td>$27K</td>
<td>$83K</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>$108K</td>
<td>$2K</td>
<td>$106K</td>
<td>23.5%</td>
<td>$25K</td>
<td>$81K</td>
</tr>
</tbody>
</table>

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, we 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.

---

54 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 factors such as the standard deduction. See I.R.C. § 63(c) (allowing taxpayers to deduct a fixed amount of money in lieu of itemizing deductions).
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.\(^{55}\) 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, we 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 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.\(^{56}\) We are 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 we 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 its 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.\(^{57}\) The law of large numbers is essential not only for statistical validity, but also because revenue

\(^{55}\) For a proof, see https://stats.stackexchange.com/questions/265622/sample-size-proportion-per-control-vs-experiment-group (last visited March 1, 2018).

\(^{56}\) Below, we will consider the caveat that the treatment group’s tax change might have some indirect effect on the control group. See infra Part I.A.4.

\(^{57}\) 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).
neutrality means that the tax liability of each treatment group participant depends on all other participants’ tax returns.\(^{58}\)

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 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 our purposes; a simple multiple linear regression model might work well,\(^{59}\) or the government might perform a nonlinear regression\(^{60}\) or even use machine learning techniques, such as a neural network regression\(^{61}\) or a decision forest.\(^{62}\) 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

---

\(^{58}\) 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.

\(^{59}\) See generally PAUL D. ALLISON, MULTIPLE REGRESSION: A PRIMER (1st ed. 1998).

\(^{60}\) See generally GEORGE A.F. SEBER & C.J. WILD, NONLINEAR REGRESSION (2003).


\(^{62}\) See Weida Tong et al., Decision Forest: Combining the Predictions of Multiple Independent Decision Tree Models, 43 J. CHEM. INF. COMP. SCI. 525 (2003).
on the money for the period the government held it.\textsuperscript{63} 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.\textsuperscript{64} If final reconciliation of the experiment takes a while, for example because some taxpayers fail to file their tax returns on time,\textsuperscript{65} the government could make an initial adjustment a few months after the 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.\textsuperscript{66} A taxpayer who plans to reduce its entertainment deductions for reasons not apparent based on available data would be particularly likely to opt in.

\textbf{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

\begin{itemize}
\item \textsuperscript{63} The IRS often pays interest on overpayments. See I.R.C. § 6611.
\item \textsuperscript{64} Id. § 6654.
\item \textsuperscript{65} Taxpayers who file late must pay penalties. Id. § 6651. Such penalties, excluding interest, could be included the comparison between the treatment and control group, to account for the possibility that the tax regime may affect the timeliness of filing.
\item \textsuperscript{66} ROBIN BROADWAY, FROM OPTIMAL TAX THEORY TO TAX POLICY 50 (2012) (discussing the importance of asymmetric information to tax policy analysis).
\end{itemize}
between success and failure, particularly because what matters is the *marginal* effect of the tax reduction on gross income.\(^\text{67}\)

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 businesses have long been allowed only partial deduction of certain classes of entertainment expenses.\(^\text{68}\) Ideally, a process of experimentation might lead the government to hone 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.\(^\text{69}\) 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 fisc. 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 questions of how the government can scale up an experiment, a question to which we will soon return.\(^\text{70}\) 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, our 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,\(^\text{71}\) especially if public officials receive benefits.\(^\text{72}\) The lower the effects on third parties, the stronger the case that a revenue-neutral tax reform will be welfare neutral as to nonparticipants.

---

\(^\text{67}\) 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.”).

\(^\text{68}\) I.R.C. 274(n) (disallowing 50% of the deduction for most meal expenses).

\(^\text{69}\) Cf. infra Part II.A (considering generalizing experiments beyond initial participants).

\(^\text{70}\) See infra Part II.A.

\(^\text{71}\) See LEONARD J. BROOKS & PAUL DUNN, BUSINESS & PROFESSIONAL ETHICS FOR DIRECTORS, EXECUTIVES & ACCOUNTANTS 388 (7th ed. 2015).

