

An Optimal Tax System*

LOUIS KAPLOW

Abstract

A notable feature and principal virtue of *Tax by Design* is its system-wide perspective on different elements of the tax system. This review essay builds on this trait and offers a more explicit foundation for the report's general approach, drawing on a distribution-neutral methodology that is developed in other work. This technique is then employed to illuminate and extend *Tax by Design*'s analysis regarding the VAT, environmental taxation, wealth transfer taxation, and income transfers.

*The title is inspired by Slemrod (1990). I would like to thank Steven Shavell and the referee for helpful comments and the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support.

The second consideration is the crucial insight that the tax system needs to be seen as just that – a system. While we address the impact of each tax separately for simplicity of exposition, we focus throughout on the impact of the system as a whole – how taxes fit together and how the system as a whole achieves government’s goals.

Tax by Design (ch. 1, p. 2)

I. Introduction

Perhaps the signal feature of *Tax by Design* is that it takes a systemic view of optimal tax reform, both for clarity of thought and presentation and also for advancing a pragmatic proposal for the UK tax system. Specifically, although analysis and concrete revisions unavoidably focus on one element of the system at a time, they are guided by the understanding that each piece is part of a larger whole and that the role of a given component depends on the operation of the others. This methodological choice should be applauded, loudly and without reservation.

The central theme of this review essay is to reinforce this construct by making even more explicit the structure that underlies *Tax by Design*’s vision. The key elements are largely present, but not always fully developed and related to each other. Elaboration will both lend further strength to this reform document and also show how the method has some additional implications for the subjects considered therein.

Another significant aspect of *Tax by Design* is that it largely takes a distribution-neutral (and revenue-neutral) approach toward tax reform. In this regard, the rationale is not always entirely clear: There appears to be some mixture of concern for practical political constraints and for keeping separate multiple and possibly conflicting considerations to facilitate communication. This dimension would also benefit from more direct attention, for distribution-neutral analysis is particularly apt for executing the systemic approach embodied in *Tax by Design*.

Section II develops these two related points, which are central to my recent book, Kaplow (2008a). Tax design importantly involves a problem of instrument choice, one that is familiar to economists. Generally, it tends to be optimal to employ each instrument to address that objective to which it is best suited, leaving other objectives to other instruments. Regarding the tax and transfer system, this principle tends to imply that the focus of revenue-raising and distributive efforts should be on the income tax and transfer system, whose traits are best suited for these purposes (although, as is familiar, the match is not first best). Under this view, other instruments, such as environmental taxes or inheritance taxes, should focus on the particular margins to which they are addressed, essentially setting revenue and distribution to the side. This outlook is much in the spirit of Musgrave (1959), who advanced a conceptual separation of the distributive and allocative functions of government.

This compartmentalized approach can be justified by considering a particular distribution-neutral policy experiment, one that *Tax by Design* employs in its concrete analysis of the elimination of exemptions in the VAT, such as for food purchases: Undertake the VAT reform in question (or that of any other instrument) and simultaneously adjust the income tax and transfer system such that the package as a whole is distribution neutral. When that is done, one can legitimately set distribution to the side. Moreover, in the basic case (in which labor is weakly separable, as elaborated below), labor incentives are unaffected. In short, the standard equity-efficiency tradeoff involved in redistribution policy is effectively held fixed. All that remains is the pure efficiency effect, which if positive leaves the government undertaking this experiment with a surplus that, for example, could be rebated pro rata. The result is actually a Pareto improvement; hence, the seeming regressivity of elimination of the VAT exemption for food purchases is moot, for the poor are made better off along with everyone else. Section II further discusses how, when actual reforms do have distributive effects, this distribution-neutral mode of analysis remains quite helpful.

Section III applies this approach to a number of the subjects addressed in *Tax by Design*. First, it elaborates the VAT illustration since that is one of the more important reforms advanced, in the process explaining how moving to uniformity is likely to be even more advantageous than the report suggests. Then it examines environmental taxation, showing how a more rigorous foundation is provided, one that offers a sharper indication of whether, why, and by how much various environmental schemes (such as cap-and-trade and petrol taxes) optimally deviate from the first-best Pigouvian prescription that marginal taxes (or permit prices) should equal marginal external harm. Next, it addresses wealth transfer taxation, showing again how the suggested framework helps organize thinking about a complex set of issues, in some instances leading in different directions. Finally, it considers the analysis of benefits reform, providing more direct justifications and also some refinements and extensions in light of more systematic application of the proposed framework.

II. Distribution-Neutral Tax Design

As suggested, the design and reform of all manner of fiscal instruments are usefully analyzed employing a particular, distribution-neutral policy experiment. The basic setting starts with the optimal income tax and transfer framework pioneered in Mirrlees (1971), in which the government taxes the labor income of individuals whose abilities to earn income vary, but in ways that are not observable to the government. As *Tax by Design* reminds us, the need to tax labor income as a proxy for earning ability is the root cause of the labor distortion due to income taxation. The optimal redistributive income tax problem is to design a nonlinear income tax schedule (that includes a uniform transfer) so as to maximize some social welfare function, subject to a revenue target and individuals' private optimizing responses to taxation, often modeled as incentive compatibility constraints. As is standard, we will suppose that individuals' preferences are the same, so they differ only in earnings ability – an assumption that will be explored

in subsection 3. In addition, it will be assumed throughout this section that individuals' preferences are weakly separable in labor, a subject reopened in subsection III.1 on the VAT.

The present analysis will differ from optimal income tax analysis in key ways. Notably, rather than determining the optimal extent of redistribution, the given distribution will be taken as given, which, as mentioned, is also largely done in *Tax by Design*. Note, importantly, that this method does *not* assume that the present income tax and transfer system is set optimally. Optimal tax theorists had long thought that this agnosticism disrupted the ability to examine the optimal design of other tax instruments, which, if true, would cast a serious cloud over the *Tax by Design* project since its entire mission is to do just that. However, a new line of work develops an alternative and actually simpler, more intuitive way to assess policies, utilizing a distribution-neutral approach. The seminal paper is Hylland and Zeckhauser (1979), and much of my work in the last two decades, culminating in Kaplow (2008a), is devoted to generalizing and applying this method to a broad range of fiscal tools.

