New Economic Perspectives on Telecommunications Regulation

William P. Rogersont


INTRODUCTION

During a decade of work culminating in the publication of A Theory of Incentives in Procurement and Regulation in 1993, Jean-Jacques Laffont and Jean Tirole wrote a pathbreaking series of papers that established a new principal-agent paradigm for analyzing incentive problems in procurement and regulation. Although they have gone on to slay major theoretical dragons in other areas of economics, they have continued to devote at least a fraction of their talents and energies to regulation. Over the last five or six years, they have produced a series of original and interesting papers analyzing new issues and problems that have arisen in telecommunications markets as technology rapidly changes and governments around the world attempt to introduce competition into these markets. As this book makes abundantly clear, the authors have devoted an enormous amount of time and energy over the past five years to gaining an in-depth understanding of competitive and regulatory issues in telecommunications markets. Putting a sophisticated and up-to-the-minute understanding of real world facts in the hands of two of the world’s best theorists creates a dynamite combination and this book is the de-
lightful result. They describe the important institutions in the telecommunications industry and provide an overall framework for thinking about the key incentive issues in this industry. Besides presenting the results of their own recent work, they sketch out a much broader theoretical overview of the industry. They include thoughtful and clear summaries of other recent research and also provide clear and accessible treatments of some of the old standard economic ideas (such as Ramsey pricing) when these ideas play an important role in their thinking. Topics covered include how access to a monopoly bottleneck should be regulated when the monopolist supplies access not only to itself but to potential competitors, how universal service can be supplied in a competitively neutral fashion once entry of competitors has occurred, and whether or not there is a need to regulate prices of interconnection when competing telecommunications networks exist.

The structure of the book will greatly increase its accessibility to a broad audience. The main text of the book is essentially free of formal models, and the book is written so that it can be read independently of these models, which are contained in boxes set off from the main text. I have read other books that follow this format where it was clear that the main concern of the authors was with the formal models in the boxes, and attempting to read the text without reference to the formal models was very frustrating and difficult. This is not the case with this book. They consistently present complete and logically consistent verbal developments of their arguments and analysis in the text. Therefore I think that this book will be useful to a very broad audience.

Graduate students and professors engaged in building formal models of regulation will find a lively, engaging treatment of the hottest and most recent theoretical work in the area together with a wonderfully thorough set of references to the literature. Applied economists and policymakers will find a clear and insightful framework for thinking about incentive issues in telecommunications. By describing the important incentive issues in telecommunications and presenting what seem to them to be the important and fascinating questions that remain unanswered, Laffont and Tirole have undoubtedly set an influential agenda for future research and policy debate. Everyone interested in either the theoretical or policy aspects of telecommunications regulation will want to read this book.

In the remainder of this review, I will discuss four topics that Laffont and Tirole focus on in their book. I will present my perspective on what the issues are, describe some of Laffont and Tirole's conclusions, and suggest other approaches that might complement some of their work.
I. EFFICIENT PRICING IN TELECOMMUNICATIONS

One of the early and significant contributions of economic theory to regulation policy is the theory of Ramsey pricing, and Laffont and Tirole devote a large part of chapter two (pp 60–84) to a clear and thoughtful explanation of this theory. The concept of Ramsey pricing is relevant in situations where there are large fixed costs of production so that the marginal cost of producing another unit of output (which does not include any allocation of the fixed costs) is significantly lower than the average cost of production (which includes an allocation of fixed costs). The most basic idea in welfare economics and price theory is that it is efficient for a good to be produced and consumed if the marginal cost of production is less than the marginal benefit of consumption. Accordingly, if prices are set equal to marginal costs of production, consumers will make efficient consumption decisions because they will purchase the good if and only if their marginal benefits from consumption exceed the marginal cost of producing the good. The problem, of course, is that the firm will not be able to cover its fixed costs if it sets prices equal to marginal cost. One solution that may be possible in some circumstances is to have the regulated firm recover its fixed costs through levying connection charges of some sort on consumers. Ramsey pricing essentially deals with the case where connection prices are not possible so that the fixed costs of production must be recovered through raising prices above marginal costs.

If the firm produces only a single product, there is no further analysis to be done to determine the second-best solution. The firm simply assigns all its fixed costs to the single product and charges a price equal to the average cost of production. Consumption is less than the efficient quantity, but there is nothing else to be done about it. However, when the firm produces multiple products, the decision of how the fixed costs should be allocated among products for purposes of cost recovery still remains. The idea behind Ramsey pricing is to choose an overhead allocation method that minimizes the social loss from inefficient consumption decisions. Frank Ramsey first suggested this optimization problem, and the prices that result are labeled Ramsey prices. Standard accounting practice and perhaps even some notions of fairness would probably suggest that overhead should be allocated over products in proportion to marginal costs or some similar scheme. The key qualitative property of Ramsey prices turns out to be

---

2 Chapter 2 covers some basic principles of regulation. The remainder of this chapter contains a very clear and accessible presentation of the basic idea underlying the authors' earlier research using principal-agent models to analyze incentives and the moral hazard problem in regulation and procurement.

that fixed costs are allocated across products in inverse proportion to the elasticity of demand for the products. That is, more overhead is allocated to products if demand is unresponsive to price and less overhead is allocated to products if demand is responsive to price. This is relatively intuitive. Since the problem with allocating overhead to products is that this causes demand to fall below the efficient level, the optimal method for allocating overhead is to allocate more overhead to products where there will be a smaller effect on demand.

Laffont and Tirole go on to explain Vogelsang and Finsinger’s more modern conclusion: Ramsey prices can be implemented by delegating the choice of relative prices to the regulated firm through a price cap system (pp 66–67). Under a price cap system, the regulated firm is allowed to choose relative prices any way it wishes subject only to the constraint that its average price (using last period’s quantities as weights to calculate the average) is no higher than the period before. As Laffont and Tirole stress, this is an important result, because it addresses the criticism of Ramsey pricing that the regulator will never be able to estimate demand elasticities precisely enough to directly calculate Ramsey prices (pp 91–96, 131–32). The Vogelsang-Finsinger result shows that the regulator does not need to know demand elasticities to implement Ramsey prices. By choosing an appropriate price cap scheme, the regulator can induce the regulated firm itself to choose Ramsey prices.

The fact that a firm will choose markups in inverse proportion to demand elasticities under a price cap plan is actually quite intuitive. Under a price cap plan, a firm is always allowed the option of raising one price so long as it lowers some other price to keep its average price the same. Suppose, for example, that we begin with a situation where a firm is selling equal amounts of two products, called products A and B, and that the firm is charging the same markup above marginal cost on both products. The price cap rule would allow the firm to raise the price of one product by a dollar if it lowered the price of the other product by a dollar. Suppose the demand of product A is more elastic than the demand for product B. In particular, assume that a one dollar price change in product A would cause demand for A to change by (say) three units, but a one dollar change in the price of product B would only cause the demand for B to change by (say) two units. If the firm raised the price of product B by one dollar it would lose profits of

---


5 Of course, this argument assumes that the regulated firm is able to estimate price elasticities. However, as Laffont and Tirole point out, firms in all markets generally make pricing decisions based on their estimates of demand elasticities, and there is no reason to believe that regulated firms would be particularly unable to do this.
two dollars because of lost sales of B. However, if it lowered the price of product A by one dollar if would gain profits of three dollars due to increased sales of A. Therefore, price cap schemes inherently give the regulated firm an incentive to raise the price on products with less elastic demand (because the loss of sales will be small) in order to lower the price on products with more elastic demand (because the gain in sales will be high).

Laffont and Tirole observe that Ramsey pricing is almost never used in regulated markets, and, in particular, that it is not used in telecommunications markets. They provide a careful and thoughtful discussion of the practical as well as theoretical pros and cons of Ramsey pricing and conclude that, on balance, there is a strong and clear case for greater use of Ramsey pricing in regulated markets, and, in particular, in telecommunications markets (pp 73–80, 84–96).

I would like to offer three comments about this conclusion. First, I think that the book’s abstract discussion of Ramsey principles could have been more interesting and more relevant to telecommunications if the authors had focused more attention on a specific feature of local telephone networks. In local telephone networks, a large portion of the cost of providing telephone service is the cost of building the loop (i.e., the connection between the end user and the telephone company’s switching office), and the cost of building the loop does not really vary at all with the amount of time that the end user uses the loop. This unusual cost structure raises a number of interesting pricing issues in the telecommunications industry.

I will begin by introducing a minimal amount of background information about the nature of telephone networks. A local network is usually thought of as consisting of three parts. The first part, called the loop, consists of pairs of copper wires that run from each end user to central offices owned by the local telephone company. The second part of the network is the central office and the switch that resides in the central office. The switch is the intelligent part of the network that directs calls to the correct location. The third part of the local network, called the transport, consists of high capacity transmission lines that carry signals between switches.

There are two important facts to know about the loop. First, the loop is by far the most expensive part of the network and is the part of the network where economies of scale are most strongly present. Second, the cost of installing a loop to an end user is essentially completely independent of the minutes of use that the end user makes of

---

Some experts think that it might be relatively economic to have multiple carriers install their own switches and transport. However, there is wide agreement that it would be extremely inefficient to have multiple carriers build their own parallel sets of copper loops in most areas except for the most dense business areas.
his telephone. That is, the same pair of copper wires must be installed to permit an end user to use his phone one minute per day, one hour per day, or twenty-four hours per day. In the parlance of economists, then, loop costs are completely fixed with respect to output when output is measured in minutes of use. The historic practice in the United States has been to recover a large share of loop costs by adding a per minute fee onto long distance telephone calls. That is, when a customer of one local telephone company makes a long distance call to the customer of another local telephone company, both local telephone companies charge the long distance company a per minute fee for either originating or terminating the call, and the long distance company in turn passes these charges on to the person making the call. The practice has been to add the same per minute charge to all types of calls. These charges are typically referred to as access charges, and the local telephone companies are typically referred to as supplying the product of “local access” to the long distance companies.

I will now describe the aspect of the efficient pricing problem in telecommunications that the abstract Ramsey formulation does not fully capture. The Ramsey problem can be viewed as focusing on the issue of how a firm should recover a fixed cost if it incurs the fixed cost to serve the needs of multiple consumers. This is because if a firm incurred a fixed cost to serve the needs of one and only one consumer, the firm could always recover this fixed cost in a completely nondistorting fashion by charging a fixed connection fee. However, I think that much of the pricing problem in telecommunications can be usefully understood as a failure to charge the consumer fixed fees to recover fixed costs incurred directly to serve only that consumer. From an economic perspective, the user of the loop is being charged a fee per minute for using the loop (for long distance calls) even though the cost of supplying the loop to the user is completely independent of the number of minutes that the loop is used for. In economic terms, the cost of supplying the loop to the consumer is a completely fixed cost and the marginal cost of supplying a minute use of the loop is zero. However, the local telephone company is allowed to recover a portion of the costs of building the loop by charging a per minute fee to the consumer, which reduces use of the loop below the efficient level. Attempting to recover the cost of building a facility (the loop) by charging a consumer for how often he uses the facility (minute of use access charges) even though there is no cost to using the facility creates what

---

7 Switching and transport costs are more sensitive to minutes of use because larger switches and transport lines have to be installed if phones are used more. In these cases, costs are probably more affected by peak usage than by average usage, however.
many economists view as perhaps the most significant pricing error in telecommunications markets.

Suppose, for example, that the average cost of building a loop in a particular neighborhood could be recovered by a fixed charge of $20 per month per loop. Then the simple principle of not recovering facilities costs through usage fees would suggest that each loop be charged a $20 per month fee but that there be no per minute charges for use of the loop for long distance calls. All of this analysis can be done without any recourse to Ramsey pricing principles.