\(^\text{72}\) A taxpayer may not deduct business expense payments made in violation of state or federal law. I.R.C. § 162(c)(2).
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. The problem arises in medical experiments when participants may be able to deduce which group they are in. In 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.

“Hawthorne effects” occur when members of the treatment group behave differently because they know that they are in an experiment. Subjects might, in a tax experiment, focus more on the relevant tax issue, giving it outsized importance. Or subjects might be regret averse, that is, they wish to avoid feeling regret for the decision that they have already made to opt into the experiment. In the entertainment deduction experiment, an exaggerated cutback on 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. 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

---

73 See, e.g., Jefferson M. Fish, The Trouble with Double-Blind Placebo Studies, PSYCHOL. TODAY, Nov. 23, 2010 (noting that patients often can determine whether a pill is a placebo or biologically active), available at https://www.psychologytoday.com/blog/looking-in-the-cultural-mirror/201011/the-trouble-double-blind-placebo-studies.
74 Double-blind medical experiments typically include three groups: a control group, a placebo group, and a treatment group. See, e.g., id.
75 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).
77 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.
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 taxpayers might take shortcuts in reporting. Meanwhile, even if reporting is required for individuals in the treatment group and they face liability for misreporting,79 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 often 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 either 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.80 For example, the experiment might feature eleven groups in all, 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.81 Hawthorne and John Henry effects seem likely to be most

---

79 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. Commissioner, 100 T.C. 374, 380 (1993).

80 This is often described as a dose-response design. See William M. Holmes, Using Propensity Scores in Quasi-Experimental Designs 18 (2014).

81 See, e.g., Simon Bate & Natasha A. Karp, A Common Control Group—Optimising the Experiment Design to Maximize Sensitivity, 9 PLoS ONE (2014), available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263717/ (showing that with multiple treatments, sensitivity for comparison with the control is maximized with a larger control group).
pronounced at the extremes, when a taxpayer concludes that it 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 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, then 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 matched 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. 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

---

82 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, e.g., Ruta Brazauskas & Brent R. Logan, Observational Studies: Matching or Regression?, 22 BLOOD MARROW TRANSPLANT 557 (2015). Whatever design is used, taxpayers will be affected by those with similar characteristics.

83 See supra notes 59-62 and accompanying text.
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 deductibility altogether is likely to reduce transaction costs.\(^{84}\) If for political or transaction costs reasons, the law seems likely to settle at an extreme,\(^ {85}\) 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.

\(b\). \textit{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.

\(^{84}\) For a discussion of transactions costs in the tax system, see Kneave Riggal, 17 VA. TAX REV. 295, 307-08 (1997).

\(^{85}\) The current treatment of entertainment expenses is at the extreme of complete disallowance. I.R.C. § 274(a)(1). But meals with clients are 50% deductible. \textit{Id.} § 274(n).
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 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 if we assume that
being invited to participate has no direct effect on behavior,\textsuperscript{86} 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 simply 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 more sophisticated still. 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 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

\textsuperscript{86} 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 78 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.
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.  

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

1. Deductions
   
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. It is useful in part because it is relatively trivial and uncontroversial, allowing us to focus 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. A justification for this reform is that the tax law previously advantaged equity relative to debt, because equity is taxed twice. President Obama proposed to eliminate the deduction in part because it increases leverage in the economy as a whole, making it “more susceptible to

---

88 See Pub. L. No. 115-97, § 13301 (amending I.R.C. § 163(j)).
severe downturns." Others have argued that it would be better to eliminate the double taxation of equity, or to allow a deduction based on a corporation’s combined debt and equity. Meanwhile, there is a strong theoretical argument that business interest should be deductible. 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 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 changes in the statute, the imposition of caps on the deductibility of state and local taxes. 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” that lean Democratic rather than Republican. 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, eliminating the tax allows taxpayers to be taxed on the consumption benefits that they receive. There exist counterarguments and compromise proposals. 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

---

91 Id. at 9.
92 See Dubat, supra note 89 (arguing that this should be preferred to eliminating the deduction).
100 Kaplow, supra note 98, at 486 (noting that the deduction may promote spending on undersupplied public goods).
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 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. 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. 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. The theory is that a parent

---

102 See supra notes 40-42 and accompanying text.
who works at home in effect is paying herself to take a job, but her income from doing so is not taxed. Economists sometimes argue that imputed income should be taxable, 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 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. Taxpayers with older children who can be left alone might prefer this experiment as well.