1. Analysis

For concreteness, consider commodity taxation: In addition to the income tax, the government can levy (linear) taxes or subsidies on each good or service that individuals might consume. Atkinson and Stiglitz (1976) famously showed that, with weak separability of labor, uniform taxation is optimal; that was assuming an optimal income tax but, as just suggested, this restriction will not prove to be essential. See Kaplow (2004, 2006; 2008a) and Laroque (2005). This choice is a good launching point because it directly answers the question of optimal relative rates under a VAT. Additionally, it is a good start on many other issues. For example, Atkinson and Stiglitz (1976, 1980) explained how the taxation of savings (capital) can be assimilated into this framework because consumption in each time period can be viewed as a different commodity, so savings taxation is tantamount to a differential tax, falling more heavily on future consumption. Various of my prior papers (e.g., Kaplow 1996, 2001, 2004, 2012) and my book (Kaplow 2008a) extend the framework to environmental taxation, wealth transfer taxation – gifts and bequests can be viewed as additional commodities – and public goods – providing more (less) of a public good than dictated by the Samuelson (1954) formula is analogous to a relative subsidy (tax) on that good.

Begin with a nonuniform commodity tax scheme – say, a VAT with exemptions for necessities like food – and an arbitrary income tax and transfer system, perhaps the current composite system in the United Kingdom. Consider a reform that (A) eliminates all tax differentials in the VAT (say, by setting all rates at the existing standard rate) and (B) simultaneously adjusts the income tax in a distribution-neutral fashion. To be more precise regarding the latter, set the income tax at each level of income so that, if individuals continue earning that level of income, they will have the same utility as

before.¹ This income tax adjustment will involve tax reductions to offset the additional taxes that individuals' must pay under the VAT, although it would involve an income tax increase if some individuals at the outset spent much of their disposable income on goods subject to above-standard rates that were reduced. This adjustment also will offset the utility effects of individuals adjusting their consumption choices since, as stated, the tax adjustment is designed to hold utility constant.²

Analysis of this composite reform proves to be straightforward. First, it obviously is distribution neutral, by construction, so distributive concerns can legitimately be set to the side. Second, in the present setting, there also will be no labor supply adjustments. The reason is that the reform keeps utility constant at its pre-reform level *for each choice of income level (and thus of labor effort) that an individual might make*. Since each choice of labor effort, both before and after the reform, maps by construction to the same level of utility, the same choice of labor effort that initially maximized utility will continue to do so after the reform.

What's left? The answer is that the distribution-neutral scheme just posited raises revenue. The reason has to do with the second source of the adjustment to the income tax schedule. The first was to offset the change in individuals' payments on commodity taxes (say, under the VAT). This revenue loss (gain) from the income tax precisely offsets the revenue gain (loss) from the commodity tax reform. The second source of income tax adjustment was to wash away the utility effects due to individuals' rearranging their levels of consumption. Now, we know that these utility effects must be positive, by revealed preference. Hence, individuals' must all be paying higher income taxes on this account, to absorb (offset) the utility gains from their consumption adjustments. Accordingly, this second aspect necessarily raises revenue.³

So far, we have seen that the distribution-neutral experiment holds everyone's utility constant but results in a revenue surplus to the Treasury. Thus, the reform appears to be beneficial. To complete the story, and adhere to revenue neutrality as well, let us rebate the surplus pro rata to the population. The consequence will be that everyone (every income group) is better off, a Pareto improvement. To confirm this, recall that the original distribution-neutral income tax adjustment kept all individuals' utilities at their

¹ Observe that this modification is to the entire income tax schedule, and (contrary to the impression of some readers of my work) it involves neither a uniform lump-sum tax or subsidy (except in the special case in which the reform has precisely the same monetary effect on utility at all income levels) nor individualized lump-sum taxes. The changed income tax and transfer schedule, applicable to all individuals, will at each income level have marginal rates that in general differ from those initially prevailing, and the effect of the difference in the tax schedule on labor supply at each income level will be equal in magnitude and opposite in sign to that of the basic reform. For formal elaboration and further discussion, see Kaplow (2006; 2008a, ch. 6).

² Policy experiments that provide compensation, to keep individuals on the same indifference curve, are familiar from the use in other settings of partial equilibrium analysis that isolates substitution effects.

³ There is an omitted third step concerning how rearrangements of consumption affect commodity tax revenues. Since we are now, by hypothesis, under a uniform system, there is no effect. However, if we were moving, say, from a uniform to a nonuniform scheme, there would be a revenue effect, and it would be negative (individuals shift from higher to lower taxed goods); moreover, this revenue loss would exceed the direct utility gain from consumption realignment.

pre-reform levels. From that point, everyone now receives an additional rebate, so it is clear that everyone's utility rises.

There exists another perspective from which to view this distribution-neutral approach to fiscal reform. Specifically, the foregoing analysis implies that failing to set each fiscal instrument (other than the income tax and transfer schedule) efficiently in the narrow, traditional sense is a Pareto dominated strategy. If one fails to rectify an existing inefficiency or implements a reform that creates an inefficiency, it is possible to adopt a (further) reform that makes everyone better off, as just described. That reform, of course, is one that corrects the inefficiency and uses the income tax to offset the distributive impact, thereby creating a surplus that can be rebated pro rata. Eschewing efficiency is leaving gains – for everyone – on the table. This teaching from elementary microeconomics turns out to extend to a world with costly redistribution because, as explained, distribution-neutral implementation holds the equity-efficiency tradeoff fixed in the basic case under consideration.

This result of the distribution-neutral approach, although hardly without qualification (see below), holds despite a number of contrary intuitions. First, the result follows no matter how regressive is the initial reform, viewed in isolation from the offsetting income tax adjustment. Any regressivity in the VAT modification is offset by the income tax adjustment that is designed to render the package as a whole distribution neutral. Second, the key result does not (in the basic case with weak separability) depend on labor responses to income taxation because the distribution-neutral first stage keeps labor supply constant.