How does Ramsey pricing enter the picture or change this analysis? Although the average cost per loop may be twenty dollars per month, the marginal cost of a loop may be only three or four dollars per month. This is where Ramsey pricing comes into play. Ramsey pricing theory would suggest that if consumers' demands for loops were responsive to the per month fee they were charged for loops, then consumers with higher elasticities should be charged higher monthly fees. While every consumer should pay at least the marginal cost of three or four dollars, the remainder of the loop costs should be recovered by charging monthly fees that are higher for users whose demand for loops is less elastic.

How might Ramsey pricing principles alter the simpler policy prescription that all loops should be charged the same $20 per month connection fee? It is not clear that there are significant elasticity differentials among residential consumers in the demand for first lines that could be usefully taken into account. This suggests that charging the same connection fee to all residential consumers for a first line probably is consistent with Ramsey pricing principles. However, it seems likely that the residential demand for additional lines to the same house could be much more elastic (i.e., whether or not there is a need for an extra line for a computer or fax machine or the children might be the sort of luxury good for which demand is price sensitive). If this is true, then it might be efficient to charge significantly lower connection fees for additional lines to a residence. In a similar vein, since the number of lines demanded by a business is probably sensitive to price, it might be efficient to charge lower per line connection fees to businesses demanding multiple lines. Note that the efficient prices in this case might strike some people as being quite unfair. For

Note that the cost of supplying switching and trunking may be usage dependent, and I am only focusing on recovery of loop costs in this discussion, which are clearly usage independent.

An extra factor at play in the case of loops is that there are economies of scale in building loops. That is, once a network of loops is in place in a neighborhood, the cost of creating a loop for a particular homeowner is much less than the average cost of creating loops for the neighborhood as a whole.
example, it might appear to some that less wealthy homes that demand a single telephone line would be subsidizing more wealthy homes that demand multiple lines. This brings me to my second comment.

My second comment relates to fairness concerns. Laffont and Tirole very clearly explain that one of the main reasons that regulators have been slow to adopt Ramsey principles is that Ramsey pricing often seems unfair to them (pp 132–36). While in some cases, demand curves may be inelastic because people are rich (rich people continue to purchase even if price rises), in other cases demand curves may be inelastic because people have no alternatives (people with no alternatives continue to purchase even if price rises). The idea of charging more to people that have fewer alternatives is anathema to many regulators. To them, treating people fairly assumes a far greater importance than potential aggregate efficiency gains for the economy as a whole, especially if these efficiency gains are likely to be largely directed to owners of businesses.

Therefore, from a policy perspective, simply finding policies that could make every person better off (in the sense that aggregate efficiency is increased) may not be enough. It may also be important to go on to devote more attention to finding policies that simultaneously effect income transfers between groups so that every person actually is made better off. I will return to this theme later in my discussion of universal service, since policy developments in this area provide a very nice example of policymaking that takes this concern into account.

My third comment concerning Laffont and Tirole's discussion of the potential benefits of Ramsey pricing relates to the effects of such pricing on entry and the development of competition. As the authors correctly point out, one of the factors that affects the elasticity of demand for a product is the extent to which competitive entry will occur in response to price rises (p 104). If a price rise will induce significant entry by new firms that steal business from the incumbent, then demand will be more elastic. From this perspective, one of the effects of Ramsey pricing is therefore to minimize the amount of competitive entry that occurs. This is because prices are raised for products where entry will not occur and lowered for products where entry would otherwise occur.

This, of course, is perfectly efficient from the point of view of a static economics model. That is, the entry that occurs because prices are raised above marginal costs is actually inefficient in the sense that production could be accomplished more cheaply by the incumbent;

---

10 The student of economics will recognize this difference in goals as the difference between Kaldor-Hicks efficiency and Pareto efficiency.
therefore, minimizing this type of entry is actually efficient. However, from a more dynamic perspective, encouraging entry that appears to be inefficient in the short run may be highly desirable if the effect of this entry is to increase the amount of competition to the point where regulation is no longer necessary in the long run. That is, if it is actually optimal to subsidize entry of new firms in the short run in order to escape the costs of regulation in the long run, then pricing policies that violate the Ramsey rule and actually encourage entry may be desirable.

This point applies, perhaps even with more force, to another aspect of the efficient pricing problem. An important feature of loop costs is that they can vary dramatically over different regions of the country. One of the main determinants of loop costs in an area is simply the population density of the area; the average cost per loop is much smaller in more densely populated areas. For example, the average cost of a loop for the United States as a whole is thought to be approximately $20 per month. However, loop costs in dense business areas are often in the neighborhood of $5 per month while loop costs in rural areas can easily exceed $80 per month. Local telephone companies generally charge the same fixed monthly fee for local service to all their customers regardless of their location. Therefore, customers are being charged prices that can differ dramatically from the actual costs of serving them. Principles of efficient pricing, of course, would require that end users in higher cost locations pay higher prices.

In particular, for the case of telephones, it has been widely observed that the policy of charging averaged prices has created an artificially large incentive for firms to enter into low-cost, dense business areas. From a static short-term perspective this is inefficient, since entry is occurring to some extent simply because the local incumbent telephone company is being forced to charge prices higher than its costs.

However, one of the major goals of telecommunications policymakers in the United States is to encourage the development of competition in local telephone markets so that regulation will no longer be needed. It may well be, then, that artificially handicapping incumbents in the most profitable areas of their territories is actually a reasonably good way of encouraging such entry. On the other hand, it also seems

\[1\]
In fact, the situation is actually worse than this. Residential customers are generally charged less than business customers. Since residential customers are generally in less dense areas than business customers, customers in less dense areas are therefore charged less, on average, than customers in dense areas.

\[2\]
For example, it is widely recognized that precisely this sort of situation created large incentives for MCI and other firms to enter the long distance market during the late 1970s and early 1980s. This entry in turn eventually resulted in the unregulated competitive long distance
possible that, even if it is desirable to subsidize entry in order to encourage competition, more efficient and desirable ways of accomplishing this exist than implicitly handicapping the incumbent in some areas of its territory.

Clearly, the entire subject of whether it makes economic sense to purposely distort regulated prices in order to encourage the development of competition is a topic that deserves more economic research and that is particularly relevant to telecommunications markets today.

II. ONE-WAY ACCESS

When AT&T was broken up in 1983 into a single long distance company and seven separate local telephone companies (often referred to as the Baby Bells), the main justification for this division was that long distance markets were quickly growing competitive but that local markets were still clearly natural monopolies requiring regulation. The main concern was that an enormous problem would be created if a single firm was allowed simultaneously to operate a regulated local telephone company and a deregulated long distance company. The problem was that the firm would have a powerful incentive to evade regulation of its local monopoly by degrading access to other long distance suppliers (p 163) and then raising the price of its own long distance service. It was felt that there were many ways for a local company to degrade the connections of its rivals to the local network and that regulators could not possibly monitor this large number of practices. Therefore, the only way to avoid this problem was to isolate the regulated local network in a separate company that was not allowed to supply long distance service. Another potential problem that concerned regulators was that a single firm providing both regulated local service and unregulated long distance service would have an incentive to engage in accounting and managerial cross-subsidies which might be difficult to prevent fully.

The Telecommunications Act of 1996 allowed entry of the Baby Bells into long distance markets once certain specific conditions were satisfied, in particular that the local market was open to competition. Once a local market becomes competitive and the Baby Bell serving this market is no longer subject to any sort of regulation, the fear that the Baby Bell could escape regulation by vertically integrating into market that exists today.

---

13 See United States v American Telephone and Telegraph Co, 552 F Supp 131 (D DC 1982), aff'd as Maryland v United States, 460 US 1001 (1983). This consent decree is commonly known as the "MFJ" (Modification of Final Judgment).

14 See the exclusionary measures listed on pp 164–65.

15 Pub L No 104-104, 110 Stat 56 (1996), codified in various sections of title 47.

16 See 47 USC § 271(c) (Supp 1996).
long distance would, of course, no longer be relevant. Arguably this is what Congress had in mind when it added this condition to the Telecommunications Act of 1996. While this may be the rationale that Congress had in mind, the law it passed is not consistent with this rationale. In particular, the so-called competitive checklist of conditions that must be satisfied in order for a Baby Bell to be allowed to enter long distance markets actually requires the development of only minimal amounts of competition. In particular, even when these conditions are satisfied, the Baby Bell will still generally have significant amounts of monopoly power, and regulation of prices will still be required. For example, Bell Atlantic recently received the first approval of any Baby Bell to enter the long distance market to serve its customers in the state of New York. However, there is absolutely no plan afoot to end price regulation of Bell Atlantic in New York. This is because it obviously still has significant market power for many classes of customers in many regions of the state. This means that, generally speaking, once the Baby Bells are allowed to enter long distance markets under the provisions of the 1996 Telecommunications Act, their large incentives to evade regulation of their local markets by degrading access to long distance competitors and raising their own long distance prices will, once again, become a significant issue.

Laffont and Tirole describe this issue in chapters three and four of their book and more generally explore the question of how a regulated firm’s access price should be set when it is allowed to enter an unregulated market that uses this access as an input. They report the results of two of their papers. Both papers assume that the regulator is allowed to regulate not only the incumbent’s local telephone service but also the price that it charges for long distance service. Their first paper shows that regular Ramsey pricing principles can be applied to this problem. They show that, in general, so long as there are fixed costs of producing local telephone service that must be recovered through per minute charges, these costs should be recovered both through access charges that the local telephone company charges other long distance companies, as well as through charges on end users for local service. In particular, this means that access prices should not generally be set equal to marginal cost. Their analysis also sheds

17 See 47 USC § 271(c)(2)(B).
19 For example, Bell Atlantic is still the only facilities-based provider of local telephone service to the vast bulk of residences in New York.
considerable light on the nature of the efficient component pricing rule ("ECPR"), a pricing rule for access that has been widely advocated in various regulatory proceedings. In particular, Laffont and Tirole’s analysis characterizes conditions under which the ECPR would be desirable.

Their second paper is perhaps even more interesting. They define a global price cap to be a price cap formula that includes both the price of the incumbent’s long distance service and its local service, and show that a global price cap can be used to induce the incumbent to set the Ramsey-optimal prices (pp 170–78). Furthermore, the incumbent will have no artificial incentive to attempt to degrade access to its competitors under the optimal price cap. Therefore, the main concern of regulators is solved by such a scheme.

These are fascinating results that clarify a number of issues related to one-way access pricing. However, one important limitation of the results is that they assume that the regulator is allowed to regulate not only the incumbent’s price of local telephone service but also its price of long distance service—even though the long distance market is thought to be competitive and none of the other firms supplying long distance is regulated. As a practical matter, I think that it is inconceivable that the FCC would be allowed to regulate the long distance prices of Baby Bells (even by including the long distance price in a price cap) when they were not regulating the long distance prices of any other carriers. Therefore, an interesting question for policy purposes concerns how the incumbent’s price of access should be regulated if its price of long distance service is unregulated and whether there are any other ways of dealing with the incentive of the incumbent to degrade access.

III. TWO-WAY ACCESS

As entry into local telephone markets has progressed, the issue of what sorts of prices one telephone network should be allowed to charge another network for interconnecting with the network has assumed increasing importance. In the long run, as the government’s regulatory role in local markets diminishes and prices become largely


22 Armstrong and Vickers have made some progress on this problem. See Mark Armstrong and John Vickers, The Access Pricing Problem with Deregulation: A Note, 46 J Indus Econ 115 (1998). They characterize the optimal access price when the incumbent’s long distance price is not regulated. Perhaps their most interesting result is to show that regulating the absolute level of the access price is always preferable to regulating the margin between the access price and the long distance price. At least some people have interpreted the ECPR rule as suggesting that regulators should regulate the margin between the access price and the long distance price instead of directly regulating the absolute level of the access price.
deregulated, the issue of whether interconnection prices competitively set in oligopoly markets are socially efficient will assume greater importance. This latter question is already of paramount importance in the internet backbone market, which is largely unregulated.