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. A tax experiment might seek to move tax to the “baseline tax system,” 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.” Such an experiment might help eliminate distortions leading to excessive owner-occupied housing. In principle, similar experiments could test other forms of imputed income.

---

106 Taxpayers who would be below the standard deduction even with child care expenses would not benefit from deductibility. See I.R.C. § 63(c) (providing for a standard deduction).
107 This highlights that tax policies may have externalities. See infra Part II.D. Imputing income for child care might result in parents leaving their children alone more often, for better or more likely for worse.
109 Id. at 9 (arguing that imputed income is taxed in the baseline).
110 Id. at 10.
111 This thus implicates some of the same externality issues as the home mortgage interest deduction. See infra notes 222-225 and accompanying text.
b. Taxing Work Amenities

Beyond the question of imputed income, experiments on definition of income can explore fundamental questions about whether the classic Haig-Simons 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. Brooks points out that “[i]there is enough heterogeneity of psychic benefits across jobs and individuals to make universal assumptions unreasonable.” In theory, a tax authority might conduct a survey to rate the attractiveness of different jobs and impose 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, more than most such taxpayers, 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 meals to employee gyms.

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. A tax experiment might help evaluate some of the criticisms of a marked-to-market system, such as that it imposes substantial transactions costs, especially if the regime applies beyond publicly held securities.

---

113 Robert Murray Haig, The Concept of Income, 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.”).
114 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”).
115 Brooks, supra note 112, at 13 (citing Simons, supra note 114, at 53).
116 Id.
117 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).
118 “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).
119 Currently mark-to-market is normally available solely for securities dealers like stock brokerages. I.R.C. § 475; cf. id. § 1256 (providing mark-to-market on sophisticated financial instruments like futures contracts and foreign currency contracts).
120 See Edward A. Zelinsky, For Realization: Income Taxation, Sectoral Accretionism, and the Virtue of Attainable Virtues,
to assets such as art.\footnote{Noel B. Cunningham \& Deborah H. Schenk, \textit{Taxation Without Realization: A “Revolutionary” Approach to Ownership}, 47 Tax L. Rev. 725, 801-02 (1992) (discussing applicable rules for art objects).} A significant complication with a tax experiment of this sort is that the tax experiment must occur over a relatively long time horizon. We 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.\footnote{See infra Part II.C.2.}

3. \textit{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,\footnote{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, \textit{Detrimental Reliance on IRS Guidance}, 2015 Wis. L. Rev. 421, 431–37.} 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. \textit{Radical Tax Reform}

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

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. 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. If enacted economy-wide, then a combination of taxpayers in the value chain pay tax on all of the $150. If only some producers were required to pay tax, the system could become much more complex. Value-added taxes already create challenges of international harmonization, but this would magnify those challenges on a domestic level.

124 JANE G. GRAVELLE, CONG. RESEARCH SERV., R17413, INTERNATIONAL CORPORATE TAX RATE COMPARISONS AND POLICY IMPLICATIONS 3 (Jan. 6, 2014).
125 Dean Baker, Get Rid of Corporate Taxes, N.Y. TIMES, Jan. 13, 2016, at A21 (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.”).
126 Baker, supra note 125.
128 For a discussion of different ways of calculating the tax, see Itai Grinberg, Where Credit Is Due: Advantages of the Credit-Invoice Method for a Partial Replacement VAT, 63 TAX L. REV. 309 (2010).
129 See, e.g., id. at 321-22 (discussing some international issues).
II. CHALLENGES FOR TAX EXPERIMENTATION

This Part addresses challenges for tax experimentation. Part II.A 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. Part II.B addresses equity objections to tax experimentation, considering both horizontal equity and after-tax income inequality. Part II.C assesses whether taxpayers might be able to manipulate tax experimentation to their advantage, and Part II.D explores whether tax experiments can work with tax expenditures without adversely affecting other tax policy goals.