Finally, the distribution-neutral approach is unaffected by the possibility that some commodities may have very elastic demands and others highly inelastic ones. That is, the analysis essentially nullifies the standard Ramsey prescriptions, which *Tax by Design* (ch. 6, pp. 7–8) seems at one point to endorse. Interestingly, the only specific mention of Ramsey in the entirety of *Tax by Design* appears in a footnote to this brief discussion. This inattention runs quite contrary to longstanding public economics teaching but, as the preceding analysis suggests, is entirely appropriate in light of Atkinson and Stiglitz (1976, 1980) and the subsequent extensions of that work. *Tax by Design* does not address this major shift, contains few citations to Atkinson and Stiglitz (one of which is to the part of the 1980 text that derives the Ramsey results, not mentioning the later discussions that suggest that these results should not guide analysis in a regime with an income tax), and attends little to the literature or the specific arguments that justify the approach that it largely embraces. Nevertheless, it is notable that this newer framework is indeed adopted, with traditional Ramsey analysis cast almost entirely to the side.

Reflecting on this analysis, we can see that imposing distribution neutrality is not merely a possible political constraint but rather constitutes a great virtue for the purpose of advancing analysis. It turns out to rationalize the core strategy of *Tax by Design*, which largely involves examination of each fiscal instrument in isolation, focusing on efficiency, and often setting distribution to the side, as well as ignoring labor supply

effects. Viewed more broadly, the basic approach to instrument choice advanced in the introduction and originally endorsed by Musgrave (1959) is validated as well. Economists, including public economists, have often operated as if this methodology was correct, so it is helpful to know that, at least under certain assumptions (on which, more below), it is.

2. Non-Distribution-Neutral Reforms

The most obvious limiting assumption of the distribution-neutral approach is, well, its assumption of distribution neutrality. *Tax by Design* embraces this assumption, but what if policy-makers following its recommendations do not? That is, what if they wish, while undertaking comprehensive tax reform, also to change the extent of redistribution? Fortunately, much of what is learned under distribution-neutral analysis continues to be helpful. Moreover, that knowledge enables one to sharpen the analysis of reforms that alter the extent of redistribution.

To see this, begin with any non-distribution-neutral reform – one that has a resulting income tax schedule that fails to hold distribution constant – and decompose it into the following two steps:

1. Construct the distribution-neutral version of the reform, as described above. That is, alter the income tax and transfer schedule in the actual reform so as to make the whole package distribution neutral.
2. Change the income tax and transfer schedule from that in step 1 to the schedule in the actual reform.

A number of observations are immediate. First, this decomposition is generally possible and conceptually straightforward. Second, analysis of step 1 is something we already know how to do: It is precisely the policy experiment analyzed in subsection 1. Hence, we know that, if the reform promotes efficiency, narrowly and traditionally viewed, there will be a surplus that can be rebated so as to make everyone better off. Also, if the reform reduced efficiency, there would be a deficit – and a reform in the opposite direction would tend to be optimal.

Third, notice that step 2 involves a *purely redistributive change to the income tax and transfer schedule*. Accordingly, the proper analysis of the step 2 reform is that of the standard optimal redistribution problem, as in Mirrlees (1971). Note further that this step is generic, which is to say that the analysis is the same regardless of the underlying reform that generated it – that reform was handled in step 1.

This decomposition has a variety of virtues, the first and foremost of which is conceptual clarity. A second, related value is to policy analysts, whether economists doing theoretical modeling, empirical analysis, or policy simulation, or any others: Namely, specialization is facilitated. Using step 1, we can analyze all matter of fiscal tools – a VAT, capital taxation, public goods, environmental taxation, inheritance

taxation – to a significant extent without having to engage in the analysis or value judgments inherent in the optimal income tax problem.

Additionally, results can more readily and clearly be communicated among analysts and to policy-makers. For example, an analysis of the optimal petrol tax (say, in light of congestion and pollution externalities) might be developed through various simulations that embed non-distribution-neutral tax adjustments. Any bottom line will reflect not only the modeling and empirics of the petrol tax but also that of redistribution: assumptions on the labor supply elasticity and the distribution of skills in the population and also the choice of social welfare function. When two analyses disagree, to what is that disagreement to be attributed? And if two agree, might they have offsetting disagreements in different domains? When analysis is not decomposed, as suggested – and is not distribution neutral to begin with – progress and dissemination of results is made much more complex. It is therefore very valuable that the analyses and recommendations in *Tax by Design* are largely undertaken from a distribution-neutral perspective, even if those implementing various reforms care about redistribution and may choose to alter it.

The two-step decomposition also makes it easier for policy-makers to pick and choose. Suppose that the decomposition reveals a non-distribution-neutral reform to be desirable on step 1 and undesirable on step 2. In that case, it would be attractive to implement step 1 alone. Indeed, as explained, this involves a Pareto improvement; hence, it might be politically feasible as well. Likewise, consider a non-distribution-neutral reform that is undesirable on step 1 and desirable on step 2. Then, adopting only step 2 is appealing. Why incur the inefficiency by including step 1? Again, its exclusion could be accomplished in a manner that makes everyone better off.⁴ If the real virtue of a reform lies in its distributive character alone, and the rest involves avoidable cost, then purely redistributive adjustments would be superior.

Analysts are sometimes drawn to politically feasible reform packages, such as some that may be under consideration at a given moment, rather than to distribution-neutral implementations. The foregoing suggests that this tendency can be a mistake. In addition, most policy analysts are not experts in predicting the political process. What is proposed today may be modified tomorrow, before it is implemented. Also, Pareto improvements do have some prima facie political plausibility, as suggested. Moreover, in the long run, the political forces that affect the extent of redistribution may well have their sway, even if in the short run some particular reform is implemented that runs against the tide. Hence, distribution-neutral analysis may be as or more pragmatic than any particular alternative an analyst might choose to study. Finally, given the aforementioned benefits of conceptual clarity, specialization, and communication, it has much to commend it. *Tax by Design* should be highly praised in this regard.

3. Other Distributive Dimensions

⁴ Simply consider implementing the entire package and then reversing step 1. Step 1, taken alone, is distribution-neutral. Moreover, if it is inefficient, reversing it would generate a Pareto improvement.