The FCC's rules governing the pricing of interconnection between networks are a confusing hodgepodge of inconsistent rules with no single underlying rationale. In some circumstances, when one carrier drops off a call to be terminated on another carrier's network, the originating carrier must pay the terminating carrier a per minute fee. In other circumstances, the terminating carrier must pay the originating carrier a per minute fee. In still other circumstances, neither carrier is required to make any payment to the other carrier. The FCC is currently considering a rule that would allow the terminating carrier to charge a terminating fee directly to the customers of the originating carrier in some cases. Of course, once we consider the possibility that the terminating carrier might charge the customers of the originating carrier, this raises the question of whether or not regulated carriers are allowed to flow back interconnection charges to their customers when they are charged interconnection fees by the other carrier. Once again, under current regulations, the answer varies tremendously based on circumstances that do not seem to have much economic significance.

I think that the most interesting research that Laffont and Tirole describe in their book are the two papers that they have written describing how interconnection prices are determined in an unregulated oligopoly environment. These are among the first papers on this subject, and the contribution of these papers is as much to develop a framework for thinking about these issues as to derive any particular results. They begin by explaining why there is a possibility that competing firms might agree to high interconnection prices as a way of facilitating collusion between themselves, but then go on to identify a range of other factors that must be considered as well (pp 187-207).

24 See Federal Communications Commission, Calling Party Pays Service Offering in the Commercial Mobile Radio Services, 64 Fed Reg 38396 (1999) (amending 47 CFR part 20). This is the so-called “calling party pays” rule. The FCC is considering adopting this rule for calls that terminate on cellular networks. The argument is that people would be more likely to leave their cellular phones on if they did not have to pay terminating charges themselves, and that this would therefore increase the usage of cellular phones and make them stronger competitors with wireline phones.
26 Interestingly enough, in the internet backbone market, large firms have apparently not tried to take advantage of this possibility. Large internet backbone firms have generally agreed to exchange traffic with one another for free.
Although their specific model makes a number of special assumptions about the nature of the interconnection pricing rules that firms use, the chapter contains a fascinating discussion and set of conjectures about the possible results from allowing more general rules. I expect that building models to explore fully these sorts of questions will yield a fascinating body of research over the next few years. I also think that models built to explore similar sorts of questions in a regulated environment will yield equally interesting results.

IV. UNIVERSAL SERVICE

As discussed above, the cost of providing a loop to an end user can vary dramatically over different regions of the country, depending largely upon population density. Traditionally, the government has felt an obligation to guarantee that telephone service is available to residents of all areas of the country at the same low "affordable" rate. When local telephone service was provided by a single regulated monopoly, the government accomplished its goal quite simply by requiring the regulated monopoly to charge the same price to all end users regardless of the actual cost of serving them. By requiring the monopoly to charge averaged prices to all its end users, the government essentially levied implicit taxes on low cost end users and paid implicit subsidies to high cost end users.

With the advent of competitive entry into local markets, this scheme has come under increasing strains. In particular, it generates clear distortions in the pattern of entry. Ideally, we would like entry to occur in any region where the entrant can supply customers at a lower cost than the incumbent can. If the incumbent were allowed to charge prices equal to costs in different areas, this is the pattern of entry that would result. However, when the incumbent is required to charge the same averaged price in all regions that it serves, there will be incentives for "too much" entry in low cost, dense areas and incentives for "too little" entry in high cost, less dense areas. Even if one is not concerned with the problem of entry distortions, this system still needs to be changed. This is because competitive entry is eroding the source of funding for the government's implicit universal service program. Low cost end users are essentially able to avoid paying the implicit universal service tax by leaving the incumbent monopoly and instead purchasing services from entrants at lower unregulated prices that reflect the true costs of serving them.

Congress's answer to this dilemma in the Telecommunications Act of 1996 was to instruct the FCC to create a system of explicit universal service taxes and subsidies to replace the system of implicit
taxes and subsidies. The type of system that policymakers have in mind is conceptually quite clear and simple. The FCC would begin by allowing the regulated incumbent telephone companies to charge higher prices in less dense regions and lower prices in more dense regions to reflect better the actual costs of providing service. For high cost regions, the FCC would estimate the difference between the actual cost of providing telephone service in the region and some "benchmark" price that was thought to be "affordable." The difference between the estimated actual cost of service and the affordable rate would be made available as a per customer subsidy to any firm able to win customers in that region. For example, suppose that the government estimated that the true cost of providing service in an area was $40 per month per line and that the affordable rate was $24 per month. Then the government would pay a subsidy of $16 per month per line to any company (the incumbent or new entrants) that managed to win business in this region. Since the subsidy is made available to new entrants as well as the regulated incumbent, it is said to be "competitively neutral." Therefore, in theory at least, Congress could continue to provide subsidies for high cost users without distorting patterns of entry. I think this is a nice example where economic theory has helped Congress identify a method of accomplishing its desired equity goals, while minimizing efficiency losses.

There are at least two very interesting economic issues concerning such a scheme. The first is whether or not some type of competitive auction process could be used instead of a cost model to determine the size of subsidies paid in different areas. The advantage of this, of course, is that the government would not have to rely on a cost model (with all of the problems that any cost model has) to determine the correct amount of subsidy required in any area. Laffont and

---

27 See 47 USC § 254 (Supp 1996). The FCC has been very slow in making any progress towards implementing such a system. This is largely due to bickering among various interest groups that stand to win or lose large amounts of money depending upon how the outcome can be manipulated, and not because the essential policy itself is that complex or difficult to implement.

28 To keep the discussion conceptually clear, I will abstract away from the fact that the federal and state governments share jurisdiction for regulating telephone prices. See William P. Rogerson and Evan Kwerel, A Proposal for Universal Service and Access Reform, lecture notes for presentation at AEI Conference on Innovations in Universal Service for Telecommunications (May 17, 1999) (on file with author), for a detailed discussion that considers this complication as well as many others.

29 In policy circles, a figure of approximately $24 per month has come to be accepted as an "affordable" rate.

30 Any cost model constructed by regulators would necessarily contain errors and misestimates of costs. Furthermore, since regulators would presumably use actual incurred costs to estimate the required levels of subsidy, firms might have an incentive to purposely inflate their costs to increase the subsidy they receive.
Tirole discuss this subject extensively and describe work by Milgrom and a number of coauthors on this subject (pp 243–64). One of the interesting issues that arises when one considers universal service auctions is that, generally speaking, there may be advantages to allowing more than one carrier to win the right to provide services at subsidized rates in any given region. However, the only way to provide firms with an incentive to bid lower subsidy levels is to commit to allow only the lowest bidder (or some set of the lowest bidders) to be eligible to receive the subsidy. (Suppose regulators asked firms to bid on the per line subsidy they would require to serve an area and set the subsidy offered per line equal to the lowest bid. However, suppose that all carriers were eligible to receive the announced subsidy. Then no carrier would have an incentive to make a low bid because making the low bid provides no advantage.) Therefore, part of the auction mechanism design must be to determine how many winners are selected based on the bids that are received.

One proposal advanced by GTE that Laffont and Tirole describe would allow there to be a single winner if the lowest bid was “far enough” from all the other bids, but would allow multiple winners if there were a cluster of lowest bids “close enough together.” It turns out that economic theory has not generally considered the problem of having auctions determine not only a price but also a market structure. Laffont and Tirole provide a framework for analyzing this problem, relate it to the limited literature that exists, and provide some initial results. These initial results are relatively negative, in the sense that it turns out to be difficult to create auctions that guarantee that the correct market structure results. Nonetheless, given the problems with cost models, it seems likely that regulators will eventually turn to auction mechanisms for setting universal service subsidies, and determining how these auctions should be run is an interesting and important subject for future research.

The second issue about universal service is how any subsidy scheme ought to treat sunk costs (p 260). For example, it might be the case that, starting from scratch, the cheapest method of serving some less dense rural areas is to use a fixed wireless technology. However, if an existing incumbent has already sunk costs in creating a wireline system, then the incremental cost of simply continuing the wireline service may be much lower than the additional cost of installing an entirely new wireless network. One issue, then, is whether or not explicit


32 For example, to the extent that regulators cannot monitor quality perfectly, competition among multiple providers for subsidies will help guarantee quality of service. See pp 251–60 for a discussion of this and other reasons.
subsidy schemes might encourage new entry that is really not efficient given the nature of sunk costs in the industry. One can build models that suggest that an incumbent with large sunk costs will always offer lower prices to customers than a new entrant (since the incumbent will rationally understand that its sunk costs should be ignored for purposes of calculating its incremental profit from winning customers). However, this, in turn, raises the issue of whether or not such a subsidy scheme would essentially result in a taking, that is, whether or not such a subsidy scheme would essentially mean that the incumbent was not allowed to recover its investment costs. Therefore, the entire issue of how sunk and stranded investments are treated when explicit subsidies are used to attract entry needs further attention.

CONCLUSION

Rapid technological change and deregulation have fundamentally altered the economic landscape of the telecommunications industry, and Laffont and Tirole’s new book provides a lively up-to-the-minute perspective on how economic theory can help us understand these changes and formulate appropriate policy responses to deal with them.
If Taxpayers Can’t Be Fooled, Maybe Congress Can: A Public Choice Perspective on the Tax Transition Debate

Kyle D. Logue†


INTRODUCTION

It is a truism that changes in government policy can produce losers and winners. And nowhere is that truism more clearly revealed, or more closely studied, than in the context of tax rule changes. Consider the individual who makes an investment in real estate or business equipment or some sort of financial instrument (such as stocks or bonds) or even in human capital, and who does so in reliance on a special tax deduction or credit which applies to that investment. If the relied-upon tax preference is unexpectedly repealed, the investor may suffer what is sometimes called a “transition loss,” a decline in the value of the pretransition investment.

The story can just as easily go the other way: if an individual makes an investment before the enactment of the special deduction or credit (and oblivious to the possibility of its enactment), she may enjoy a windfall “transition gain” when the new rule is unexpectedly adopted. And both effects—transition losses and gains—can result from any number of tax rule changes (including changes in the tax base as well as in tax rates) as well as from non-tax-law policy changes of every variety, ranging from developments in common law tort doctrine to changes in environmental laws or regulations to changes in government spending programs.

That unexpected rule changes produce losers and winners is uncontroversial. Disagreement arises over the question of what, if anything, should be done about these transition effects. To put the question slightly differently, what should our “transition policy” be? Or put differently still, to what extent should the government deliberately try to alleviate the “retroactive” effects of rule changes?

† Professor of Law, University of Michigan Law School. I am grateful for comments received from Steve Croley, Saul Levmore, and Daniel Shaviro. All remaining errors are my own.

1 The concept of retroactivity can be defined in many different ways, and the definition
However you put it, the question has spawned a substantial literature, which can be understood as having developed in three general phases (pp 2–3). Phase one was the "old view" scholarship, which relied on general fairness concerns to argue against uncompensated retroactive rule changes. In phase two, applying the assumptions and methods of modern economics (and the policy goal of social welfare maximization), "new view" scholars offered a decidedly more favorable take on retroactivity, and hence less favorable view of transition relief. Phase three marked something of a revival of the old view (or a reaction to the new view) in which scholars mounted welfarist (social-welfare maximization) defenses of transition relief in certain contexts. In addition, and somewhat separately, a number of tax policy scholars have for years encouraged the use of retroactive taxation precisely because such taxes have the capacity to surprise investors and thus to raise revenue without distorting incentives.