A. Transitions

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, 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. 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, for example, a higher proportion of eligible taxpayers 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, it can be seen as tax reform in and of itself, though of a different form from what one might

---

130 See supra Part I.A.4.
131 The possibility of externalities is addressed infra Part II.D.
132 See supra Part I.A.4.b.
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. Tax law already allows taxpayers to elect various options. 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 in to 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. 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.

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. Such experimentation is generally rare,

133 See, e.g., I.R.C. § 451(c)(2) (allowing taxpayers to elect a particular approach to advance payments).
134 See Miller & Brody, supra note 26; 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).
135 See Alice M. Rivlin & P. Michael Timpane, Introduction and Summary to ETHICAL AND LEGAL ISSUES OF SOCIAL EXPERIMENTATION 1, 1 (1975) (discussing a wide range of experiments).
but when the government has engaged in it, usually there is some benefit (e.g., a welfare program, \(^{136}\) a child’s eligibility for a school voucher program \(^{137}\)) 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.\(^{138}\)

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 responses depending on whether an experiment is opt-in or opt-out.\(^{139}\) 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.

---

\(^{136}\) See, e.g., KERSHAW & FAIR, supra note 16.

\(^{137}\) See, e.g., Paul E. Peterson et al., School Vouchers: Results from Randomized Experiments, in THE ECONOMICS OF SCHOOL CHOICE (Caroline M. Hoxby ed., 2003).

\(^{138}\) I.R.C. § 167 (allowing depreciation deductions for assets held for business or investment purposes); I.R.C. § 168 (providing method for calculating depreciation deductions for tangible property, typically over multiple years, even decades).

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

Some taxpayers who might not opt in to 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, and individuals are required to obtain

---

140 See Understanding Minimum Car Insurance Requirements,
health insurance under the Affordable Care Act for this reason. 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, 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 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. We thus turn to that concern.

B. Equity

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. The concern that tax experimentation might violate horizontal equity reflects the more general argument that randomization of government policy is inherently unequal. If the government randomizes taxpayers who have opted-into an alternative tax regime into a control

https://www.thebalance.com/understanding-minimum-car-insurance-requirements-2645473 (last visited March 1, 2018) (noting that 47 states require automobile insurance)

141 The Tax Cuts and Jobs Act eliminated the tax for noncompliance by setting the penalty to $0. See Pub. L. 115-97 § 11081 (amending I.R.C. § 5000A).

142 This concern is listed as a legislative finding in the Affordable Care Act. See 42 U.S.C. § 18091(I) ("If there were no requirement, many individuals would wait to purchase health insurance until they needed care.").

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

144 See Boris I. Bittker & Lawrence Lokken, Federal Taxation of Income, Estates, and Gifts ¶ 3.1.4 (1999 & Supp. 2017); cf. 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—the extent to which equals are treated equally—is a worthy goal of any tax system.")

145 For a 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).
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.\footnote{146} The general consensus in the literature has been that experimentation is acceptable if there is a sufficient justification for the difference in treatment.\footnote{147} Involuntary experimentation is most common in contexts in which individuals are thought to have lost their rights as a result of committing crimes.\footnote{148} In one notable criminal justice experiment, domestic violence perpetrators in the Bronx were randomly assigned to one of four different treatment programs.\footnote{149} That experiment concluded that treatment programs commonly employed throughout the nation for batterers may not be effective.\footnote{150} 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.\footnote{151} For example, experiments on supplements to the Earned Income Tax Credit (EITC)\footnote{152} have given the benefits to only some eager to participate.\footnote{153} Because the government expects to lose money on each member of the treatment group, revenue constraints limit the number of people included.  