As explained at the outset of this section, the analysis has proceeded under the assumption that individuals' preferences are homogeneous, the only difference between people being in their earning abilities. Introducing heterogeneous preferences can have two types of effects on the analysis.

First, preference differences may interact with labor supply in ways that influence the distortionary cost of taxation. See, for example, the dimensions explored in Mirrlees (1976), Saez (2002), and Kaplow (2008a, sec. 5.C.2). As one illustration of this sort of phenomenon, if some goods (say, attendance at the opera or symphony) are disproportionately preferred by high-ability individuals (even if at the same income level), a higher tax on such commodities is attractive because it enables somewhat more redistribution for a given level of labor supply distortion – or, instead, less labor supply distortion for a given degree of redistribution.

Second, and receiving relatively more attention in *Tax by Design*, individuals may simply have different tastes: Some prefer chocolate, others vanilla. As a consequence, eliminating, say, a commodity tax preference for chocolate will relatively disadvantage chocolate lovers. These sorts of taste differences, it turns out, do not have obvious implications for optimal commodity taxation or other policies. See, for example, Kaplow (2008b) and Weinzierl (2009). Preference heterogeneity does make literal Pareto improvements unlikely. Instead, one might achieve approximate Pareto improvements, such as by income class. See Ng (1984). *Tax by Design* aptly notes that this sort of heterogeneity hardly constitutes a case against reform: Not only is there no particular reason society should prefer chocolate lovers, but if one had to pick a benchmark, neutrality seems more appealing than a discriminatory status quo.

Another reason that literal Pareto improvements may not be feasible is due to the fact that real tax and transfer systems tend to employ only a handful of brackets rather than allowing any, general, nonlinear scheme. Accordingly, a fully distribution-neutral tax adjustment may require, say, a 1.3% increase in one part of a given bracket but only a 1.1% increase in another part. When averaged to a 1.2% increase throughout the bracket, some will be paying 0.1% more than under a perfectly distribution-neutral scheme and others 0.1% less. Clearly, these idiosyncratic departures from distribution neutrality will not ordinarily be terribly consequential; if they were, one could always introduce a new bracket or adjust the break points.⁵ Moreover, the distribution-neutral construct is primarily an analytical device, not a specific policy prescription to be followed precisely or otherwise.

⁵ Many tax reform analyses, including *Tax by Design*, seem to favor fewer tax brackets on grounds of simplicity. However, this is one of the features of the system that, in itself, is not particularly problematic. Whether with fairly exact withholding by employers on behalf of the government, as in the United Kingdom, or individual tax preparation, as in the United States, these computations are performed by computer in any event. Any nonlinearity, even the use of a zero bracket or exempt taxpayers, does introduce incentives to game the system by income shifting, but this motive is based on the extent of the difference in marginal tax rates between brackets or across taxpayers. The number of brackets is not, in itself, material.

III. Applications

1. VAT

The VAT was already employed in section II for illustrative purposes in explaining how, in a basic setting, moving to uniform commodity taxation – equivalent to a VAT at uniform rates – tends to be efficient, even Pareto optimal with distribution-neutral implementation. The corresponding chapters in *Tax by Design* are among its best, both because of the importance of such reforms – great efficiency benefits and administrative gains to boot – and because of the clarity and sophistication of the analysis. Worthy of special note is the construction of an approximate distribution-neutral implementation and the discussion of preference heterogeneity, as just mentioned.

VAT reform is also an excellent vehicle for introducing important additional qualifications. One involves externalities, which will be postponed to the next subsection on environmental taxation.⁶ Another is labor separability, which assumption is employed above in showing that a distribution-neutral move to uniform commodity taxation will not affect labor supply. As is familiar since Corlett and Hague (1953), it tends to be advantageous to tax leisure complements and subsidize leisure substitutes because these actions help to offset the labor-leisure distortion induced by income taxation.⁷ The income tax inefficiently reduces labor supply, so other interventions are helpful if they push in the opposite direction.

To illustrate, most books (excluding occupational texts and the like) are leisure complements: It takes time to read; indeed, the value of the time spent reading most books greatly exceeds their direct cost.⁸ Thus, *Tax by Design*'s proposal to eliminate the VAT preference for books understates the efficiency gain. Not only is there a distortion between books and music (etc.), but the encouragement of books actually reduces labor supply as well, exacerbating the labor-leisure distortion. Indeed, both books and music are optimally be taxed at a higher than standard rate. (Now, most academics and other elites would cringe at the thought, perhaps for good reason due to some sort of externality, but perhaps also out of self interest or narrowness of perspective.)

Much more important, the favorable impact of eliminating the exclusion for food could, along similar lines, be greater because expenditures on food are so large. One must not forget the obvious point that purchasing raw food (flour, vegetables, and so forth) must, to a significant extent, be mixed with time in order to produce meals. Aborwerth and Whalley (2002) indicate that the optimal taxation of groceries on this

⁶ As *Tax by Design* indicates, similar analysis might be applied to “internalities” as well.

⁷ Some confusion on this matter has been caused by opposite statements of this maxim, with apparent support from formal derivations, in Atkinson and Stiglitz (1976) and elsewhere; for confirmation of the conventional wisdom and technical clarification, see Kaplow (2010).

⁸ This point about relative time intensity is important. Luxury cruises also take time, but involve intense expenditures; hence, they may well be leisure substitutes: They make increased labor effort more attractive because one needs high disposable income, relative to available time, to go on a luxury cruise instead of staying home to read a book.

account significantly exceeds the standard VAT rate, making the common practice of granting exemptions much more inefficient than suggested by the standard analysis of distortion of commodity choices.

The particular point about leisure complements turns out to be of real significance for analyzing VAT reform. It also illustrates a broader lesson: the value of more systematic analysis than even the system-oriented perspective reflected in *Tax by Design*. Analysts and reformers should methodically begin with the income tax and transfer system, a commodity tax regime, and any other features pertinent to the particular reform in question. Next, one should undertake a distribution-neutral analysis, employing standard simplifying assumptions like labor separability. Finally, one should relax any of the assumptions that might be relevant to the problem at hand. Following this approach in each and every instance, rather than merely having some looser sense of relevant factors in mind, streamlines analysis, clarifies thinking and the presentation of results, and sometimes pays added dividends by revealing important channels that might otherwise be overlooked.