And now Professor Daniel Shaviro, building on the new view approach but adding his own distinctive spin, has written what will likely be considered the definitive study of legal transitions from the utilitarian or welfarist perspective, and perhaps from any perspective. In When Rules Change: An Economic and Political Analysis of Transition Relief and Retroactivity, Shaviro takes the various strands of the existing literature on retroactivity and weaves them together, applying his unique combination of legal expertise, political pragmatism, and theoretical sophistication in public finance economics as well as political science. The result is a subtle, balanced, and scholarly treatise on transition relief and retroactivity that should serve as the starting point for all future research in the field.

In its stated objectives, the book is admirably ambitious. As Shaviro puts it in the introduction,

This book seeks to reframe the scholarly transition debate, and possibly even influence political practice, through two distinct contributions. The first is to advance economic understanding of

---

that a particular scholar has in mind can be crucial to his or her transition policy recommendation. For now, it is enough to understand that there are degrees of retroactivity; that is, a new rule can be applied more or less retroactively. See Saul Levmore, The Case for Retroactive Taxation, 22 J Legal Stud 265, 266–72 (1993). More precise definitions will be offered as the Review unfolds.

Shaviro explicitly adopts the "social welfare norm of utilitarianism, which seeks to maximize social welfare" (p 16). I will therefore use the term "welfarist" to describe Shaviro's approach to transition issues. For a precise account of Shaviro's normative assumptions, see pp 16–19. For a comprehensive defense of social welfare maximization as the appropriate norm for evaluating public policy, see Louis Kaplow and Steven Shavell, Principles of Fairness Versus Human Welfare: On the Evaluation of Legal Policy, John M. Olin Center for Law, Economics, and Business, Harvard Law School, Working Paper No 277 (Mar 2000), available online at <http://www.law.harvard.edu/programs/olin_center/> (visited May 12, 2000).
the transition issues posed by rule changes generally, albeit with particular reference to issues in the federal income tax. The second is to integrate the economics with a political science analysis that is distinctive to the present federal income tax system, in order to propose norms that should generally guide transition practice in this area (pp 1–2).

Are these ambitious goals achieved? Remarkably, to a large extent they are. For example, Shaviro advances the “economic understanding” of transition issues in a number of different ways, perhaps most significantly in his sustained argument against the use of surprise retroactive taxation as a nondistortionary source of revenue and in his introduction of a new taxonomy of rule changes.

However, it is with respect to the book’s second stated objective (“to integrate the economics with a political science analysis that is distinctive to the present federal income tax”) (p 2) that Shaviro’s most imaginative contributions come. For example, he develops and applies a public choice perspective to the tax transitions question, a perspective that includes a refreshingly realistic (if somewhat cynical) assessment of the tax lawmaking process. In addition, he makes bold and potentially controversial use of the “widespread consensus among tax policy thinkers” that the welfare-maximizing tax system is one that approaches the comprehensive tax base (“CTB”) ideal by doing away with all (or almost all) tax preferences (p 93).

Although Shaviro’s framework is cleverly conceived and defended, the “concrete policy payoff” it generates is something of a disappointment (p 14). After making a great show of the extent to which the new view transitions framework fails to incorporate the reality of political decisionmaking, Shaviro ends up adopting a version of the new view recommendations, albeit on different grounds (p 229). The reason for this anticlimactic conclusion, however, lies not with a lack of nuance or subtlety in Shaviro’s analysis. It is difficult to imagine a more nuanced and subtle analysis. Rather, the reason lies in the nature of the transition problem, especially in the tax-legislative context, as well as a curious characteristic of public-choice-oriented normative tax policy scholarship.

This Review will, in a broad sense, follow Shaviro’s characterization of the book’s objectives. Part I will summarize the existing economic framework for analyzing legal transitions and retroactivity issues (with major emphasis on tax transitions). Although Shaviro makes numerous interesting contributions to this framework, this Review will examine only the two most significant of these. That will be done in Part II. Then, Part III will focus on Shaviro’s principal innovation: his introduction of a thoroughgoing public choice perspective and CTB ideology to the tax transitions debate.
I. THE EXISTING ECONOMIC FRAMEWORK

A. Choosing the Optimal Transition Policy

What is a “transition policy” anyway? The idea is to exert some influence over the decisions of political actors, to affect how they decide to implement new rules. Shaviro provides the best articulation of this idea in his concept of a “constitutional norm.” As Shaviro points out, a constitutional transition norm need not be, and usually is not, legally binding on the political decisionmakers in question—the ones who are making the policy changes to which the norm will apply. But a transition norm, to be effective, must at least be a “stable long-term rule or aspiration” that is somewhat “binding,” at least in the sense of constraining or influencing the behavior of government actors (p 92). Thus, in the legislative context, the transition policy would be the prevailing norm, whether implicit or explicit, that tends to govern (“influence”) legislators’ decisions as to how transition losses and gains will be handled in the event of rule changes.

The next question is the one to which normative transition scholarship has largely been addressed: What should our transition norm be? And should it be different in different contexts? Before economic analysis was applied to that question, normative scholarship on the subject tended to focus on the “reasonable expectations” of investors and on the unfairness of making rule changes apply to those who had relied on prior law. Thus, the traditional view (again, sometimes called the old view) disapproved of retroactivity generally, especially in the legislative context. Economic analysis changed the terms of the discussion and reversed the presumption, introducing a generally more favorable view of retroactivity and less favorable view of transition relief.

By far, the most comprehensive statement of the existing (pre-Shaviro) economic framework for analyzing legal transitions can be

---


4 It is in this sense that transition scholars often speak of a government’s “binding commitment” or “precommitment” to follow a particular transition policy. See, for example, Kyle D. Logue, *Tax Transitions, Opportunistic Retroactivity, and the Benefits of Government Precommitment*, 94 Mich L Rev 1129, 1140, 1148, 1157 (1996).

5 One of the principal contributions of Shaviro’s book is to summarize the existing tax transition norms that appear to be currently in effect with respect to certain types of tax law changes (pp 98–110).

found in a 1986 article by Louis Kaplow.⁷ According to that framework, the choice of an efficient transition policy turns, at least initially, on two general questions: First, what transition norm would optimally allocate the risk of a change in government policy, that is, the risk of transition losses and transition gains? And, second, what transition norm would give investors the optimal incentives concerning whether and how much to rely on any government policy?

B. The New View: Focusing on Risk Incentives

Standard economic analysis generally assumes that individual investors dislike risk, all else being equal. How do we know this? Investors regularly buy insurance policies or diversify their stock portfolios as ways of reducing the variance of outcomes that they face. Just as investors appear to want to shed themselves of such market-created risks, presumably they would like to do the same with government-created risks, such as the risk of experiencing an unexpected transition loss or gain caused by a change in government policy. A number of scholars have observed this equivalence.⁸ And it is an observation that carries some weight, because it is also generally assumed that private investors should be responsible for finding their own ways to deal with market-created risks. Thus, if there is a general equivalence between market-created risk and government-created risk, the starting assumption should be the same: private parties should be expected to find ways (through insurance policies or portfolio diversification or whatever) of dealing with the risk of uncompensated changes in government policy that would accompany a retroactivity norm.⁹

But perhaps there is something different about transition risk. Perhaps there is some reason why, with respect to government-created risk, private markets working alone are not sufficient. The economic framework, primarily in Kaplow's article, looks at this question and concludes that, in fact, the market-risk/government-risk equivalence generally holds. Specifically, part of Kaplow's analysis is devoted to working through a number of reasons why we generally expect private insurance mechanisms to be more effective than government-provided insurance at dealing with the standard insurance inefficien-

---

⁷ See Kaplow, 99 Harv L Rev at 520–66 (cited in note 3). Although the Kaplow article provides the most comprehensive statement of the economic framework for analyzing legal transitions, the article does not reach definitive conclusions as to what the welfare-maximizing transition policy would be for any particular type of rule change.


cies caused by the asymmetrical information between insurer and insured; that is, the problems of moral hazard and adverse selection. And he explains why those problems are present with respect to transition risk as well, an assessment with which Shaviro appears to agree.

The implication of this analysis is that any risk-allocation defense of an anti-retroactivity norm—or a norm favoring transition relief—can largely be ignored. Since we can count on investors to deal with transition risk on their own, our transition norm need not be concerned with it. That conclusion seems to have gained considerable sway in the literature. Indeed, even Shaviro asserts that “the transition risk analysis often lacks significant policy implications” (p 42).

In addition to worrying about optimal risk allocation (at least long enough to dismiss it), the standard economic analysis of legal transitions is concerned with giving private parties the incentive to anticipate rule changes. And this concern is not dismissed. The basic argument is straightforward: a transition policy that credibly promises not to alleviate the retroactivity effects of rule changes will encourage investors to take into account the possibility of such changes—even to anticipate them. And this will generally be a good thing, if we assume that government policy, over time, will tend toward the maximization of social welfare.

Although the best example of the welfare-enhancing incentive effect of a retroactivity transition norm is that of a product ban (which, if made retroactive, gives manufacturers the incentive to pursue optimal safety for pretransition as well as post-transition production), it is

10 See Kaplow, 99 Harv L Rev at 540–42 (cited in note 3). Shaviro seems to agree with this conclusion at p 42.

11 For example, in none of his work on retroactivity does Saul Levmore discuss the risk-allocation question. He focuses instead exclusively on the incentive dimension of the transition issue, which is discussed in the next section. See Saul Levmore, Changes, Anticipations, and Reparations, 99 Colum L Rev 1657, 1661–86 (1999); Levmore, Retroactive Taxation at 340 (cited in note 3); Levmore, 22 J Legal Stud at 279–91 (cited in note 1). Furthermore, I am aware of no article or book that explores the government-risk question, as Kaplow poses it, in any level of detail. Therefore, Shaviro’s conclusion that the significance of the transition risk issue “has mainly been exaggerated” is puzzling (p 42).

12 Both Kaplow and Shaviro, however, mention a possible exception to this dismissal of transition risk concerns. Specifically, they note that, in situations in which extremely low-probability events are systematically underestimated by individuals, those individuals may tend to underinsure, with potentially drastic consequences if the event comes to pass. Kaplow, 99 Harv L Rev at 602–03 (cited in note 3); Shaviro (pp 33–42).