\footnote{146} See supra Part I.A.4.b. 
\footnote{147} See, e.g., Abramowicz et al., supra note 9; Maurice Rosenberg, The Impact of Procedure-Impact Studies in the Administration of Justice, LAW & CONTEMP. PROBS., Summer 1988, at 13, 16. 
\footnote{148} See, e.g., Denise C. Gottfredson & M. Lyn Exum, 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). 
\footnote{150} Id. at ix (“Regrettably, our study suggests that some of the most prevalent court responses to domestic violence crime may be ineffective.”). 
\footnote{151} See, e.g., DAVID GREENBERG ET AL., SOCIAL EXPERIMENTATION AND PUBLIC POLICY MAKING 225 (2003) (stating that randomization “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”). 
\footnote{152} I.R.C. § 32. See generally BITTKER & LOKKEN, supra note 144, ¶ 37.1 (discussing the EITC in detail). 
\footnote{153} 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 (May 2014), http://www.mdrc.org/publication/paycheck-plus/file-full. This program involves both treatment groups (receiving the extra EITC) and control groups (who do not). Id. at 4.
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 being placed in the control group because we are not sure that the treatment is better than placebo. 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. 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 of whether the experiment may succeed in producing useful information. One might argue that tax experimentation is more troubling than other legal experiments because horizontal equity is an important tax-law goal. On this argument, it might be acceptable to have random experiments concerning patent policy, whose primary goal is to maximize economic efficiency, but less acceptable to allow inequalities to creep into tax policy. Some tax scholars, however, have criticized horizontal equity as a meaningless concept.

---

154 See Miller & Brody, supra note 26.
155 Abramowicz et al., supra note 147, at 965 (arguing that “random policy experimentation . . . will produce better information than nonrandomized experiments”).
156 See supra note 144.
157 See, e.g., Ouellette, supra note 13.
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.”

Tax law has long used randomization. For example, the IRS uses randomization in selecting which taxpayers’ returns to audit. 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. Increasing taxpayer utility while maintaining at least the same tax revenue is a legitimate government interest, 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.

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.165

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.166 This may happen for several reasons. First, higher-income taxpayers have

---

165 Lawrence Blume & Daniel L. Rubinfeld, *Compensation for Takings: An Economic Analysis*, 72 Calif. 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.”).

166 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 159, at 607; *see also* Louis Kaplow, *Horizontal Equity: Measures in Search of a Principle*, 42 Nat’l Tax J. 139, 140–41 (1989). “The vertical equity principle,” however, “does not prescribe whether tax rates should be
better tax advisors or 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, thus creating more opportunities for alternative tax regimes that benefit them. Third, wealthier taxpayers tend to be less risk-averse. Even if the 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, even a policy that dominantly benefits wealthy taxpayers should be seen as enhancing social welfare if other taxpayers are not harmed. Placing aside externalities, revenue-neutral tax experimentation, unlike virtually all other conceivable tax reform, is Pareto optimal, and Pareto improvements are often thought to be welfare-improving, despite experimental evidence that people sometimes are willing to accept Pareto-dominated outcomes. 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 proportionally, progressively, or regressively; nor, if progression or regression is the chosen mode, does it indicate how steep the slope should be.” Bittker & Lokken, supra note 144, ¶ 3.1.4; accord McDaniel & Repetti, supra, 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.”). 167 The simplest tax return, the 1040EZ, can only be used by taxpayers with taxable income below $100,000. See IRS, Which Form—1040, 1040A, or 1040EZ?, https://www.irs.gov/taxtopics/tc352 (last visited March 1, 2018). The most complicated return, the 1040, includes numerous schedules for more complex tax situations.