2. Environmental Taxation

The proper analysis of environmental taxation – more broadly, employing differential taxes and subsidies to correct externalities (negative and positive ones) – involves a quite direct extension of the analysis of commodity taxation just discussed, as explained in Kaplow (1996, 2004, 2008a) and demonstrated formally in Kaplow (2012). To see this, review the underlying logic of the argument that, subject to certain qualifications, optimal commodity taxation is uniform. Consumers' consumption choices are undistorted if the price ratios they face reflect the true resource costs of producing the commodities in question. Moreover, when certain (already mentioned) assumptions are made, there is nothing to be gained from distorting these decisions with regard to reducing the labor-leisure distortion due to income taxation.

Externalities can be added to this story in a straightforward manner. Simply note that the true resource cost of each commodity includes any externalities, positive or negative, that production or consumption of that commodity entails. Hence, the proposition that consumption price ratios should equal the ratios of true resource costs means that they should reflect the sum of production costs and any negative externalities (or minus any positive externalities). This result can be achieved by setting all commodity taxes (subsidies) equal to the net external harm (benefit) associated with each good. Then, moving from any other regime to such a system, when implemented in a distribution-neutral fashion, will result in a Pareto improvement, as before.

An immediate implication is that distributional concerns about environmental taxation are to a certain extent misplaced for the reasons as discussed in subsections II.2 and II.3. *Tax by Design* seems often to adopt this perspective, but its treatment of distributive concerns is somewhat less systematic than when it addresses the VAT.

This clean result also contrasts significantly with the history of discourse on optimal second-best environmental taxation, some of which is discussed in *Tax by Design*. For a collection and survey, see, respectively, Goulder (2002) and Bovenberg and Goulder (2002). Much of the debate involves whether there may be a double dividend – an environmental tax might both correct an externality and allow reduction of distortionary income taxation – or perhaps not, and there even may be a cost, if the environmental tax is more distorting of labor supply than the revenue sources it replaces.

These disagreements can, to a substantial extent, be resolved straightaway by application of the present framework. We know that none of these alternative phenomena is operative under a distribution-neutral reform and certain additional assumptions. Hence, whenever different outcomes are discovered, it must be that one of these assumptions is violated. One often at issue is distribution neutrality: As Kaplow (2012) documents, much of prior work involves non-distribution-neutral reforms and identifies changes in distortion that actually are a consequence of the implicit changes in redistribution, not the environmental taxation per se.⁹

Other possibilities are suggested by other assumptions underlying the framework. For example, it may be that some relevant commodities interact with the labor-leisure decision. West and Williams (2007) suggest that the optimal tax on petrol is above the first-best Pigouvian level because driving is a leisure complement. This may well be true, but note importantly that this qualification has nothing to do with the fact that petrol or driving may involve externalities. Even if pollution, congestion, and accidents did not exist – so the optimal corrective tax was zero – it would be optimal to tax petrol on this account, precisely the result in subsection 1's discussion of the VAT, where externalities were assumed not to exist.

Once again, we see that a single, unified, distribution-neutral framework is a multi-purpose engine that greatly facilitates analysis of a wide range of fiscal instruments. *Tax by Design* is roughly aligned with the suggested approach but might have been sharpened by pursuing it more explicitly and consistently.

As a final note, *Tax by Design* explores a number of other issues pertinent to environmental taxation, both in general terms and with regard to climate change and motoring. This part of the analysis is a mixed success: Myriad factors might have been considered, only some (but many of the most important ones) are, and the analysis does not always reflect the state of the art in the field of environmental economics.

One shortcoming is that the comparison of taxes with cap and trade is based in significant part on the Weitzman (1974) prescription, with some modification in light of

⁹ These results are often obscured by the fact that the analyses use representative agent models; hence, the change in distribution is deeply implicit. In essence, the reform packages change the net slope of the income tax schedule that the representative individual in these models faces. However, the only reason for income taxation (rather than reliance on a lump sum tax) is that individuals' abilities differ, in which case we should care not only about changes in distortion but also about the changes in the implicit degree of redistribution entailed by such reforms.

Roberts and Spence (1976). Yet these standard comparisons are often inapt, and particularly so in the climate context for the simple reason that the instruments can (and no doubt will be) adjusted over time, and both regimes (unlike traditional command-and-control regulation) provide direct feedback. Keeping in mind that greenhouse gases are a stock pollutant, the aggregate level of which changes slowly, consider that, under a tax regime, regulators will observe the quantity emitted over time; this must be so if tax revenue is to be collected. Hence, if the taxes were originally set too high or too low, this will quickly be revealed, and adjustments can eliminate the problem. Likewise, under cap and trade, regulators uncertainty over control costs may lead them to set the initial permit quantities too high or too low. But the permit market generates a price, which will reveal firms' marginal costs, enabling the regulator over time to adjust the permit quantity. Thus, both schemes essentially allow one to harness firms' control cost information and set policy accordingly. See Kaplow and Shavell (2002) and Kaplow (2011).

Another issue at the intersection of fiscal and environmental concerns is the matter of grandfathering, whether through the distribution of free permits to preexisting polluters or exemptions of certain quantities of such emissions from taxation. The fiscal consequences involve not only revenue but also distribution, and they are properly subject to the foregoing analysis, although this is not usually done and *Tax by Design* is no exception. Additionally, the environmental consequences, though obvious and striking, are often overlooked. Specifically, there is a potentially large ex ante inefficiency due to the *prospect* of grandfathering. Climate policy has been openly debated for decades, and more stringent controls are likely in the future. If firms understand that high pre-enactment emissions will be rewarded with free permits or tax exemptions, this means that for long periods of time – including right now – incentives will be muted, even perverse. See Kaplow (2011). There are potentially large gains to be realized from convincing current onlookers that such relief will not be forthcoming; even moderate nudges of expectations in the right direction would be valuable. In this respect, *Tax by Design* is a mixed blessing, disapproving of free permits on fiscal grounds (based on what turns out to be incomplete analysis) but missing the opportunity to condemn them harshly on environmental efficiency grounds.