13 Put differently, if we believe that good policy changes will in the long run outweigh bad ones, a rule that generally induces behavioral changes by investors in anticipation of rule changes will likewise enhance social welfare. This idea was first emphasized in a classic tax article by Michael Graetz and was later built into the larger economic framework by Kaplow. See Graetz, 126 U Pa L Rev at 65–66 (cited in note 8); Kaplow, 99 Harv L Rev at 533 n 65 (cited in note 3) (quoting Graetz). When Shaviro refers to this idea, he too calls it the “new view,” and he cites Graetz as the originator (pp 43, 64, 101).
not the only one, not by a long shot. Indeed, part of what makes the incentives analysis so appealing, and controversial, is the strikingly broad range of rule changes to which it can be applied. Thus, the same welfare-enhancing incentive aspect of a retroactivity norm exists with respect to many types of rule changes, including, but not limited to: a judicial decision expanding (in an efficiency-enhancing way) tort liability for product-caused accidents; the repeal of a special tax preference provision that (for whatever reason) has proved to be a bad idea; or even the taking of private property for public use. In each of these cases, so long as one assumes that the government (including the legislative, executive, and judicial branches) is generally moving in the right (welfare-enhancing) direction over time, the new view incentives analysis would apply.\footnote{As will be discussed more fully in Part III.A below, Shaviro observes that there may be categories of rule changes for which the desirability of this anticipatory incentive effect will not hold.}

C. The Old View Revisited: Incentive Subsidies and Opportunistic Retroactivity

It was to this basic new view economic framework (and explicitly conditioned on its assumptions) that I made a small contribution in an article published in 1996.\footnote{See Logue, 94 Mich L Rev at 1129 (cited in note 4).} In that article, I offered an efficiency-based justification for providing transition relief with respect to a specific category of rules. My argument was that, if there is an \textit{identifiable} category of rules that the government uses as a means of inducing specific reliance by investors \textit{and} that we want to preserve for the government's future use (a category of rules that I call "incentive subsidies"), then adopting a transition policy that permits the imposition of transition losses upon repeal of such rules would come at a cost. The problem is this: investors, knowing of such a transition policy, would demand a larger subsidy than if the transition policy forbade retroactive repeal. I call this the "default-premium effect."\footnote{Id at 1140.} Indeed, at the limit, if the retroactivity norm were sufficiently extreme—if the norm essentially said that whatever incentive subsidy the government provides will be opportunistically confiscated ex post (through nominal retroactivity or outright theft if necessary), then inducing reliance on the government’s incentive subsidies would become impossible.

Given my adoption of the new view assumption of a benevolent, long run, social-welfare-maximizing government, one might ask why the concern for government opportunism. Can a benevolent government engage in theft? In fact, from a purely welfarist perspective, a
certain sort of government opportunism is entirely consistent with the benevolent-government assumption. According to the literature on the so-called "time-inconsistency" problem, developed by scholars within the rational-expectations school of macroeconomic theory, an otherwise social-welfare-maximizing government will behave opportunistically (will become a wealth expropriator) if (a) it cannot make binding promises not to do so, and (b) despite this fact, investors have no clue that the government would ever do such a dastardly thing. Under those assumptions, it can be shown that a policy plan of promising not to enact opportunistic expropriative retroactive rule changes, and then promptly turning around and doing just that, is in fact "optimal." 18

The lesson of rational-expectations theory, however, is that such a policy plan, though optimal if it were to succeed, cannot in fact possibly succeed. Investors are not that stupid. They will anticipate such opportunism, perhaps even before they have been burned once but certainly afterward. And the government's bait-and-switch plan will break down over time. In fact, the certainty of its long run breakdown will feed back into the original decision by the government to adopt the plan in the first place. In other words, once rational expectations are taken into account, such an optimal plan is said to be "time-inconsistent." 19 A nonstarter.

The default-premium effect that I have associated with the retroactive repeal of incentive subsidies is analogous to this time-inconsistency problem. Fearing expropriative retroactive repeal, investors will demand a default premium in their incentive subsidies, or may not be willing to rely on them at all. Thus, according to my analysis, the welfare-maximizing transition norm with respect to incentive subsidies may be one that forbids (or at least discourages) the government from opportunistically expropriating previously provided incentive subsidies. That norm, in other words, may include an implicit precommitment (again, a constitutional norm) that the legislature will provide some minimal level of transition relief should a given class of incentive subsidies be repealed.

Notice I said "may." Whether such a special transition norm for incentive subsidies would in fact be welfare enhancing depends on a

17 The classic article on the subject is Finn E. Kydland and Edward C. Prescott, Rules Rather than Discretion: The Inconsistency of Optimal Plans, 85 J Pol Econ 473, 474-75 (1977). Shaviro refers to this literature and the idea of opportunistic "rent extraction" briefly (p 89).
18 That such an outcome could conceivably be called "optimal" is what drives some scholars to doubt the welfarist approach.
19 See, for example, Stanley Fischer, Dynamic Inconsistency, Cooperation and the Benevolent Dissembling Government, 2 J Econ Dynamics & Control 93, 94, 98 (1980) (demonstrating the benefits of credible government precommitment to optimal tax policy plan).
series of difficult conceptual and empirical questions. For starters, how does this old view goal of encouraging taxpayer reliance on certain government actions balance out against the new view goal of encouraging investors to anticipate rule changes? The answer may depend on our confidence in the initial decision to use the incentive subsidy in the first place and in our assessment of the likely need to use such incentive subsidies in the future.²¹

Another objection that I have heard to my time-inconsistency/default-premium argument runs as follows: when the government needs to make a binding commitment to induce specific acts of reliance on the part of investors, it should do so through the use of a legal contract, and only through the use of a legal contract. That is, after all, what government contracts are for: to allow the government to make binding promises about how much it will pay, to whom, and for what. Incentive subsidies, especially in the form of tax preferences, should never be used in that way. Rather, the argument goes, with any type of policy tool besides a government contract, no government commitment should be made as to whether payment will be forthcoming. To put the argument in quintessentially new view terms, with respect to such incentive subsidies, it is better to put investors on notice that such tax preferences, like any other law, can change in any way imaginable, including nominally retroactively, and at any time so that investors adjust their expectations accordingly and take steps to anticipate changes.

In response to this objection, I can only fall back on my own intuition, since a definitive resolution of the issue must await empirical investigation. And my intuition is this: if we believe that the government should be able to make somewhat binding precommitments with respect to its contractual arrangements,²² it is likely that the government should also be able to make precommitments (albeit only implicit ones and only binding in the sense in which Shaviro talks of the...

---

²¹ For example, as will be discussed more fully in Part III.C below, Shaviro makes an argument that suggests that, for public choice reasons, we should adopt a transition policy that discourages the use of all (or almost all) tax preferences (p 98). If one takes that argument seriously, the threat of opportunistic retroactive repeal of such provisions might be just the ticket. How far Shaviro wants to push the argument is unclear.

²² How “binding” we want even government contracts to be is something of an open question, for the same sort of public choice reasons that (as Shaviro emphasizes) we should be cautious about “guaranteeing” tax preferences. For recent public choice oriented arguments that government contracts should be considered less binding on the government than private contracts are on private parties, see, for example, Gillian Hadfield, Of Sovereignty and Contract: Damages for Breach of Contract by Government, 8 S Cal Interdiscipl L J 467, 469 (1999); Daniel R. Fischel and Alan O. Sykes, Governmental Liability for Breach of Contract, 1 Am L & Econ Rev 313, 327-32 (1999). Even these scholars, however, assume that government contracts will be binding to some extent. See Hadfield, 8 S Cal Interdiscipl L J at 469 (proposing to analyze, not if, but how the government is bound).
binding effects of "constitutional" norms) in some other settings in which specific investor reliance is needed but entering into individual contracts with investors is not practical. For example, it would presumably be extremely difficult to use a series of direct government contracts to replicate the type of broad-based reliance effect that a tax credit or tax deduction can induce—that is, reliance on the part of many investors in many parts of the country with relatively little in the way of red tape and other transaction costs. Thus, if government contracts should be considered binding, perhaps incentive tax provisions should be treated similarly, as being accompanied by an implicit precommitment to provide transition protection in the event of repeal.

In some respects, Shaviro seems to agree with my default-premium concern. For example, in his discussion of the transition-risk analysis, with a concept he calls "transaction flexibility," Shaviro captures the core of my precommitment idea (pp 37-40). He suggests that, where individually negotiated government contracts are impractical, the government might "offer the same rule in a variety of 'flavors,' which differ only in the details of their transitional guarantees" (p 38).

D. When Surprising Investors Is the Whole Point

The final element of the existing economic framework for analyzing legal transitions is the idea of the ex post capital levy. Among economists, it is sometimes argued that certain types of retroactive taxes are desirable precisely because they are a surprise to investors—that is, precisely because they are not anticipated and have little or no effect on taxpayers' incentives. The argument derives from the optimal-tax literature, which defines an optimal tax as one that produces the least possible distortion of taxpayer or investor decisions—specifically, the least possible effect on the labor-leisure tradeoff.

According to the optimal-tax perspective, then, the best tax is one that the taxpayer can or will do nothing or very little to avoid. Such a tax is sometimes referred to as a lump-sum tax. The head tax would be
one example: each individual taxpayer pays the same amount of tax. Period. The problem is that a head tax is regarded as distributively unjust. So are many other quasi-optimal taxes, such as excise taxes on inelastically demanded consumer products. In part for that reason, we have chosen at least at the federal level to rely primarily on income and wage taxes. But taxing income or wages means also introducing some distortion, as people can choose to earn less, substituting leisure for work.

Enter the surprise retroactive tax. If the new tax is genuinely a surprise to taxpayers, then, by definition, they can do nothing to avoid it. And the bigger the surprise, the better. Plus, it is possible to make a retroactive tax somewhat progressive by making the tax base something that correlates with current income or wealth, say, past income or wealth. Thus, for one example, Saul Levmore notes the potential optimal-tax benefits of applying a tax rate increase or tax deduction repeal nominally retroactively, to tax years prior to the year of enactment.2

As was already discussed above in connection with the default-premium effect, the problem with trying to surprise taxpayers with a retroactive tax, from an incentives perspective, is that taxpayers are not easily fooled that way.27 If there is a transition policy in effect that permits the adoption of nominally retroactive income tax rate increases, or of ungrandfathered wealth-tax increases, such a policy will eventually be found out; the news will work its way down to the masses, leading to just the sort of anticipatory behavioral effects that optimal-tax theory tells us to avoid. Again, this is the lesson of the rational-expectations approach.30

II. SHAVIRO'S SPIN ON THE EXISTING ECONOMIC FRAMEWORK

A. Taking Rational Expectations Seriously: Surprise Taxation Can Never Work

One of Shaviro’s most important contributions to the existing economic framework is his comprehensive and sustained attack on the optimal-tax justification for surprise retroactive taxes. In Shaviro’s view, the findings of the rational-expectations literature, mentioned above, justify a somewhat extreme and potentially controversial con-

28 See Levmore, 22 J Legal Stud at 274 (cited in note 1).
29 Levmore discusses this possibility. See id at 274, 276. One of his responses is that even a little surprise is better than none. See id. My retort was this: “True enough. However, having been hit once with a ‘bolt from the blue,’ and knowing that Congress will be tempted to try the same trick again, taxpayers will begin to watch the sky for the next one.” Logue, 94 Mich L Rev at 1172 n 140 (cited in note 4).
30 See Logue, 94 Mich L Rev at 1172 n 140 (cited in note 4) (“This is the quintessential example of the time-consistency problem.”).
clusion: that the use of surprise taxes can never have long-run efficiency benefits.

In developing this position, Shaviro points out that rational-expectations theory does not claim that investors will never be surprised by rule changes. Of course, they will. Rather, the claim is somewhat more modest: given that individuals will learn from experience, we should assume that they will not be fooled by a program of surprise macroeconomic policy changes, at least not in a systematic way. That is, even if some investors are fooled, there is no obvious direction that the errors will tend to take. Some will underestimate the likelihood of the policy change; others will overestimate it. In any event, the under- and overestimation errors will cancel each other out, thereby canceling out the lump-sum benefits of a long-term program of surprise policy changes.

From these seemingly modest premises, Shaviro concludes that the analysis of transition issues should proceed on the assumption that it is impossible ever to fool investors systematically in the long run. Moreover, although Shaviro acknowledges the possibility that cognitive biases on the part of individual investors (such as systematic underestimation of retroactive policy changes) could in theory undermine his strong rational-expectations presumption (pp 22-25), he ultimately dismisses that possibility.