169 For example, consider the example alternative tax regime in section Error! Reference source not found., where taxpayers pay a lump-sum in exchange for lower marginal rates. Although the high-earnings-potential taxpayer H may expect to benefit from this alternative regime, if H loses her job, then H will be left much worse off than under generally applicable tax law, having to pay the lump-sum but getting no benefit from the lower marginal rates.

170 See infra Part II.D.


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


on relative wealth. 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 inequality is measured solely through a measure like the Gini coefficient, inequality is reduced. In this case, experimentation might produce greater inequality of happiness, but not greater financial inequality.

Suppose, however, that tax experimentation is thought to increase inequality under some definition sufficiently to 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, 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 effect redistribution. 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 experiments in several ways. First, taxpayers who know about an experiment may change their

---


177 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 (Sept. 2017). Household income would not include deductible business expenses.


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

behavior to become eligible to be invited 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. We 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 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. 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. 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

("Uncommon transactions that are taxed inappropriately become common as taxpayers discover how to take advantage of them.").

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 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. 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. 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 if the experiment failed for this reason, that would simply be an indication that the underlying hypothetical tax provision (elimination of a deduction but only for health professionals) is unworkable because

182 IRS, INTERNAL REVENUE MANUAL § 4.1.3.2 (“DIF mathematical formulas are confidential and for official use only. The DIF score assigned to a return should not be disclosed.”).
183 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 (2013).
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 a year 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 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.

These problems can be addressed. 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 would be developed of the behavior of control group and treatment group taxpayers, and 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, an experiment could even be designed 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 1.

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, the experiment is guaranteed to be revenue neutral over

---

\[184\] 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.
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 be advised to rigorously apply the existing tax-law doctrines that prevent shifting gross income and deductions between years, such as the constructive receipt doctrine\textsuperscript{185} and the economic performance doctrine.\textsuperscript{186} 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,\textsuperscript{187} and so the government might prioritize tax experiments that apply to only ordinary income (i.e., income that is not capital gains).\textsuperscript{188}

3. Status Changes

Individuals can get married or divorced; corporations can merge with others or divide into multiple corporations (e.g., 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—but $A$ 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 it makes no economic sense for $A$ to acquire $T$, aside from tax considerations. 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 $A$.

\begin{footnotesize}
\begin{enumerate}
\item 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, \textit{supra} note 144, ¶ 5.9.
\item 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). \textit{See generally Stephen Gertzman, Federal Tax Accounting} ¶ 4.04[3] (2016 ed.).
\item Joseph J. Cordes \& Harvey Galper, \textit{Tax Shelter Activity: Lessons from Twenty Years of Evidence}, 38 \textit{Nat’l Tax J.} 305, 322 (1985) (when capital gains rates are low, investments shift to activities generating capital gains).
\item Many other countries’ tax systems have very different treatment for income from capital and income from labor. \textit{See, e.g., Klaus Sieker, \textit{Portfolio 7140: Business Operations in Germany} ¶ IV.B.1 (1st ed. 2017 rev.)} (discussing Germany’s special treatment of capital income like dividends and interest); \textit{cf.} I.R.C. § 64 (defining “ordinary income”).
\end{enumerate}
\end{footnotesize}
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, 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.

But tax law already deals with similar problems of status changes. Individuals’ divorces and corporations’ dividing 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, then both the original corporation and the spun-off company will remain subject to the alternative regime. 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. 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

---

189 See, e.g., I.R.C. § 382 (dealing with corporate tax attributes like Net Operating Losses in the context of corporate acquisitions).
190 E.g., I.R.C. § 355 (providing tax rules for corporate separations like spin-offs, split-offs, and split-ups).
192 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 their calendar year and the other spouse using a different taxable year, I.R.C. § 6013(a)(2), and one spouse being a nonresident alien, id. § 6013(a)(1). See generally Bittker & Lokken, supra note 144, ¶ 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)).
taxpayers suffer a “marriage penalty,”193 others enjoy a “marriage bonus,”194 and under current U.S. tax law, the married-filing-separately status is generally seen as disadvantaged relative to the others.195 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.197 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,198 and there are strong arguments that the United States should switch as well.199 Of course, a tax experiment might be used to determine whether to eliminate this option in the United States. Meanwhile, there is no impediment to tax experimentation in countries with individual filing.