3. Wealth Transfer Taxation

Wealth transfer taxation is usually considered to be a subject largely unto itself, and *Tax by Design* is no different. Nevertheless, although the problem is more complex than some others, analysis is greatly facilitated if one employs, again, the distribution-neutral framework. See Kaplow (2001, 2008a).

To begin, recognize that gifts – both inter vivos and by bequest – are among the ways that donors may expend their disposable income. Hence, as a first pass, we should view donors' choices between whether to spend a pound on themselves or on their descendants much like we view their choices among other commodities, broadly viewed. If indeed gifts were in all respects like other forms of personal expenditure, this would

essentially complete the analysis because we could simply treat this case as an instance of commodity taxation.

Before considering important exceptions, it is useful to note some of what is already learned as a consequence of this formulation. Most narrowly, taxation of transfers involves distortion, which makes donors worse off. Many would respond that the donors subject to transfer taxation are high-lifetime-income individuals, so the tax is attractive on both revenue and distributive grounds. But we have already learned that this view is mistaken. Specifically, we are contemplating distribution-neutral reforms. For example, we might repeal existing transfer taxation and raise top income tax rates in an offsetting manner. Or we might expand transfer taxation (raise rates or lower exemptions) and reduce top income tax rates. We can impose the same net burdens on donors, leaving them with the same level of utility, and raise the same revenue.¹⁰ However, as we have seen, eliminating the distortion in donors' allocations of their disposable incomes would produce a gain that, if soaked up by our offsetting income tax adjustment, would leave the Treasury with a surplus that can be distributed as we wish.

Hence, in important respects, we should attend more to efficiency and less to both revenue and distributive concerns in analyzing transfer taxation. Through a modest extension, we can apply similar analysis to the savings question entangled with transfer taxation. To the extent we are considering the taxation of bequests, funded out of individuals' savings, we are influencing the burden on savings. However, just as we considered a distribution-neutral adjustment to wash out distributive effects, we could also consider transfer tax reforms that adjust the tax rate on savings in an offsetting manner. Again, all that would be left is elimination of the distortion in donors' transfer decisions.

Gifts, however, are not entirely like other expenditures, so a complete analysis must attend to the differences. Notably, gifts generate two types of externalities involving donees, both of which are mentioned in *Tax by Design*, although not in a manner that directly links them to the present framework. First, there is a positive externality, namely, that donors benefit not only themselves but also donees when they make a gift. Even altruistic donors consider only their altruistic utility and hence tend to give too little.¹¹ Second, there is (usually) a negative externality due to the income effect: Given the preexisting distortion due to the income tax, donees' inclination to work less imposes an external cost on the Treasury equal to their forgone earnings times their marginal income tax rate. A priori, either externality could be larger, and it could depend

¹⁰ An important subtlety is that a distribution-neutral reform that holds donors' utilities constant does not hold donees' utilities constant. Moreover, as their lifetime accessions increase, their marginal utilities of disposable income decrease, making it optimal for net redistribution to be less favorable toward them. See Kaplow (2001, 2008a) and Kopczuk (2009).

¹¹ *Tax by Design* mentions some prior work that views this externality as a sort of double counting, but that argument is obscure; notably, it does not say which half is not supposed to count in social welfare. Are donees in fact not benefited in a socially relevant manner? Or is it that society does not value the utility that donors obtain from supporting their children? (Perhaps this is to be deemed a socially perverse form of pleasure.) For further exploration of these sorts of normative questions, see Kaplow (2008a, pt. V).

on the nature of donors' motives as well as other factors, such as the age of the donee, the donee's relationship to the donor, whether the donee is liquidity constrained, and so forth.

One particular motive receiving some attention in *Tax by Design* is the non-motive of so-called accidental bequests, which arise as a consequence of incomplete annuitization. The report's analysis is incomplete in some respects, one of which is particularly important regarding the concrete reform proposed: extension of the UK transfer tax system from bequests to inter vivos gifts. That analysis sees the differential treatment as anomalous, an instance of differential taxation that is justified, if at all, by administrative concerns. Suppose, however, that analysis and evidence suggested that the optimal tax rate on accidental bequests was very high but that on other wealth transfers was notably lower (even zero or negative). In that case, the present regime might (by chance) make a good deal of sense: Avoidance of taxation that will be imposed at the time of death, which involves effort costs and expenditures, and perhaps unpleasantness, would only occur with regard to non-accidental wealth transfers, so they would be exempt (or could be taxed at a lower rate applicable to inter vivos giving), and truly accidental bequests would be taxed at the full rate applicable to inheritances. This observation is not meant to suggest that the status quo is optimal, or nearly so, but rather that there is more work to be done to resolve a significant tension between two widely held views regarding transfer taxation.¹²

Though largely independent of the foregoing, it is also worth emphasizing a point about equality of opportunity that *Tax by Design* handles better than most prior discussions. Specifically, for ages one of the most prominent arguments in favor of wealth transfer taxation is to provide equal opportunities, sometimes described as equal starting points, to those in the donee generation. However, most wealth transfers subject to taxation – and all of them in a system confined to bequests – come late in life, typically when donees are in their 50s. This is far past recipients' starting points, and actually much closer to their retirement. Inequality in starting points is overwhelmingly caused by differences in genetic endowment, upbringing, environment, education, contacts, and other opportunities, the most important of which are bestowed by parents far earlier in life. See Taubman (1996). Most of these are exempt from transfer taxation even in those systems that tax inter vivos gifts, such as that in the United States. Empirically, most wealth for most of the population is human capital. Moreover, on a broad social scale, the relevant inequality is not merely or even primarily that between the top few percent of the income or wealth distribution, where transfer taxes hit, and everyone else, but also includes the significant inequality among the rest, especially with regard to those in the bottom deciles.

4. Income Support

The proper analysis of income support is rather different from that presented thus far because income transfers are at the core of the central redistributive apparatus.