Thus, we are left with Shaviro's operating assumption: that a program of surprise retroactive taxation cannot in the long run have optimal-tax benefits. But what about something short of a "program" of surprise taxes? What about the possibility of using this policy tool only

---

31 "Rational expectations suggests that, if one is thinking about the long run, the incentive effects of retroactive rule changes can fruitfully be analyzed by ignoring their retroactivity and treating them instead as if they were preannounced or otherwise fully anticipated" (p 22).
32 Where Shaviro reviews conclusions from the fields of cognitive psychology and behavioral or "quasi-rational economics."
33 Specifically, he argues that, until advocates of surprise retroactive taxation can offer empirical evidence that there is a specific cognitive bias that systematically distorts investors' perceptions of the likelihood of future retroactive taxes, we should assume that surprise retroactive taxation cannot be used successfully as a lump sum tax (p 25). Of course, determining which side in a debate bears the burden of empirical proof will often determine the winner of the debate. In Shaviro's defense, however, he has offered something more. He has presented the conclusions of the rational-expectations literature, a literature, according to Shaviro, whose empirical findings strongly confirm its hypothesis: that investors cannot generally be fooled systematically by a program of surprise policy changes. Although Shaviro notes that the "full-fledged rational expectations approach...has not entirely carried the day" in the macroeconomics literature, he implies that the remaining nonbelievers are merely suffering from an insufficiently long-run perspective (p 21). ("Rational expectations still holds over a longer term, however" (p 21).) Whether he is right about that conclusion I do not know.
34 This conclusion seems somewhat in tension with his willingness, elsewhere in the book, to entertain the possibility that individual investors may systematically underestimate the "risk" of uncompensated policy changes (p 40).
on rare, exceptional occasions—such as when we finally (once and for all) manage to shift from the current income-based tax to something closer to a consumption-based tax? Maybe that would be a good occasion for a onetime-only wealth tax, a capital levy accompanied by a promise never to do anything like that again, at least not anytime soon.

Shaviro heaps scorn on this idea, and perhaps rightly so. The problem with the "just this once" approach, of course, is that the government’s very use of the retroactive tax may well undermine the credibility of the promise not to repeat the trick. To the contrary, it might have the perverse effect of alerting investors to the possibility of retroactive taxation, a possibility that many of them may never have considered before." Thus, in Shaviro’s words: "Today’s ‘surprise party’ may therefore prompt tomorrow’s ‘surprise non-party,’ resulting in compensating errors in anticipation of retroactive change that prevent the government from using surprise to its advantage on balance" (p 22).

He may be right about that as well. But then again, he may not. To the best of my knowledge, the precise findings of the rational-expectations literature do not speak directly to the question at hand: that is, would it be possible for Congress, at the time it enacts a surprise retroactive tax, to make a credible commitment not to do so again for a sufficiently long period of time that the incentive effect would be trivial? In fact, we do not have any empirical evidence, one way or the other, on that question. Given this situation of profound uncertainty, what should our conclusion be?

If Shaviro’s position on surprise taxation represents one extreme answer to that question, presumably the other extreme—the view that surprise taxes could be used effectively on a regular basis—has few adherents. An interesting question is what about the middle ground? That is, given the absence of empirical studies demonstrating specific cognitive biases on the part of investors with respect to retroactive taxes (and given that such studies may not be forthcoming anytime soon), should we consider an "only occasionally" or "just this once" retroactive tax strategy in the meantime? I, for one, find the possibility of a just-this-once retroactive tax somewhat attractive, although I obviously see the dangers that such a strategy would entail.37

36 See id. "What’s more, taxpayers may be less likely to believe Congress’s commitment if it comes on the heels of the first surprise retroactive tax.” Id. “Moreover, given the salience of the first retroactive tax, taxpayers may have a tendency to overestimate the chance of its reoccurrence.” Id at 1172 n 140.
37 As I noted above, perhaps the just-this-once promise would be somewhat credible if the retroactive tax is enacted as part of a presumably onetime-only shift from an income tax to a
B. Developing the Policy Change/Accounting Change Distinction

The other major contribution that Shaviro makes to the existing transitions framework is the addition of a taxonomy for classifying rule changes. The most important aspect of this taxonomy—the one that he uses throughout the remainder of the book—is the distinction between the “policy content” and the “accounting content” of rules. According to Shaviro’s taxonomy, every tax rule can be divided into its policy content (the efficiency and distributional aspect of the rule) and its accounting content (the details of the rule’s implementation “that could in principle be changed without affecting its policy content”) (pp 53–54). Shaviro uses these ideas to distinguish between “policy change” retroactivity and “accounting change” retroactivity.

Policy change retroactivity arises in situations in which the anticipatory (new view) incentive effects of retroactivity would be desirable if and only if the rule change itself is desirable. Thus, the merits of retroactivity in such cases depend on the “steady-state merits” of the new rule itself (p 48). By contrast, accounting change retroactivity arises in situations in which anticipatory incentive effects would be bad, and should be discouraged, even if the rule change itself is good—that is, regardless of the new rule’s steady-state merits (p 50).

To exemplify policy change retroactivity, Shaviro offers the standard product ban example. In addition, he mentions the example of the repeal of the tax exemption for municipal bond interest. In both cases, whether making the repeal retroactive (applicable to pretransition investments) will have beneficial incentive effects depends on the efficiency of the incentive effects that the new rules themselves are expected to have. Thus, if we think the product ban and the elimination of the bond exemption will improve incentives—that is, if we think the government has made social-welfare-increasing decisions by adopting those rules, then making those decisions retroactive will induce efficient behavioral responses for pretransition investments as well. And if we think the rule changes are bad, retroactivity would likewise be bad (p 48).

The starkest example of accounting change retroactivity is the hypothetical shift from a tax-accounting year that is based on the calendar year (ending, of course, on December 31) to one that is based on a fiscal year (ending on June 30). Absent some sort of deliberate transition relief, such as the introduction of a special onetime short ac-

38 And here we must again be careful in defining retroactivity. In general, when Shaviro uses the term retroactivity—such as with his idea of a “policy change retroactive tax”—he means something like nominal prospectivity without grandfathering. When he means nominal retroactivity, which he thinks is and should be discouraged, he uses that term explicitly, and so will I.
counting year, some income may be taxed exactly twice in the transition year, and some may escape tax entirely (p 50). The reason this example is so powerful, and why Shaviro’s ultimate conclusion that an antiretroactivity norm should apply to such cases, is that there is no apparent (or even conceivable) policy justification for such a random result. That is, one cannot easily imagine what beneficial anticipatory behavior a retroactivity norm could have in such a case.

The utility of the policy change/accounting change distinction, however, will depend ultimately not on its application in the obvious case, such as the accounting rule change, but in the more difficult cases. To take perhaps the most important example in Shaviro’s book, consider the accounting change and policy change aspects of proposals to shift from our current income tax to a more consumption-based tax regime (ch 9). Shaviro rightly, and skillfully, elucidates a number of ways that several leading consumption-tax proposals conflate an array of tax-rule changes that can usefully be disaggregated into various policy change aspects and accounting change aspects (pp 181–95).

For example, Shaviro notes that the essential difference between an income tax and a consumption tax is the treatment of the “return to waiting.” Thus, if we imagine two households with the same wage flows over time, “the income tax imposes a higher tax burden on the later than the sooner consumer, whereas the consumption tax imposes the same tax burden, in present value terms, on each” (p 171). Thus, the policy aspect of shifting from an income tax to a consumption tax would amount to a decision to exempt the return to waiting. In Shaviro’s taxonomy, therefore, whether that particular change should be made retroactive (that is, whether retroactivity would have desirable anticipatory incentive effects) depends on the desirability of the consumption tax idea itself, the idea of exempting the return to waiting. (For example, if we think that taxpayers are saving too little, and that a consumption tax will tend to encourage more savings, making the move to a consumption tax retroactive might induce desirable anticipatory savings decisions.)

Most major consumption tax proposals, however, also include a number of accounting changes as well, which are not clearly identified as such (p 184). Admirably picking on someone his own size, Shaviro cites as an example of this problem the scholarship of eminent public finance economists Alan Auerbach and Laurence Kotlikoff, whose work in tax policy and tax-transition policy fields is, truly, “in many ways path breaking” (p 173). According to Shaviro, Auerbach and Kotlikoff define a consumption tax as “a wage tax plus a onetime...
wealth tax upon its introduction” (p 173). Shaviro then rightly observes that this wealth tax really derives from an accounting change not a policy change, and he suggests further (again rightly) that understanding of the issues would be enhanced if such a wealth tax were not “built into” (or, made to seem an essential part of) the definition of a consumption tax. Thus, Shaviro suggests that the policy change/accounting change distinction, if nothing else, should be kept in mind for the sake of conceptual clarity and rigor.

In addition, however, Shaviro wants to make a somewhat bolder use of his policy change/accounting change idea. He wants to argue that, just as retroactivity can produce only undesirable anticipatory behavior in the paradigm cases of accounting changes, making consumption-tax-related accounting changes retroactive can likewise only be a bad idea. Ergo, the implication is that the onetime wealth tax in the Auerbach and Kotlikoff consumption tax proposal (and in many other such proposals) is a bad idea. In one sense, a onetime wealth tax (whenever it is imposed) has obvious similarities to the accounting-year change mentioned above: taxpayers who happen to be wealthy at the time of the Auerbach and Kotlikoff transition get hammered; whereas, those whose wealth will be accumulated primarily after the transition enjoy a windfall gain. That distinction seems pretty arbitrary.

In another sense, however, a onetime tax on all wealth existing at the time of the shift to a consumption tax looks somewhat less arbitrary than the failure to provide transition relief in the accounting-year situation. That is, a onetime tax on wealth could conceivably be structured so as to be roughly consistent with our notions of distributive justice or progressivity. The same could not as easily be said about the tax that would be produced by a retroactive shift from an annual to a fiscal accounting year. Moreover, if taxpayers could be convinced that this tax truly is a onetime-only thing (just as the dramatic shift from an income to a consumption tax could plausibly be sold as a onetime-only move), it might have the optimal-tax benefits discussed in Part I.D above. That possibility, of course, provides a plausible justification for Auerbach and Kotlikoff’s approach of link-

41 The example of an accounting change wealth tax that Shaviro most often refers to as being “built into” consumption-tax proposals, and the one that gets the most attention in the literature, is the onetime elimination of basis (“the tax system’s record of outlays that qualify for eventual cost recovery but have not yet been deducted”) upon the adoption of the consumption tax (p 10). As Shaviro points out, a basis “wipeout” could be done entirely within the income tax system and without a shift to a consumption tax (pp 10, 184).
42 Especially if accompanied by a shift to a more steeply progressive wage tax, to improve progressivity with respect to post-transition wealth accumulation.
ing the wealth tax to the shift to a consumption tax. As we have seen, however, these arguments do not fly with Shaviro. He just does not believe that such surprise retroactive taxes can work, even if “hidden” in the form of an ostensibly (though not actually) necessary element of a massive “one time” tax policy change. 

III. SHAVIRO’S PRIMARY INNOVATIONS: INTRODUCING POLITICAL REALITY AND TAX POLICY EXPERT OPINION

A. Attacking the Benevolent-Government Assumption

Recall from Part I.C the assumption, which underlies the new view incentives analysis, that the long run path of government decisionmaking is upward, towards the overall improvement of social welfare. Not that every rule change has to be for the better; there will be mistakes and corrections and even mistaken corrections. But the general trend will be in the right direction. There will be progress. Thus, at its core, the new view is optimistic about the capacity of individuals, collectively through their government, to identify the common good and to achieve it.