In corporate mergers,200 where the target company is subject to an alternative regime but the acquiring company is not, there are two possible options. First, the target may be required to maintain separate corporate form as the acquirer’s subsidiary,201 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.202 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, with transfer pricing regulation seeking to ensure that transactions reflect what parties would agree to in arms-length negotiations.

Second, the target may be allowed 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. The proportion of nominal tax liability attributable to the treatment group entity would be used 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 their stocks’ market capitalization can be used to determine valuation. 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, exist to properly calculate taxpayer income. But other tax provisions are “tax

\[\text{to the dividends received deduction at I.R.C. § 243(a)(3).}\]

\[\text{Transfer pricing also may become an issue as the IRS seeks to limit the scope of passthrough income rules. See Kamin et al., supra note 2, at 11 (“[T]hese kinds of transfer pricing games among related parties have proven very difficult to stamp out in other contexts.”).}\]

\[\text{See, e.g., I.R.C. § 482.}\]

\[\text{An analogous regime is used by I.R.C. § 382(b)(1) to limit the use of tax attributes (like net operating losses and built-in losses) of an acquired corporation in proportion to the value of the acquired corporation. See generally BITTKER & EUSTICE, supra note 191, ¶ 14.42[3] (explaining functioning of § 382’s limitations).}\]

\[\text{Treas. Reg. 1.382-2(a)(3)(i) (providing that in calculating “value”, control premiums or similar considerations are ignored).}\]

\[\text{E.g., I.R.C. § 162.}\]

\[\text{Stanley S Surrey & Paul R. McDaniel, The Tax Expenditure Concept and the Budget Reform Act of 1974, 17 B.C. INDUS. & COM. L. REV. 679, 679–80 (1976) (distinguishing “the structural provisions necessary to the application of a normal income tax, such as the definition of net income” from “special preferences found in every income tax”).}\]
expenditures” that further some additional goal, such as creating positive externalities or reducing negative externalities. 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 credit is often defended as encouraging scientific research that will benefit society as a whole, because companies cannot appropriate the full value of their research efforts. Similarly, tax expenditures subsidizing higher education are justified by its positive externalities. Other examples include the various credits, deductions, and exclusions subsidizing energy-efficiency and clean-energy, all of which are justified as reducing the negative externalities from fossil fuels.

An argument can be made for preventing alternative tax regimes from removing such tax expenditures. Similarly, the absence of some deductions could be justified on externality grounds. For example, the absence of a deduction or exclusion for the cost of gas used commuting by car might be justified partly by

---


211 I.R.C. § 41; 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).

212 See, e.g., Martin A. Sullivan, Economic Analysis: Can a Patent Box Promote Advanced Manufacturing?, 147 TAX NOTES 1347, 1347 (June 22, 2015) (noting that “decades of research by leading economists indicates that externalities from R&D not only exist but are very large”).

213 See CONG. RESEARCH SERV., 112TH CONG., TAX EXPENDITURES: COMPENDIUM OF BACKGROUND MATERIAL ON INDIVIDUAL PROVISIONS 104 (Comm. Print 112-45, 2012) (“[B]usinesses in general are unlikely to invest in R&D in amounts consistent with its social returns.”).

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

215 See CONG. RESEARCH SERV., supra note 213, at 633.

216 I.R.C. § 45 (credit for renewable energy generation, as from wind).

217 I.R.C. § 179D (deduction for all or part of the cost of energy-efficient commercial building property).

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

219 See CONG. RESEARCH SERV., supra note 213, at 113–253.

environmental considerations. Allowing 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. 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 home-owners 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 (i.e., 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-in to 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. Of course, if the primary goal of an experiment were informational, then a home

221 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).