¹² A broader lesson is that it is not sufficient to attend to different transfer motives; one must also tie that analysis of motives directly to the broader framework, as my prior work advocates.

Nevertheless, the subject is best understood by explicit linkage to the formal framework.¹³ For transfers, we really are in the domain of the optimal income tax literature launched by Mirrlees (1971), who in his seminal paper remarked on the direct applicability to transfers. Indeed, to know optimal transfers, we should simply look at the lower end of the optimal income tax and transfer schedule, at which point we would appear to be done.

This approach is too little employed. It is, to be sure, incomplete, but the correct response is to extend it directly rather than to cast it aside. More modern work, including some built upon in *Tax by Design*, aims to do just this. Drawing on this methodology, as developed in Kaplow (2007, 2008a), a few ideas should be added to those in the report.

First, most tax and transfer systems, and much of the analysis in *Tax by Design*, is directed toward differential provision to those with different needs, notably, families with (especially young) children and the disabled. The proper way to analyze this feature is to conceptualize the optimal income tax problem extended to two or more groups, say a high need group and everyone else. Now, one can simply apply standard optimal income tax analysis to each group, giving each its own income tax schedule, including its own intercept, which is to say, the level of the transfer at zero income. (Formally, these two schedules are linked by a common multiplier, representing the shadow value of a pound in meeting the revenue requirement.)

Several implications follow. First, traditional phaseout analysis is inapt. Specifically, if a greater lump-sum transfer is made to the high-need group, there is no necessary requirement to make up for it by higher marginal tax rates, perhaps attempting to withdraw it quickly lest the expense become excessive. This steep phaseout approach may or may not be optimal, but it involves a choice because the transfer may (and optimally would) be funded to some, possibly a large, extent by taxation on the non-needy group.¹⁴

Second, consider the common prescription, endorsed in *Tax by Design*, that steep marginal tax (inclusive of benefit phaseout) rates at the low end of the income distribution should be avoided, so as to prevent discouraging labor effort. This point can readily be overstated. Significantly lowering marginal tax rates on low levels of income mean revenue losses not just for those individuals (which losses are moderate) but also

¹³ The same prescription applies to *Tax by Design*'s briefer discussion of the taxation of families, where attention is given primarily to incentives but the analysis is not linked explicitly to the optimal income tax framework that directly considers distributive effects and how they trade off with incentives, a problem that is more complex when considering different family configurations. See Kaplow (2008a).

¹⁴ A priori, it seems plausible that optimal marginal tax rates at the low end run contrary to standard phaseout thinking. Suppose that there are two groups, the regular population and one that is generally less able to work. The former will optimally receive a low lump-sum transfer, yet high marginal rates at the low end tend to be efficient because the density of the skills distribution is not that high and there are many individuals with higher skills for whom the high rates would be inframarginal. The latter group will optimally receive a high transfer, yet relatively low marginal rates at the bottom tend to be efficient because, within this low-ability group, the density of the skills distribution will be relatively high and, compared to the former group, there will be fewer individuals with higher skills for whom the low rates would be inframarginal. See Kaplow (2007, 2008a).

huge revenue losses with regard to all taxpayers with higher incomes, for whom these would be inframarginal income tax rate cuts. Additionally, the income effect further reduces the labor supply of higher-ability types. Therefore, lowering these tax rates can be extremely expensive. This point is not novel; indeed, starting with Mirrlees (1971), most optimal income tax simulations find it optimal to employ quite high marginal rates at the low end – well under 100%, but often quite high. A related reason for this result is that the lost productivity from discouraging the work of those least able is modest. There may be offsetting externalities, such as in low-ability or high-need parents setting examples for their children, but these effects would have to be massive to favor very low marginal rates. Furthermore, they would need to offset any opposite externality in raising children due to parents' increased absence from home if they spend more time in the workforce.

Another consideration mentioned in *Tax by Design* concerns some recent work favoring low or even negative taxes (marginal earnings subsidies) at the bottom, motivated by analysis suggesting that much labor supply response is at the extensive margin. Although work in this area is important and is providing valuable insights, we should be careful in overly applying results of such models. First, as just explained, the result involves large inframarginal revenue losses on all those with higher income, most of the rest of the income distribution. Second, these results are generated in significant part in models that assume both that higher-ability individuals cannot take the jobs of low-ability individuals and that higher-ability individuals cannot work part time. Were it otherwise, these subsidies to very low paying jobs would cause many of somewhat higher ability to reduce labor effort. However, while a doctor may not be able to take the job of a machinist, it is hardly clear that a very low skilled worker cannot take the job of a very, very low skilled worker; after all, these are truly unskilled jobs that are under consideration. Moreover, part-time work is common, including at low income levels, and scale economies involved in entering the workforce are not always present or large.¹⁵ Hence, although further study is certainly in order, it seems too early to endorse this type of reform as optimal.

One further point of conceptual and practical importance is worth mentioning. *Tax by Design* is typical in its focus on explicit income transfers to the exclusion of direct government spending on goods and services. Yet these expenditures, including on education and health care, contribute significantly, and perhaps disproportionately, to the well-being of individuals at the lower end of the income distribution. Moreover, like direct transfers, they affect labor supply. A more complete implementation of *Tax by Design*'s systemic approach would integrate this dimension of fiscal activity as well.

¹⁵ Blundell et al. (1999) present UK data indicating that part-time employment at all levels is very common for both single mothers and women in couples. Blundell and Walker (2001) present evidence for the United Kingdom that a substantial portion of childcare obtained by single parents involves the use of friends and relatives and that the marginal cost of childcare rises with hours of work as cheaper forms of childcare are exhausted first, suggesting that marginal costs of working may exceed rather than fall short of average costs for these individuals.

IV. Conclusion

Tax by Design and the accompanying *Dimensions of Tax Design* are publications to be celebrated – by academics, policy analysts, government decision-makers, and the public at large. The prospect that citizens in the United Kingdom, or anywhere, might be governed by a tax system that not only looks like it was consciously designed, but is formulated with the insights of modern economic analysis that are well displayed in these reports, is most welcome.