Shaviro rains on this parade. Drawing on an influential branch of political science, he summarizes three “core defects of public political choice that discourage adopting optimistic new view assumptions about government decisions” (p 66). These defects, which he terms the problems of aggregation, organization, and information, are well known. The problem of aggregation includes various voting paradoxes (including Arrow’s theorem and the cycling phenomenon) that call into question the possibility of aggregating individual preferences in a coherent way. The problem of organization includes interest group theory, which sees government as a means of transferring wealth from the disorganized many to the well organized few. And the problem of information simply recognizes that decisions tend to be worse when made in ignorance and that, for all sorts of reasons (including defects one and two), many public political decisions are made in ignorance (pp 67–73). Given these defects in the political decisionmaking pro-

---

43 Again, as I suggested above, Shaviro may be right about that conclusion, but whether he is depends on the merits of his anti-surprise-tax arguments (discussed in Part II.A), and not on the accounting change/policy change distinction.

44 It is important to note that, although Kaplow, Graetz, and some of the other architects of the existing economic/welfarist framework have used the above-stated benevolent-government assumption in their analyses, those scholars are by no means unaware of the sort of political concerns that Shaviro makes the focus of his book. For one example, in Kaplow’s article, there is a long section on what he terms “institutional considerations,” which include many of the public choice sort of concerns that Shaviro raises. See Kaplow, 99 Harv L Rev at 566–76 (cited in note 3). In addition, Levmore’s article on retroactive taxation is, at its core, a public choice take on the surprise-retroactive-tax question. Still, Shaviro’s book represents the most thorough study to date of the implications of public choice theory for a wide variety of transition issues.
cess, "the mere fact of a policy change's adoption does not provide strong evidence of whether it [and therefore the incentive effect of making it retroactive] is good or bad" (p 73).

What follows from this observation? What is its implication for the choice of a welfare-maximizing transitions policy? Shaviro admits that, in the abstract, the answer is unclear. It is only in particular applications that the "policy payoff" will come (p 81). Most of Shaviro's book is devoted to working out the implications of these defects in the tax lawmaking context.

B. The Reality of the Tax Lawmaking Process: Worse than Sausage

If you think the legislative process in general is screwed up, wait until you get a load of the tax lawmaking process. It is especially ugly, says Shaviro. In support of this claim, he asserts that "[f]or decades, there has been a widespread consensus among those who closely follow tax policy that the legislative process in this area is seriously flawed, even by the standards of collective political choice generally" (p 86). He then goes on to argue that the tax lawmaking process exhibits the problems of organization and information (mentioned above) to a substantially greater degree than do other types of lawmaking. Again, this assertion is supported by the widely held view that "tax politics tends to be dominated by interest groups that seek favors for themselves and that, through a norm of logrolling, almost never oppose favors for each other" (pp 86–87) (citations omitted).

It is not that Shaviro claims that tax lawmaking is unique in having these pathologies, only that it is one of several problem areas, one of the areas of "distributive" politics, that is plagued by "particularly destructive interest group influence" (p 87). Other examples of distributive politics that he mentions are tariff legislation and pork-barrel spending (id). Whether Shaviro is right that tax lawmaking is exceptional in this regard, or whether he is being unjustifiably tax-centric in his public choice critique of the legislative process, is a question that is beyond the scope of this Review. However, as will become clear below, Shaviro's tax-exceptionalism arguments are central to his reasoning and ultimately to his transition policy recommendations.

C. The First Thing We Do, Let's Kill All the Tax Preferences

Shaviro believes that the exceptional flaws in the tax policymaking process (that is, the chronic and severe problems of organization and information) have produced a divergence between the tax system that every tax policy expert worth his or her salt would recommend and the tax system we actually have. Again invoking "a widespread consensus among tax policy thinkers," Shaviro urges that "society
would benefit from the use of a more comprehensive tax base ("CTB") of either the income tax or consumption tax genre" (p 93).  He then explains why the choice between the two CTBs—income or consumption—turns out not to be a momentous decision (with everything coming down to the question whether or not to have tax burdens turn on a taxpayer’s preference for current or future consumption). Rather, the important tax policy question is whether to allow divergences from whichever tax base is chosen, whether to allow the use of special tax preferences—credits, deductions, exclusions—that favor some same-period investments over others. And for died-in-the-wool CTBers such as Shaviro that question is a no brainer: the answer is a resounding no. Keep the tax laws clean of tax preference provisions, henceforth and forever.

Here’s the rub: the actual income tax code happens to be (and has always been) loaded with tax preferences. That fact is, unsurprisingly, a source of vexation for CTBers. Again, the problem from Shaviro’s perspective is that any reasonably well-organized interest group can, with a little perseverance, persuade Congress to enact that group’s pet tax subsidy. After all, those who stand to lose from the enactment of such tax preferences—that is, the taxpayers in general—are unorganized and therefore “little heard” (p 87). What’s more (and this will become important below), Shaviro argues that taxpayers, because of a manifestation of the endowment effect, do not seem to understand the equivalence of a dollar that the government spends directly and a dollar spent indirectly via tax revenues foregone because of some special “tax expenditure” provision (pp 87–88). The result is that, although the general good would be served by the adoption of a clean, comprehensive, tax-subsidy-free income or consumption tax (or, put differently, by the wholesale and permanent repeal of all existing tax preference provisions), we just never seem to get there.

And here is where Shaviro’s tax exceptionalism argument should be emphasized again: he does not contend that Congress should never intervene in the economy to encourage one type of investment over another. Indeed, he expressly rejects that conclusion. But he is clearly arguing that Congress should not, for the reasons just stated, intervene through the tax system to encourage one type of investment over another. The conclusion of this reasoning presumably is that Shaviro

45 Although he postpones “for another day” a full-fledged defense of the consensus view of the CTB idea, Shaviro clearly agrees with it and expressly adopts it for the purpose of his tax transitions analysis (p 94).
46 “Within the present income tax, the CTB norm suggests that tax preferences and dispreferences that discriminate between alternative forms of same-period consumption or investment are generally undesirable” (p 98).
47 One might even say a source of “sour grapes” (p 86).
should favor any transition policy or practice that makes any tax preference less attractive to taxpayers. As I will suggest below, however, Shaviro seems to pull back from this conclusion at the last minute, for reasons about which I can only speculate.

The next several steps in Shaviro's analysis are devilishly clever and exceptionally (almost exceedingly) subtle. In effect, Shaviro sets out to design a tax transition policy that will maximize the beneficial effects of shifts in the direction of the CTB norm and minimize the harmful effects of movements in the other direction. In other words, he argues for a transition policy that will let the air out of tax preferences once and for all, even though (and, in some sense, especially because) Congress really likes using tax preferences. Therefore, in one sense, Shaviro wants to develop a constitutional tax transition norm that would ratify the current consensus among tax policy experts about what our tax code should look like, despite Congress's apparent disagreement with that consensus. Obviously, Shaviro has set no small task for himself.

He then starts out by making clear what his first-best tax transition policy would be. Recall that under Shaviro's framework, with respect to the type of rule change that we are now discussing—a policy change as contrasted with an accounting change—the choice of the optimal transition norm depends on the desirability of the new rule itself. Thus, a decision that the CTB norm is the best steady-state tax policy has important implications, in theory, for the choice of an optimal tax transitions policy. Specifically, Shaviro's analysis to this point would suggest that the first-best constitutional tax transition norm would hold as follows: policy changes that represent movements in the direction of the CTB ideal (for example, any curtailment, including repeal, of a tax preference provision) will be made retroactive; whereas, any policy changes that move away from the CTB ideal (for example, the enactment or expansion of such provisions) will not. In other words, transition relief should be provided in the latter cases, but not in the former (p 99).

As Shaviro concludes, however, such a transition policy would be politically unrealistic. Consider the general message that such a norm would send to lawmakers: if you (Congresspersons) decide to enact a good change in the tax law (say, repealing a pesky tax preference), please make that change retroactive. That is precisely the sort of

---

48 In keeping with the subtlety and balance exhibited throughout the book, in the applications of this approach to specific cases, Shaviro notes that the CTB norm "does not provide the correct answer all the time" (p 132).
49 See note 38 and accompanying text.
50 Again, any accounting changes, in his view, should never be made retroactive, for the reasons already discussed in Part II.B.
change that we (CTBers) would like to see taxpayers trying to anticipate. But if you decide to enact a bad rule change (for example, if you enact, against our advice, yet another tax preference), well, make that change apply prospectively only, to minimize the harm done.

Such a transition norm, Shaviro recognizes, would be neither stable nor binding. Indeed, it would be entirely self-defeating. It is almost impossible to imagine that the legislative majority that voted to approve a new tax subsidy would take kindly to the idea of intentionally blunting the effect of its new rule, merely because it is seen as a bad idea in the eyes of the tax policy cognoscenti (p 99).”

D. Offsetting the Transition Gain/Loss Asymmetry

By acknowledging that his idealized tax transition norm is a non-starter, however, Shaviro does not mean to abandon his hope of using tax transition policy to maximize the benefits of the CTB norm and minimize the harm caused by departures from it. It just means that he must take a more indirect route to achieving his goal. And here is the path he takes: first, he recognizes that he must either go with an old view approach (generally opposing retroactivity) or go with a new view approach (generally supporting retroactivity), even though he disagrees with key assumptions underlying both approaches. As a practical matter, he concludes, that is the best that can be done (pp 98–110).

Therefore, the question becomes which approach—old view or new view—would do the most to further the CTB norm. And Shaviro answers that question by reference to an observation regarding what he sees as the existing tax transition norm, an observation that is as important to his eventual policy recommendations as is his embrace of the CTB norm itself: he argues that there is currently a bias in Congress’s tax transition constitutional norm in favor of compensating transition losers (if the losses result from the repeal of a tax preference) and against taxing away or otherwise eliminating transition gains (if the gains result from the enactment of a tax preference) (pp 88–91).

51 Kaplow, in justifying his benevolent-government assumption, noticed this peculiarity with the alternative of making transition policy turn on the steady-state merits of particular rule changes. See Kaplow, 99 Harv L Rev at 566–76 (cited in note 3).
52 See id.
53 Shaviro also discusses other types of tax rule changes, such as changes in marginal tax rates. Because rate changes are often used to alter the distributional consequences of the tax laws, the choice of a transition policy (effective dates and such) will turn on the same distribution-efficiency tradeoffs that motivate the rate change in the first place (p 120). See also Logue, 94 Mich L Rev at 1170 (cited in note 4); Kaplow, 99 Harv L Rev at 519 (cited in note 3); Graetz, 126 U Pa L Rev at 82–83 (cited in note 8). The same can be said of changes in the social security laws, the subject of Shaviro’s chapter 10.
How do we know this is so? Shaviro offers both theoretical and empirical support. On the theory side, he argues that interest group theory would strongly suggest such a conclusion (p 89). The intuition, which seems entirely plausible, is that the political forces that would likely arise to oppose the imposition of a specific, identifiable transition loss would tend to be stronger than forces that would oppose the allowance of a transition gain.\footnote{With transition losses, the costs are concentrated on a relatively small and identifiable group; whereas, with transition gains, the costs, in terms of lost productive efficiency, are spread over all taxpayers.}

On the empirical side, he offers two types of evidence. First, he notes that everyone who writes about transition issues, including the academics, tends to focus on transition “relief” for transition losses, even though he or she may acknowledge the gain/loss symmetry of transition issues.\footnote{See, even I did it. I mentioned the gain/loss symmetry in the introduction and then focused on transition losses throughout most of this Review. Shaviro must be onto something.} So there must be a bias. Second, and more important, he looks at actual congressional practice and provides a couple of examples that support his gain/loss asymmetry hypothesis. For one, he observes that the so-called “binding contract” grandfather rule tends to be applied only when rule changes are made to the taxpayers’ detriment (p 90).\footnote{This special grandfather rule holds that “property placed in service after a new provision’s effective date will nonetheless be taxed under the provision that it replaced, so long as the taxpayer had a binding contract to construct or acquire the property by a specified date” (p 90).} Thus, it is applied when depreciation deductions are curtailed, but not when they are expanded (id). And, of course, “rifle-shot” transition rules—which apply to only one or a few taxpayers—are always used only to prevent transition losses, never to take away transition gains. Shaviro concludes that, although these two “asymmetric effective date rules represent a discrete practice of limited general significance, they arguably bespeak an important broader attitude” (id).