222 I.R.C. § 163(h)(2)(D).


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

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 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, we discussed one solution to adverse selection that makes taxpayers hesitant to participate: designating an experiment opt-out or even mandatory. But we 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

---

225 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).

226 See supra Part II.A.
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

Participation inducements also can be used to facilitate experiments on tax rates. Consider, for example, a tax experiment in which some taxpayers will be randomly selected to pay an extra two percent in taxes, while others receive a three percent 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, but randomized studies have been unavailable, and interpretive issues make it virtually impossible to generalize from existing evidence about long-term effects of marginal tax rate changes. 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.

---

227 See, e.g., Saez et al., supra note 18 (providing a review of the literature).
228 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.”).
229 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.”).
230 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 § 12.3, available at https://catalog.flatworldknowledge.com/bookhub/21?e=rittenberg-ch12_s03. If only marginal tax
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, we have seen above that concerns about horizontal inequity may be overstated. Taxpayers might purchase insurance against being randomized to the group that receives a tax increase, 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, an experiment can often be considered 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, success might be measured 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 already on the wrong side of the Laffer Curve 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 but that 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 rates change, then there should be no income effect, but virtually any tax reform will affect both inframarginal and marginal tax rates.

---

231 See supra Part II.B.1.b.
232 See Arthur B. Laffer, The Laffer Curve: Past, Present, and Future (Heritage Foundation Backgrounder No. 1765, June 1, 2004) (discussing the curve, which postulates that tax revenues will be zero at both 0% and 100% tax rates).
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 experiments\textsuperscript{233} 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.\textsuperscript{234} 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.\textsuperscript{235} 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 downturn.\textsuperscript{236} 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 the tax cut should be made permanent.

A reasonable objection is that it will be difficult to craft in advance of a tax experiment a measure that fully captures its 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 proxy for broader economic effects.

One form of 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

\textsuperscript{233} See Abramowicz, supra note 9, at 985-87 (discussing the possibility of self-execution).

\textsuperscript{234} See Senate Republicans Scramble to Find Revenue for Tax Bill with Vote Expected Friday, N.Y. TIMES, Nov. 30, 2017.

\textsuperscript{235} Some of these problems are political. See, e.g., Stephen K. Cooper, Senate Leaders Agree to Add Fiscal Trigger to Tax Reform Bill, TAX ANALYSTS, Nov. 29, 2017 (reporting Sen. Kennedy’s reaction to the prospect of automatic tax increases).

\textsuperscript{236} See generally Alan J. Auerbach & Daniel Feenberg, The Significance of Federal Taxes as Automatic Stabilizers, 14 J. ECON. PERSP., Summer 2000, at 37.
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. We noted above that we might be interested in whether businesses affected by a tax law change increase or decrease their hiring levels. Moreover, even when 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,”\(^\text{237}\) which the Supreme Court upheld as a tax,\(^\text{238}\) but this individual mandate was removed in the most recent tax reform.\(^\text{239}\) Tax returns might tell us 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. We have already seen one example of this, the possibility of a revenue-neutral tax experiment for the home mortgage interest deduction, where home owners’ associations could decide to participate or not. 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.\(^\text{240}\)

\(\text{\textsuperscript{237}}\) I.R.C. § 5000A.


\(\text{\textsuperscript{239}}\) Pub. L. No. 115-97 § 11081 (“Elimination of Shared Responsibility Payment For Individuals Failing To Maintain Minimum Essential Coverage”).

\(\text{\textsuperscript{240}}\) The National Center for Health Statistics, for example, collects much relevant data. See National
Such tax experiments reduce but do not eliminate the danger that a tax experiment will simply shift economic activity. 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 separately might seek to ensure that income appears on the tax form of the member of the couple who enjoys the lower tax rate. Such games are more difficult to play when a tax is applied to an entire city, county or state. There is, 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.

IV. 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.

241 See supra Part II.C.2.
242 I.R.C. § 1(d) (allowing married individuals to file a return that is not join with his or her spouse).