The present essay aims to highlight some of *Tax by Design*'s most important contributions while, at the same time, providing more explicit grounding for its central, system-wide approach. As explained in section II, its distribution-neutral ground rule is actually a powerful analytic device that supports much of the analysis and provides additional insights as well. Section III demonstrates these lessons with regard to a number of the key areas of focus and policy recommendation that are contained in the report. Much of *Tax by Design* is on the mark, in some cases even more so than may be apparent. As always, there is room for improvement and further extension, and the framework advanced here is meant to be helpful in advancing this cause.

References

- Ab Iorwerth, A. and Whalley, J. (2002), 'Efficiency considerations and the exemption of food from sales and value added taxes', *Canadian Journal of Economics*, vol. 35, pp. 166–182.
- Atkinson, A. B. and Stiglitz, J. E. (1976), 'The design of tax structure: direct versus indirect taxation', *Journal of Public Economics*, vol. 6, pp. 55–75.
- Atkinson, A. B. and Stiglitz, J. E. (1980), *Lectures on Public Economics*, New York: McGraw-Hill.
- Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (1999), 'Evaluating in-work benefit reform: the working families tax credit in the UK', (Preliminary Draft.)
- Blundell, R. and Walker, I. (2001), 'Working families' tax credit: a review of the evidence, issues and prospects for further research', Inland Revenue Research Report No. 1.
- Bovenberg, A. L. and Goulder, L. H. (2002), 'Environmental taxation and regulation', in A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, vol. 3, pp. 1471–1545, Amsterdam: Elsevier.
- Corlett, W. J. and Hague, D. C. (1953), 'Complementarity and the excess burden of taxation,' *Review of Economic Studies*, vol. 21, pp. 21–30.
- Goulder, L. H., ed., (2002), *Environmental Policy Making in Economies with Prior Tax Distortions*, Cheltenham, U.K.: Edward Elgar.
- Hylland, A. and Zeckhauser, R. (1979), 'Distributional objectives should affect taxes but not program choice or design,' *Scandinavian Journal of Economics*, vol. 81, pp. 264–284.
- Kaplow, L. (1996), 'The optimal supply of public goods and the distortionary cost of taxation', *National Tax Journal*, vol. 49, pp. 513–533.
- Kaplow, L. (2001), 'A framework for assessing estate and gift taxation', in W. G. Gale, J. R. Hines, Jr. and J. Slemrod (eds.), *Rethinking Estate and Gift Taxation*, pp. 164–215., Washington, D.C.: Brookings Institution Press.
- Kaplow, L. (2004), 'On the (ir)relevance of distribution and labor supply distortion to government policy', *Journal of Economic Perspectives*, vol. 18, no. 4, pp. 159–175.
- Kaplow, L. (2006), 'On the undesirability of commodity taxation even when income taxation is not optimal', *Journal of Public Economics*, vol. 90, pp. 1235–1250.

- Kaplow, L. (2007), 'Optimal income transfers', *International Tax and Public Finance*, vol. 14, pp. 295–325.
- Kaplow, L. (2008a), *The Theory of Taxation and Public Economics*, Princeton: Princeton University Press.
- Kaplow, L. (2008b), 'Optimal policy with heterogeneous preferences', *B.E. Journal of Economic Analysis & Policy*, vol. 8, no. 1 (Advances), Article 40.
- Kaplow, L. (2010), 'Taxing leisure complements', *Economic Inquiry*, vol. 48, pp. 1065–1071.
- Kaplow, L. (2011), 'Taxes, permits, and climate change', in G. Metcalf (ed), *U.S. Energy Tax Policy*, pp. 168–192, Cambridge, MA: Harvard University Press.
- Kaplow, L. (2012), 'Optimal control of externalities in the presence of income taxation,' *International Economic Review*, volume 53 (forthcoming).
- Kaplow, L. and Shavell, S. (2002), 'On the superiority of corrective taxes to quantity regulation', *American Law and Economics Review*, vol. 4, pp. 1–17.
- Kopczuk, W. (2009), 'Economics of estate taxation: a brief review of theory and evidence', *Tax Law Review*, vol. 63, pp. 139–157.
- Laroque, G. (2005), 'Indirect taxation is superfluous under separability and taste homogeneity: a simple proof', *Economics Letters*, vol. 87, pp. 141–144.
- Mirrlees, J. A. (1971), 'An exploration in the theory of optimum income taxation', *Review of Economic Studies*, vol. 38, pp. 175–208.
- Mirrlees, J. A. (1976), 'Optimal tax theory: a synthesis', *Journal of Public Economics*, vol. 6, pp. 327–358.
- Musgrave, R. A. (1959), *The Theory of Public Finance*, New York: McGraw-Hill.
- Ng, Y.-K. (1984), 'Quasi-Pareto social improvements', *American Economic Review*, vol. 74, pp. 1033–1050.
- Roberts, M. J. and Spence, M. (1976), 'Effluent charges and licenses under uncertainty', *Journal of Public Economics*, vol. 5, pp. 193–208.
- Saez, E. (2002), 'The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes', *Journal of Public Economics*, vol. 83, pp. 217–230.

- Samuelson, P. A. (1954), 'The pure theory of public expenditure', *Review of Economics and Statistics*, vol. 36, pp. 387–389.
- Slemrod, J. (1990), 'Optimal taxation and optimal tax systems', *Journal of Economic Perspectives*, vol. 4, no. 1, pp. 157–178.
- Taubman, P. (1996), 'The roles of the family in the formation of offsprings' earnings and income capacity', in P. L. Menchik (ed), *Household and Family Economics*, pp. 5–45, Boston: Kluwer Academic Publishers.
- Weinzierl, M. (2009), 'Incorporating preference heterogeneity into optimal tax models: de gustibus non est taxandum', (unpublished manuscript).
- Weitzman, M. L. (1974), 'Prices vs. quantities', *Review of Economic Studies*, vol. 41, pp. 477–91.
- West, S. and Williams, R. C. (2007), 'Optimal taxation and cross-price effects on labor supply: estimates of the optimal gas tax', *Journal of Public Economics*, vol. 91, pp. 593–617.