Whether Shaviro is right about the “general significance” of these two examples, I don’t know. But for purposes of this Review, I will assume he is. Therefore, let us assume that, owing to flaws in the tax lawmaking process, the existing tax transition policy embodies a general bias in favor of compensating transition losses and against taxing transition gains. So what? The answer lies in the CTB norm again. From the perspective of a card-carrying CTBer such as Shaviro, a norm that calls for compensating those who suffer transition losses when a tax preference is repealed but that also allows transition gains for those who anticipate new tax preferences cuts in exactly the wrong direction. That is, the anticipatory incentive effects of such a transition policy could not be worse.
Imagine the message that norm sends to taxpayers: Thinking about where to invest your resources? Wondering whether to rely on those enticing tax preferences, given the possibility of their repeal? Don’t worry. Go ahead and rely on them. Even if they are repealed, your reliance interest will be protected by special transition rules, so long as you have a binding contract dated prior to the rule change. What’s more, and get this, if you can somehow predict what new tax preferences Congress will cook up next, go ahead and invest there too, because (thanks to this wonderful asymmetry in the way tax law changes are implemented) the new tax preference will be applied retroactively to your pretransition investment. You can’t lose.

This message, in effect, produces a transition subsidy for tax preferences and against the CTB norm. Shaviro, therefore, wants to change the message, and he wants to do so by advocating a new tax transition norm, which he hopes will come to replace the existing, biased one. Given that his first-best transition norm is not politically practical, and given that the norm against taxation of transition gains is unlikely ever to change, Shaviro chooses the new view alternative (which encourages policy change retroactivity), although, again, not for new view reasons but for public choice reasons.

E. A Problem for Public Choice Normative Scholarship

What is especially strange about Shaviro’s transition policy conclusion is how he expects to pitch his proposal to Congress. He recognizes that the “real reason for advocating this norm—that in practice it may reduce Congress’s predilection, through its transition policy, to favor tax preference expansion over curtailment—might lack broad appeal when stated forthrightly” (p 101). That is an understatement. Such a transition norm would have about as much chance of wooing legislators as would his first-best transition policy (that is, make your good laws retroactive, but not your bad ones).

In response to this concern, Shaviro suggests the possibility of arguing for the new view norm he proposes using a new view (“Graetzean”) justification rather than his own (“Shaviroesque”): defense. That is, he suggests arguing dishonestly, or least not “forthrightly,” using terms that “are likely to sound more pleasantly in Congressional ears” (id). This suggestion, which I assume he means to be taken seriously, puts in stark relief one of the difficulties presented by public-choice-oriented normative scholarship: what is a public choice tax policy scholar to do, when the audience he is trying to influence, whom he is trying to get to adopt his recommendations, is composed of the same group of people that his own theories suggest are not collectively capable of pursuing the public interest? The lesson of Shaviro’s analysis seems to be that, first, you figure out what the right policy is
(taking everything into account, including public choice concerns) and then you find a non-public-choice argument (which you may not really believe) to sell your recommendation to the policymaker. As far as I am aware, this approach itself is another innovation of Shaviro’s book and perhaps the most important one, because it reveals the most extreme implications for scholars of normative public choice analysis, at least in the tax context.

F. The Anti-Nominal-Retroactivity Norm

In addition to concluding that transition relief should be provided for tax law “accounting changes” but not for tax law “policy changes,” Shaviro seems to reach one other general conclusion regarding tax transition policy: that the existing constitutional norm discouraging “nominally retroactive” tax law changes should be retained.7 This conclusion is puzzling. Not because it is a bad idea; indeed, the anti-nominal-retroactivity (“ANR”) norm strikes me as eminently sensible, at least when applied to certain types of tax incentive provisions, such as investment credits (for example, the research and development credit). Were it not for an implicit, but strong, ANR norm, our investment credit provisions would have little effect.8

What is puzzling is Shaviro’s endorsement of the ANR norm, at least with respect to tax preferences. Given his apparent view that tax preferences are generally bad, I would have expected him to go out on a limb and favor nominally retroactive repeal of such provisions—the same way that he seems to endorse nominal retroactivity with respect to new product bans or the efficient expansion of tort liability doctrine. But for reasons that remain a little mysterious, he pulls back from that conclusion.

At first, I thought the mystery was solved by Shaviro’s statement that there was a well entrenched anti-nominal-retroactivity norm that simply had to be acknowledged and recognized as a “brute fact” of political life (p 109). That conclusion suggested no endorsement of the norm, just a begrudging acceptance of it. But then Shaviro goes on to offer a justification of the ANR norm, offering a rationale that is especially interesting: he defends the ANR norm as a means of limiting the potential reach of government opportunism and of enabling “Congress to be able to make credible binding commitments (and not

---

7 Again, the precise definition of the term “nominally retroactive” is hard to pin down. One plausible definition in the tax context is the application of a tax rule change to a prior year’s tax return (or to income earned or expenses incurred prior to the rule change).

8 When it comes to nominally retroactive tax rate increases, however, I am inclined toward Levmore’s position, which seems more open to the occasional use of retroactive rate increases on optimal-tax normative grounds, but which acknowledges the public choice impediments to such an approach. See Levmore, 22 J Legal Stud at 265 (cited in note 1).
just by the cumbersome practice of executing contracts)” (id). Thus, Shaviro harkens back to his discussion of “transactional flexibility” and “flavors” of rules, discussed in Part I.C above, apparently adopting a position that is consistent with my anti-opportunism arguments against unfettered retroactivity.

How does all of this fit together? On one hand, Shaviro seems to oppose the use of tax preferences, of any sort, ever. That sentiment motivates his endorsement of the CTB norm and, ultimately, his new view approach to transition losses. On the other hand, he favors the existing ANR norm, which provides some transition protection to those who invest in reliance on tax preferences. Perhaps the best way to reconcile these positions is simply to say that Shaviro is splitting the difference between two competing concerns. Although his main goal is to discourage the use of tax preferences, because of the public choice problem of the few exploiting the many, he also wants to make some concession to the opposite concern, the problem of the many exploiting the few. And the ANR norm balances those concerns, in an admittedly imprecise but arguably tolerable way (pp 108–10).

**CONCLUSION**

Let me conclude with a few words on the “widespread consensus among tax policy thinkers” (p 93) in favor of the CTB norm and against the use of tax preferences, since, as Shaviro’s book makes clear, one’s views on that consensus are important to the choice of an optimal tax transition policy. Much has been written on the CTB norm over the years, as staunch CTBers have called into question the use of “tax expenditures” and anti-CTBers have questioned the coherence and validity of the CTB ideal. In my view, although the anti-CTBers have had the better of the debate, the public choice concerns that Shaviro raises give me pause. That is, if we can overcome the conceptual difficulty of distinguishing between tax “preferences,” on the one hand, and “structural” elements of the tax base, on the other (which is the difficulty rightly emphasized by anti-CTBers), Shaviro’s worry about interest group influence seems valid.

My primary concern, however, with the CTB norm as articulated by Shaviro and incorporated into his transition recommendations is

---


the tax-exceptionalism problem. This Review is not the place to pursue the issue at length, but I suspect that the public choice critique of tax preferences would apply with equal strength to any other type of governmental preference. Although Shaviro acknowledges that there are other areas of "distributional politics" besides the tax area, he does not seem to recognize that whenever Congress seeks to encourage investments of one type over another, whether through the tax laws or through direct expenditures, *all* the same interest group concerns arise.

His response is that tax preferences are structurally distinct from direct expenditures in certain respects. The most important one, in Shaviro's analysis, is the role of the endowment effect. As suggested above, in Shaviro's view, there is a heuristic bias that "induces people to draw an exaggerated distinction between money that is never paid in to the Treasury and money that is first paid in and then taken back out" (p 87). And that bias gives tax preferences "greater political appeal" than would the same subsidy in the form of a direct expenditure (id).

But even if that is so, the normative implication is unclear. For example, if policy experts were to conclude on substantive grounds that a given subsidy—a given preference for one kind of investment over another—would be welfare enhancing, perhaps structuring the proposal in the form of a tax preference (rather than as a direct expenditure) would be desirable, precisely because of the endowment effect, precisely because it would minimize taxpayer opposition and maximize the odds of getting the change adopted.

Granted, such use of a cognitive psychology/public choice argument is cynical. But that is the nature of such arguments, and I did not start it. Moreover, at least in this case, the policy experts could be forthright with policymakers (if not with taxpayers) about the reason for choosing the tax code as the location for the preference. Of course, if the word got back to taxpayers that a trick was being played on them, the endowment effect might be lost. But if the tax-preference endowment effect were so easily eliminated, then Shaviro's objection to tax preferences in the first place would lose a great deal of its force. In that case, we could simply start by explaining to taxpayers the equivalence of tax expenditures and direct expenditures and be done with it.

Does my anti-tax-exceptionalism argument imply that we should reject the CTB norm and Shaviro's public choice concerns about tax preferences? Of course not. Those concerns are valid. They provide ample reason to be skeptical of any tax preference (putting aside for now the difficulty of defining a tax preference). My point, however, is that such concerns also provide ample reason to be skeptical of any
governmental preferences, which means the public choice problem could be a much larger one than Shaviro's tax exceptionalism suggests.

Having said all of this, I, for one, am not yet sufficiently skeptical of the legislative lawmaking process to endorse a transition policy that would undermine lawmakers' ability ever to use tax (or, taken to the extreme, nontax) preferences in the future. And, apparently, neither is Shaviro. Indeed, his bottom line transition proposal for tax preferences—including (critically) his endorsement of the ANR norm as applied to tax-preference repeals—strikes me as a sensible middle position. Under such a rule, although many types of tax preferences would be discouraged (because of the default-premium effect discussed in Part I.C above), if Congress nevertheless determines that a tax preference is called for, the use of various "up front" tax subsidies would continue to provide a default-premium-minimizing option for lawmakers.61

Of course, from the perspective of the radical CTBer, the worry would be that, under Shaviro's proposal, the availability of transition-protected up front tax subsidies is just another loophole that needs to be slammed shut. Fortunately, from the anti-opportunism perspective, the ANR norm appears to be safe, for now.

61 Part of my unwillingness to eliminate tax and nontax preferences from lawmakers' policy toolbox is that I still tend to hold a somewhat optimistic view of policymaking in the long run. For a similar sentiment, with respect to lawmaking generally, see Levmore, 99 Colum L Rev at 1662 (cited in note 11) ("In short, although there is an endless stream of pessimistic arguments about the likely character of laws, there may be a comparable—less familiar but perhaps superior—set of optimistic arguments supporting the idea that most new law is good law.").

62 "Up front" subsidies include credits or deductions all or most of whose benefits are concentrated in the first year of investment and thus for which an ANR norm provides almost full transition-loss protection. See Logue, 94 Mich L Rev at 1192–94 (cited in note 4); Kaplow, 99 Harv L Rev at 587 (cited in note 3). See also Daniel S. Goldberg, Tax Subsidies: Onetime vs. Periodic: An Economic Analysis of the Tax Policy Alternatives, 49 Tax L Rev 305, 309–12 (1994).